nature portfolio

Peer Review File

Dissecting epigenetic heterogeneity in single-cell DNA methylomes with a unified framework

Corresponding Author: Professor Shengquan Chen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript by Tang et al. described scMethCraft, which is a software package that uses deep neural networks to impute missing values in single cell methylation data to improve the quality of existing analysis or enable new types of analysis. Although the imputation methodology is reasonable, the manuscript should present a rigorous validation of the imputation result, particularly regarding predictions in regions with strong cell-type specificity, and in a way that biologists without a machine learning background can understand. Lastly, other DMR calling methods with a greater statistical robustness should be considered.

1.The design of the neural network for imputing missing DNAm data is reasonable. However, the resolution of the methylation feature was ambiguous. Did scMethCraft try to impute cytosine methylation level for single cytosines or for a certain bin size?

2.The manuscript did not provide any assessment or validation of the imputation result. Such validation should be presented in a way that biologists can easily understand. What is the correlation between imputed values and true values? Importantly, is scMethCraft able to impute regions with cell-type-specific methylation levels correctly reflecting the identity of the cell type? The last point is related to the performance of the Similarity weighing module. If the authors mask DMR regions with strong cell-type specificity, such as the gene body of the GAD2 gene in inhibitory cells or the SATB2 gene in excitatory cells, does scMethCraft correctly impute the missing values that match the known cell-type specificity?

3.Line 148, the statement that “no existing method has been specifically designed for the dimensionality reduction of scDNAm matrix” seems inaccurate. There have been quite a few single cell methylation data generation / analysis published in recent years, including Tian et al. 2023, which provided the human brain data used by the current manuscript. What method did those publications use?

4.The choice to compare to SnapATAC2 seems unusual and needs to be justified. Tian et al. 2023 used a simple method to quantify methylation level across 100 kb bins. Why not compare to the method used by Tian et al., which is a much more commonly used method in existing publications?

5.The design of the batch correction benchmarking is questionable. Although batch effect may indeed exist between MTG and Pro datasets, the expectation that L2/3-IT cells from the two brain regions (MTG and Pro) should cocluster and overlap on a UMAP is not justified. Cortical regions are known to be somewhat different so L2/3-IT cells from the two regions may have meaningful biological differences. The authors should focus on MGE and CGE-derived interneurons or non-neuronal cells to benchmark the performance of batch correction. Interneurons and non-neuronal cells are less different between brain regions.

6.The statistical method for DMR calling (rank sum test) is not rigorous. It is known to be overly lenient because it treats correlated observations (single cells from the same sample) as independent observations. Most DMR calling method uses count-based statistics, such as the beta binomial test. The authors may be constrained by a lack of biological replicates. Nevertheless, a more rigorous DMR calling approach should be used.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors present scMethCraft, a computational method designed for single cell methylation analysis that accounts for the missingness patterns intrinsic to this modality and the characteristics of this complicated data type. Their method offers a latent representation of cells that supports downstream analysis and appears robust to batch effects. Overall, this is an interesting and potentially valuable contribution to the single-cell community.

While the algorithm reports gains against algorithms benchmarked against, the benchmarking is incomplete. The imputation capabilities of the network are useful, coupled with identification of DRM detection. This method needs to demonstrate its performance against other published methods and justify the sophisticated architecture.

This should include a comparison of modelling assumptions, scalability and outputs relative to scMethCraft. In particular, MethSCAn should be included in the comparisons (it feels like a natural comparison) and, given the apparent inspiration, scBasset. The latter is due to its influence on the genomic sequence module within scMethCraft. Also Melissa as another tool: https://doi.org/10.1186/s13059-019-1665-8
Kremer, L.P.M., Braun, M.M., Ovchinnikova, S. et al. Analyzing single-cell bisulfite sequencing data with MethSCAn. Nat Methods 21, 1616–1623 (2024). https://doi.org/10.1038/s41592-024-02347-x
Kapourani, CA., Argelaguet, R., Sanguinetti, G. et al. scMET: Bayesian modeling of DNA methylation heterogeneity at single-cell resolution. Genome Biol 22, 114 (2021). https://doi.org/10.1186/s13059-021-02329-8
Kapourani, CA., Sanguinetti, G. Melissa: Bayesian clustering and imputation of single-cell methylomes. Genome Biol 20, 61 (2019). https://doi.org/10.1186/s13059-019-1665-8

In figure 2, the performance of scMethCraft is satisfactory but the gains over competing methods, especially EpiScanpy is not dramatic. Considering that scMethCraft has a complex deep learning architecture, I would be curious to find out the run times of scMethCraft and competing methods. Furthermore, it is not clear to me, from Fig2EFG panels, how well OPCs, ODCs, and MGCs cells are better separated by scMethCraft. Either the color scale needs to be adjusted or a separate supplemental figure can display this.

scMethCraft's performance on batch correction is impressive, and a clear improvement.

The manuscript states:
"The input of scMethCraft is a single-cell DNA methylation cell-by-region matrix, in which each read count represents the average methylation level across CpG sites within a genomic tile for each individual cell. "
This could be clearer.
• Is the model input percentage methylation per region per cell (#sites methylated / #sites measured) or the read count as written?
• If it is the latter, does the count represent the number of methylated reads or the total measured reads?
• Is there a minimum coverage threshold / number of sites measured below which data are excluded for a bin within a cell? (in contrast to across cell filtering which I interpret the 80% missingness threshold to refer to). Was model sensitivity to this choice explored?

Lines 116-118 state that a 50-dimensional embedding is used for each region, how was this dimensionality chosen, and was model performance explored as a function of this choice?

The phrase "Upon completion of training, the similarity matrix $\mathbf{S}$ undergoes normalization as $\mathbf{S}_{\mathbf{norm}}$" should specify the normalization used (e.g., z-score normalization, min-max scaling) to aid reproducibility and clarity.

The model architecture suggests potentially high computational costs, particularly since the cell-cell similarity matrix becomes O(N2).
• Please provide details on the computational resources used (number and type of GPUs, runtime, memory usage)
• Please include these quantities as a function of the number of cells and genomic bins used
• Comparing these costs to competitors would further strengthen this work and encourage users

The reported performance differences under clustering metrics (e.g., ARI, AMI) appear modest. However, these are paired comparisons, so relative improvements may be obscured by summary statistics (as the significance of the test statistic that accounts for this shows). Consider presenting visualizations of the relative changes (i.e., relative disimprovement compared to scMethCraft) for the key metrics.

The manuscript briefly notes the intra-batch weighting parameter, β, is set to 0.2.
• Was sensitivity to this choice explored by the authors? I would like to see this.
• Are there dataset characteristics that might inform choices of the magnitude? E.g., batch sizes, coverage.
I think a short discussion for estimating β from data, or future directions such as making β adaptive, would enhance the paper.

The imputation capabilities of scMethCraft looks promising, as cell-to-cell similarity module in the network presumably should aid in imputation. There is one clear challenge with the imputation. It seems the average methylation level's gradient within specific clusters in Fig4 EFGH become much more pronounced after imputation. These gradients are much less pronounced in raw data. Furthermore, the cells that are labeled green (TENs) seemingly get higher in average methylation after imputation (panel H), whereas blue (non-neuronal cells) get lower average methylation after imputation. This can be a by product of imputation, naturally, as cells share more information across other cells. Authors should quantify this effect by looking at average methylation per cell per cell type before and after imputation to ensure imputation does not overemphasize certain patterns in specific cell types.

Can scMethCraft also identify hypermethylated DMRs? The results focus on hypomethylation.

I appreciate the fact that authors broke their network down into different modules and show the performance. It is somewhat counterintuitive yet interesting that it is the one-hot-encoded DNA sequence modality that yields the highest improvement in clustering, whereas my expectation would be that the improvement comes from cell-to-cell similarity module, as that allows improve data sparsity. Furthermore, the DNA sequence elements exhibit no variation across cell types; thus, it is surprising to see that that information can help identify better classification. I think this is worth touching upon in the discussion further, especially if the authors have clear insights into this, as the discussion section is relatively short.

I could not find clearly where the labels for performance evaluation are obtained from, especially in for Figure 2. Did the authors generate these labels or are they obtained from the original studies?

The authors have created a clean and user-friendly github repository for this method. However, in my review I noticed that scBasset has strongly informed the genomic sequence module. I feel like this should be acknowledged explicitly within the paper given this module is a central contributor to the success of the method (based on figure 6). Proper attribution would improve transparency and clarify novelty relative to prior work.
Minor
The size of the genomic windows used in the algorithm is barely mentioned in main text. No reason why that number has to be only in the methods section, as this is an important parameter of the model.

In the intro, authors claim " 57 observed in scRNA-seq or scATAC-seq data4, 11. Finally, gene expression often inversely 58 correlates with methylation levels." and argue after that other algorithms for sc-RNA and sc-ATAC may not perform well because they are not capturing an inverse correlation. The actual concern is that methylation is typically represented by percentage of GCs methylated rather than a count based paradigm, as in RNA and ATAC. If methylation is also represented as counts, there is not reason the anti-correlation of methylation to gene expression (which is not always the case) should make standard dimensionality reduction methods less applicable.

Typo
148 knowledge, no exiting method have been specifically designed for the dimensionality reduction

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Wang et al. introduce a new neural network-based model for the analysis of single cell DNA methylation data, named scMethCraft. Their model uses a hybrid neural network architecture to predict DNA methylation within genomic bins. The major novelty of this approach is the incorporation of DNA sequence information into the modeling of DNA methylation data, achieved through the hybrid network. Overall this fills an important need in the area of single-cell DNA methylation analysis, and the authors demonstrate versatile use-cases including cell clustering, batch correction, and data imputation. However, there are a few points that require further consideration, outlined below:

There is no information given about the computational resources required to train the model. It is important that the authors can show the time required for training their model, how this scales with increasing numbers of cells, what GPU hardware is required, and how this compares with other competing methods.

The authors also present the identification of DMRs, and propose that their model is better able to identify these regions compared to other methods. However, these differential tests have been performed on the "enhanced" data from the model, meaning that the cells are no longer independent observations. This violates the assumptions of the statistical test and will result in inflated significance values and inflated false discovery rate. The methods also do not state that any multiple testing correction was performed on the p-values obtained.

The authors focus their analysis and evaluations solely on the brain; some examples in other tissues are required to demonstrate that the model can be applied to a diversity of tissues rather than only the brain.

In the evaluation of data integration results, the authors have only used examples where the same cell types are present across all datasets. An important aspect of integration performance that is therefore overlooked is the ability of the model to

perform accurate integration when some of the cell types are unshared across datasets.

In the assessment of cell annotation performance, the authors have performed clustering across a variety of resolutions until they obtain the same number of clusters as are annotated cell types. This likely inflates performance of the cell annotation method, as the authors ensure the clustering is somewhat similar to the cell annotations by optimizing the resolution. Instead the authors should show performance with clustering at a fixed resolution, chosen independently of the cell type labels.

Furthermore, it was unclear where the ground truth cell type labels used for evaluating cell annotation performance originated from.

Finally, more in-depth benchmarking comparisons with other methods are needed. This includes Epiclomal, DeepCpG, MambaCpG, Melissa.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have successfully addressed my comments, and I have no further concerns.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The revised manuscript includes a number of additional inclusions that satisfy my initial comments.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
In their revised manuscript, the authors have addressed most of my initial comments. The additional benchmarking and discussion of other tools is a good addition, as is the additional integration evaluations. However, I still have concerns regarding the differential testing methods used in the manuscript, and more broadly the implied recommendation for users to follow this approach with their own data using scMethCraft.

The authors have now used Bonferroni correction, and state that it helps mitigate the inflation of significance arising from non-independence among observations. This is not accurate: Bonferroni controls the family-wise error rate and assumes independent observations, it does not mitigate inflated significance arising from non-independent observations.

Adopting Bonferroni correction addresses the issue of multiple testing, but fundamentally the differential test is statistically flawed due to the violated assumption of independent observations. This is an important point, because the results of these tests will not be reliable, and enabling this in the scMethCraft framework is concerning as it may lead to false discoveries in the field.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
The authors have now addressed my concerns regarding the statistical testing. The added words of caution should help to inform readers of potential false positives generated through these differential tests.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

To the reviewers,

We want to further thank the four reviewers for their time reviewing and their constructive comments on our manuscript entitled “Dissecting epigenetic heterogeneity in single-cell DNA methylomes with a unified framework”. We have thoroughly revised and enhanced our manuscript based on their comments. In addition to the minor changes, we have made the following six major modifications in our revised manuscript, with the revisions highlighted in red. The revisions are summarized in Figs. 2-5, Supplementary Notes S2-S6, Supplementary Figs. S1-S5, S8-S10, S15-S31, and Supplementary Tables S1, as well as the main text and the Methods section of the revised manuscript.

1) We validated the batch integration performance of scMethCraft under a broader range of scenarios. We systematically investigated the risk of batch overcorrection and introduced an adaptive strategy for selecting the batch integration parameter. This extension enables scMethCraft to adaptively balance batch effect removal and the preservation of biologically meaningful heterogeneity across datasets with varying batch compositions (Fig. 3, Supplementary Figs. S8-S10, S28-S31, Supplementary Notes S3, S5).

2) We substantially expanded the benchmarking by incorporating additional baseline methods and dataset. We also clarified the rationale for excluding certain methods whose modeling assumptions or scalability constraints make them unsuitable for region-level single-cell DNA methylation (scDNAm) data analysis. These comparisons further demonstrate the general applicability of scMethCraft (Figs. 2, 4, Supplementary Figs. S1-S5, S19, Supplementary Notes S2, S6, Supplementary Table S1).

3) We further validated the data enhancement (imputation) capability of scMethCraft under multiple challenging settings, including additional masking experiments on cell-type-specific regions. These experiments demonstrate that scMethCraft reliably enhances scDNAm signals while preserving cell-type-specific methylation patterns (Fig. 4, Supplementary Figs. S15-S16, S21-S22).

4) We conducted a detailed benchmark of runtime and memory usage for scMethCraft across datasets of various sizes. In addition, we introduced an optional low-rank approximation extension as fast implement of scMethCraft to improve scalability for large datasets without compromising performance. These results demonstrate that scMethCraft achieves acceptable efficiency and favorable scaling behavior for contemporary scDNAm data analysis (Supplementary Figs. S18, S20).

5) We revised the DMR analysis to address statistical assumptions raised by the reviewers. Specifically, we clarified the rationale for adopting the Wilcoxon rank-sum test for DMR identification in scMethCraft and introduced appropriate multiple-testing correction. These

revisions improve the rigor of the downstream differential analyses (Fig. 5, Supplementary Fig. S17).

6) We performed additional sensitivity analyses to systematically evaluate the robustness of scMethCraft with respect to genomic bin size, region filtering thresholds, and embedding dimensionality, thereby providing practical guidance for users on parameter selection in different analytical settings (Supplementary Figs. S23-S27, Supplementary Note S4).

Next, we will clarify the comments raised by the reviewers point-by-point, with the reviewers' points in black and our responses in blue.

REVIEWER #1 (REMARKS TO THE AUTHOR)

The manuscript by Tang et al. described scMethCraft, which is a software package that uses deep neural networks to impute missing values in single cell methylation data to improve the quality of existing analysis or enable new types of analysis. Although the imputation methodology is reasonable, the manuscript should present a rigorous validation of the imputation result, particularly regarding predictions in regions with strong cell-type specificity, and in a way that biologists without a machine learning background can understand. Lastly, other DMR calling methods with a greater statistical robustness should be considered.

RESPONSE

We deeply appreciate the reviewer for the affirmation of our work and the enthusiastic assessment, as well as the thoughtful comments, which helped us to improve the manuscript. We have carefully considered and answered every comment raised by the reviewer and made revisions in the corresponding section of the revised manuscript.

COMMENT 1

1. The design of the neural network for imputing missing DNAm data is reasonable. However, the resolution of the methylation feature was ambiguous. Did scMethCraft try to impute cytosine methylation level for single cytosines or for a certain bin size?

RESPONSE TO COMMENT 1

We appreciate the reviewer's recognition of our methodology and apologize for any lack of clarity in our manuscript. First, it is necessary to clarify that scMethCraft analyzes single-cell DNA methylation (scDNAm) data using a certain bin as the basic unit. This strategy represents a very common paradigm in scDNAm data analysis[1-4], as it provides a simple and effective way to reduce data dimensionality, improve the signal-to-noise ratio, and enhance interpretability, while also yielding data formats that are compatible with widely used pipelines such as Seurat[5] and Scanpy[6]. Thus, scMethCraft is designed under this bin size-based paradigm. Its hybrid network architecture is designed to extract latent features associated with methylation levels from DNA sequences of reference genomic regions and to accurately learn and predict the methylation level of each individual cell within the corresponding region.

Secondly, in this study, we use a bin size of 10 kbp by default for scMethCraft as well as for baselines. However, the flexible network structure of scMethCraft allows it to be readily extended to different bin sizes. For example, in the ACC dataset, we partitioned the data using bin sizes of 5 kbp, 10 kbp, 20 kbp, 50 kbp, and 100 kbp. We observed that bin sizes below 100

kbp have a limited impact on visualization and clustering results (Figs. R1, R2). Nevertheless, some related cell types, such as the Vip, Sncg, Lamp5, and Lamp5-Lhx6 cells, may become more difficult to distinguish with a large bin size, as indicated by the purple boxes.

In summary, we again apologize for this ambiguity and have revised the manuscript to eliminate any potential misunderstanding for readers. In addition, we have incorporated bin size as an optional parameter in scMethCraft and added relevant discussion in Supplementary Note S4 to accommodate the needs of different users.

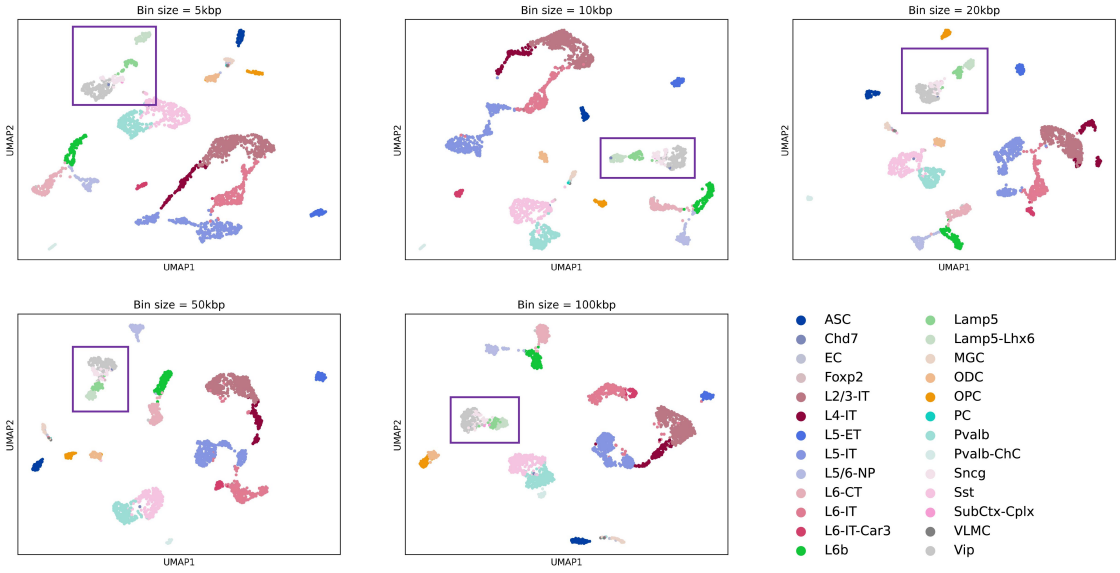


Fig. R1 UMAP visualization of the cell embedding created by scMethCraft with varying bin sizes.

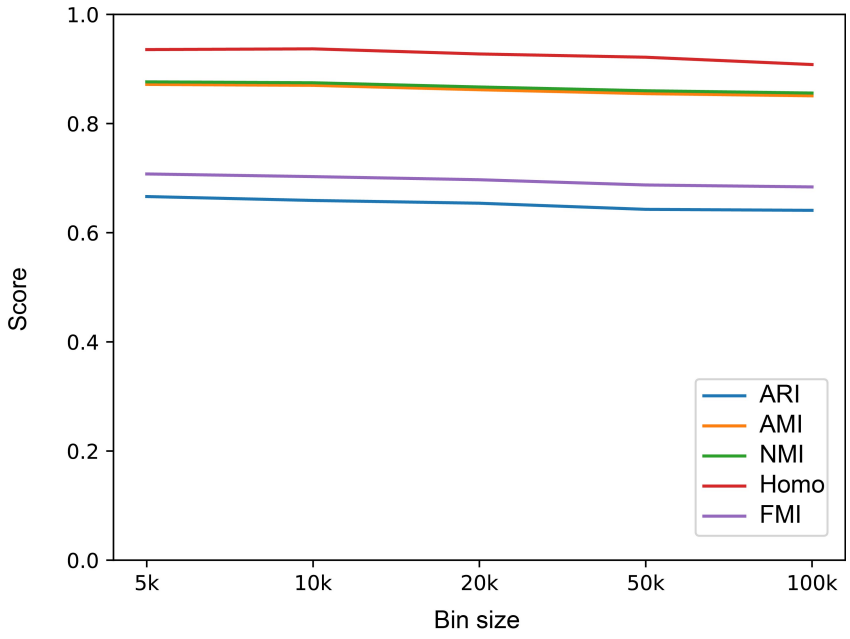


Fig. R2 Cell clustering performance of various bin sizes.

COMMENT 2

2. The manuscript did not provide any assessment or validation of the imputation result. Such validation should be presented in a way that biologists can easily understand. What is the correlation between imputed values and true values? Importantly, is scMethCraft able to impute regions with cell-type-specific methylation levels correctly reflecting the identity of the cell type? The last point is related to the performance of the Similarity weighing module. If the authors mask DMR regions with strong cell-type specificity, such as the gene body of the GAD2 gene in inhibitory cells or the SATB2 gene in excitatory cells, does scMethCraft correctly impute the missing values that match the known cell-type specificity?

RESPONSE TO COMMENT 2

We thank the reviewer for this important comment. We agree that rigorous and biologically interpretable validation of imputation results is essential, particularly for regions with strong cell-type-specific methylation patterns. Below, we address the three points raised by the reviewer.

First, regarding the correlation between imputed values and true values. The reviewer points out that a direct correlation between imputed and ground-truth values is an intuitive validation strategy. However, in real single-cell sequencing data, a noise-free and fully observed ground truth does not exist[7]. Consequently, direct correlation-based validation is most commonly performed on simulated data in prior studies[7-9].

Due to the absence of simulation frameworks that can realistically model single-cell DNA methylation data at the matrix level, we relied on evaluation strategies based on real datasets. Specifically, to enable biologically meaningful assessment, we adopted the validation strategy proposed by SCALE[7]. This approach assumes that averaging single cells belonging to the same biological cell type yields a meta-cell that serves as a reasonable approximation of the underlying ground-truth signal. Following this strategy, we computed the correlation between the imputed methylation values of individual cells and those of their corresponding meta-cells. scMethCraft substantially improves the Spearman correlation between individual cells and their respective meta-cells (Fig. R3), indicating a more faithful reconstruction of the underlying methylation distribution. These results demonstrate scMethCraft's ability to infer accurate methylation levels by jointly leveraging learned sequence features and intercellular similarity. In contrast, other scDNAm imputation methods exhibit overall weaker performance, which is closely related to their underlying design principles and implementation strategies (Supplementary Note S2).

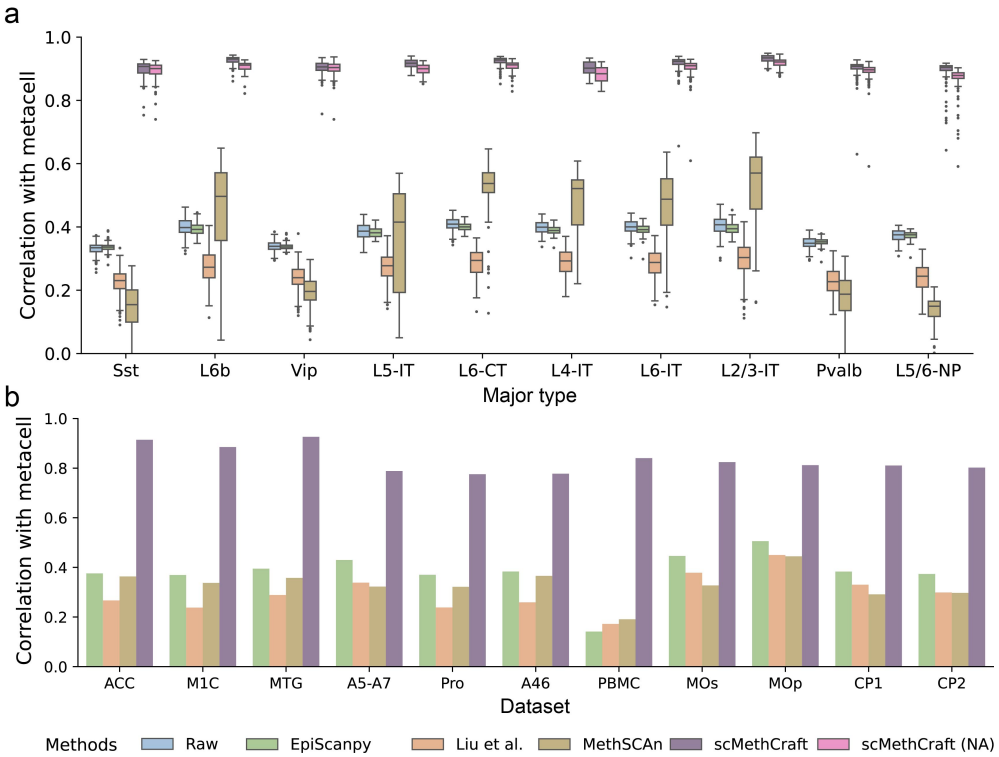


Fig. R3 Data enhancement performance of scMethCraft. a, Spearman correlation between the average methylation level of each single cell and its corresponding meta-cell, computed for the raw matrix and for enhanced matrices generated by different methods across cell types in the ACC dataset. scMethCraft (NA) denotes correlations computed using only regions originally annotated as NA in the raw matrix. **b,** Spearman correlation between enhanced datasets and their corresponding meta-cells across 11 datasets processed by different methods.

Second, regarding regions with strong cell-type-specific methylation, we have evaluated whether scMethCraft correctly reconstructs regions with cell-type-specific methylation patterns that reflect known cell identities in the original manuscript. Specifically, we first generated heatmaps of the raw scDNAm matrix and the scMethCraft-enhanced matrix (Fig. R4), focusing on the top 10 cell-type-specific hypomethylated regions identified using

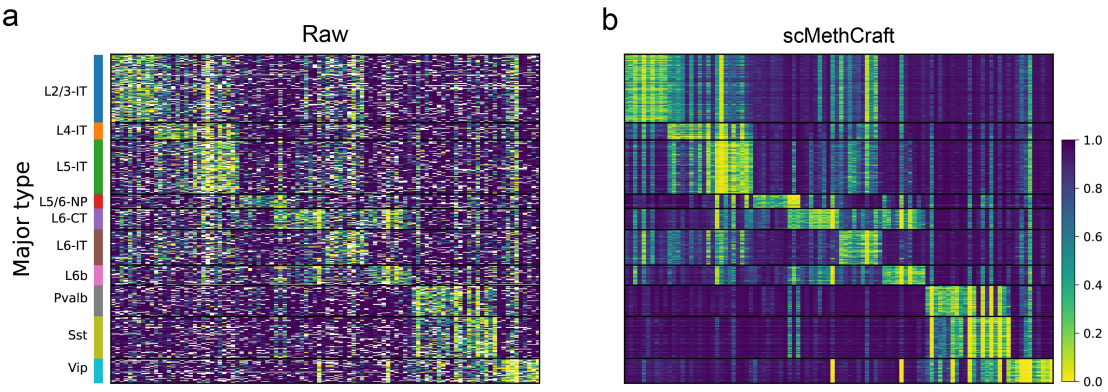


Fig. R4 Heatmaps of the (a) raw matrix and (b) scMethCraft enhanced matrix. The top 10 cell-type-specific hypomethylated regions identified using EpiScanpy are shown.

EpiScanpy on the raw data. In the raw matrix, extensive NA values and technical noise (shown as white regions) obscure cell-type-specific patterns. After enhancement with scMethCraft, NA values are effectively estimated, and distinct methylation patterns across cell types become clearly discernible.

We further examined regions overlapping the *STMN2* gene, which is labeled as a marker gene for excitatory neurons[10], and visualized methylation levels by heatmaps. As shown in Fig. R5a, in the scMethCraft-enhanced data, *STMN2* exhibits a clear cell-type-specific methylation pattern, characterized by hypomethylation in telencephalic excitatory neurons and hypermethylation in other neuronal populations. In contrast, the corresponding patterns in the raw data are considerably noisier. The results showed that scMethCraft enhances these

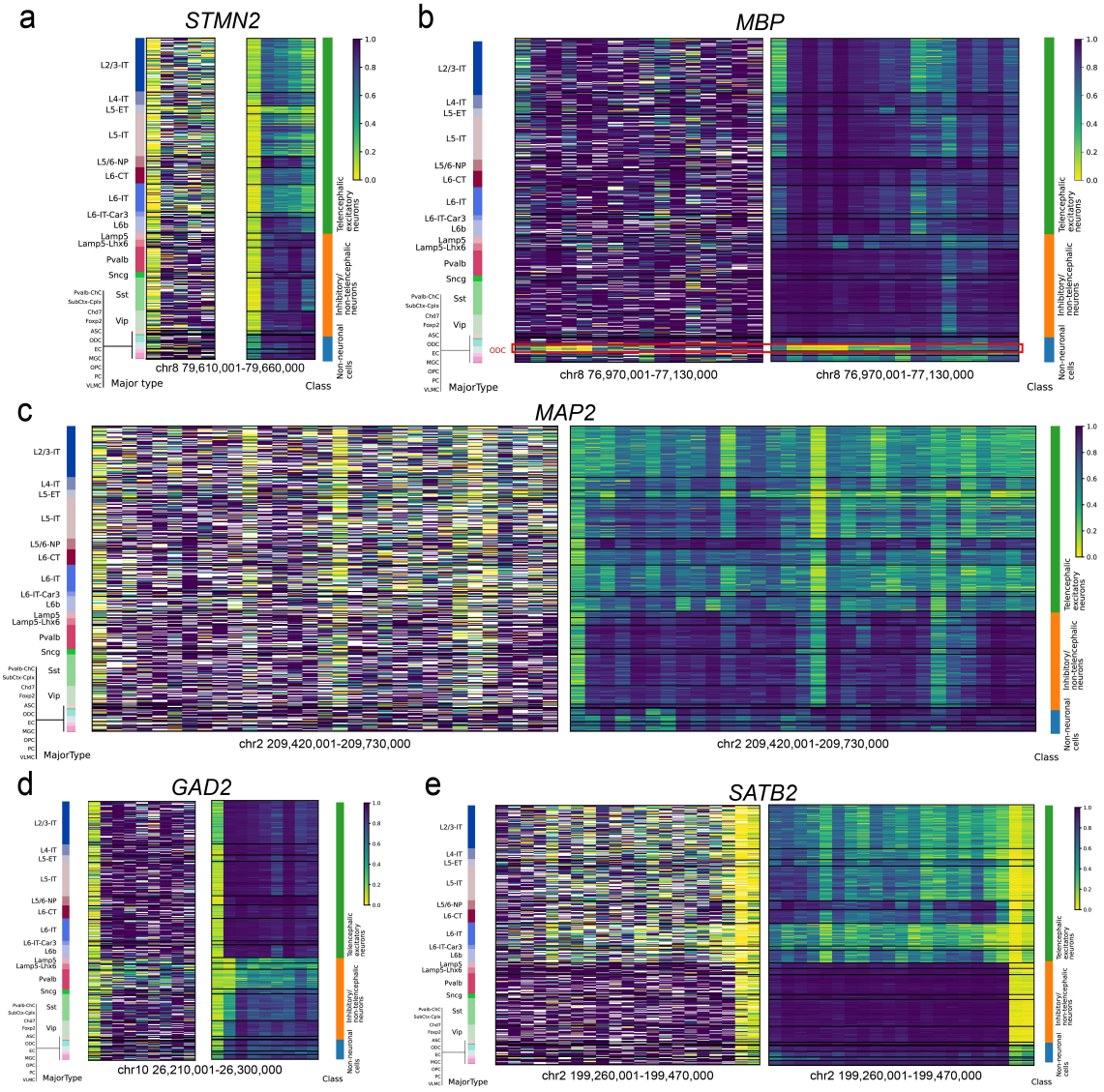


Fig. R5 Heatmaps of the raw data and scMethCraft-enhanced data. Cell-type-specific regions overlapping with the **a,** *STMN2*, **b,** *MBP*, **c,** *MAP2*, **d,** *GAD2*, and **e,** *SATB2* genes are shown**.** For each heatmap, the left column corresponds to the raw data, while the right column corresponds to the scMethCraft-enhanced data.

biologically meaningful patterns without requiring prior cell-type annotations. Similar observations were made for *MBP* in oligodendrocytes (Fig. R5b)[11], as well as *MAP2* as a neuronal marker (Fig. R5c)[12], *GAD2* in inhibitory neurons (Fig. R5d)[13], and *SATB2* in excitatory cells[14] (Fig. R5e, as suggested by the reviewer). In the raw data, these patterns are difficult to discern due to noise and missing values, whereas they become clearly resolved after scMethCraft enhancement.

Finally, inspired by the reviewer's suggestion, we conducted an additional masking experiment to assess whether scMethCraft can correctly impute fully masked, cell-type-specific regions, such as those overlapping known marker genes, while preserving their expected cell-type specificity. In the ACC dataset, we masked methylation values overlapping *STMN2*, *GAD2*, *MAP2*, and *SATB2* in the L2/3-IT and Lamp5 cell types by setting these entries to NA, and then reran scMethCraft to impute them. Figs. R6a,b and R7a,b show that even when

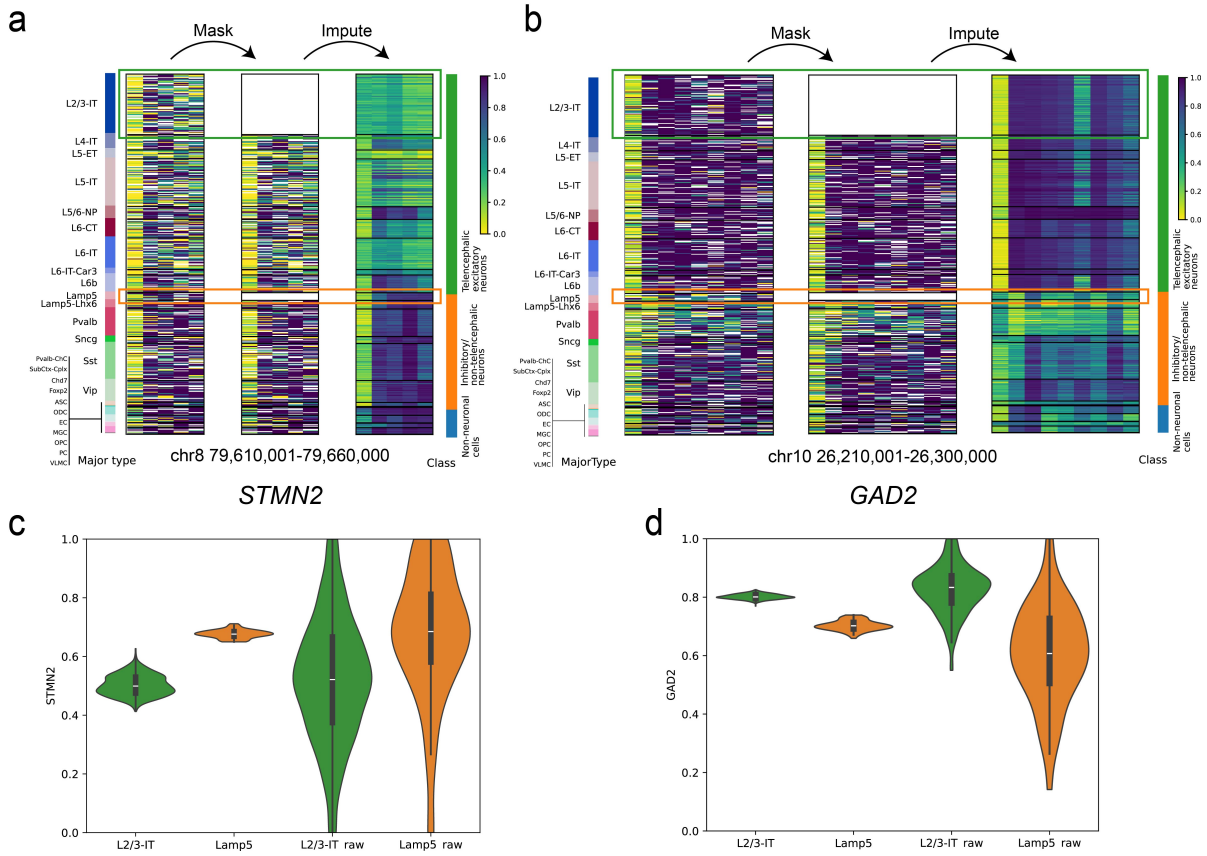


Fig. R6 scMethCraft effectively recovers masked cell-type-specific regions. a-b, Heatmaps of **(a)** *STMN2* and **(b)** *GAD2* gene-overlapping regions. For each panel, the three heatmaps from left to right correspond to the raw data before masking, the raw data after masking, and the imputed results obtained by applying scMethCraft to the masked data. Green boxes indicate L2/3-IT cells, and orange boxes indicate Lamp5 cells. **c-d**, Violin plots showing the average methylation levels of L2/3-IT and Lamp5 cells in the **(c)** *STMN2* and **(d)** *GAD2* gene overlapping regions before masking and after imputation. The maximum width of each violin is identical, and kernel density estimates have been truncated to the data range.

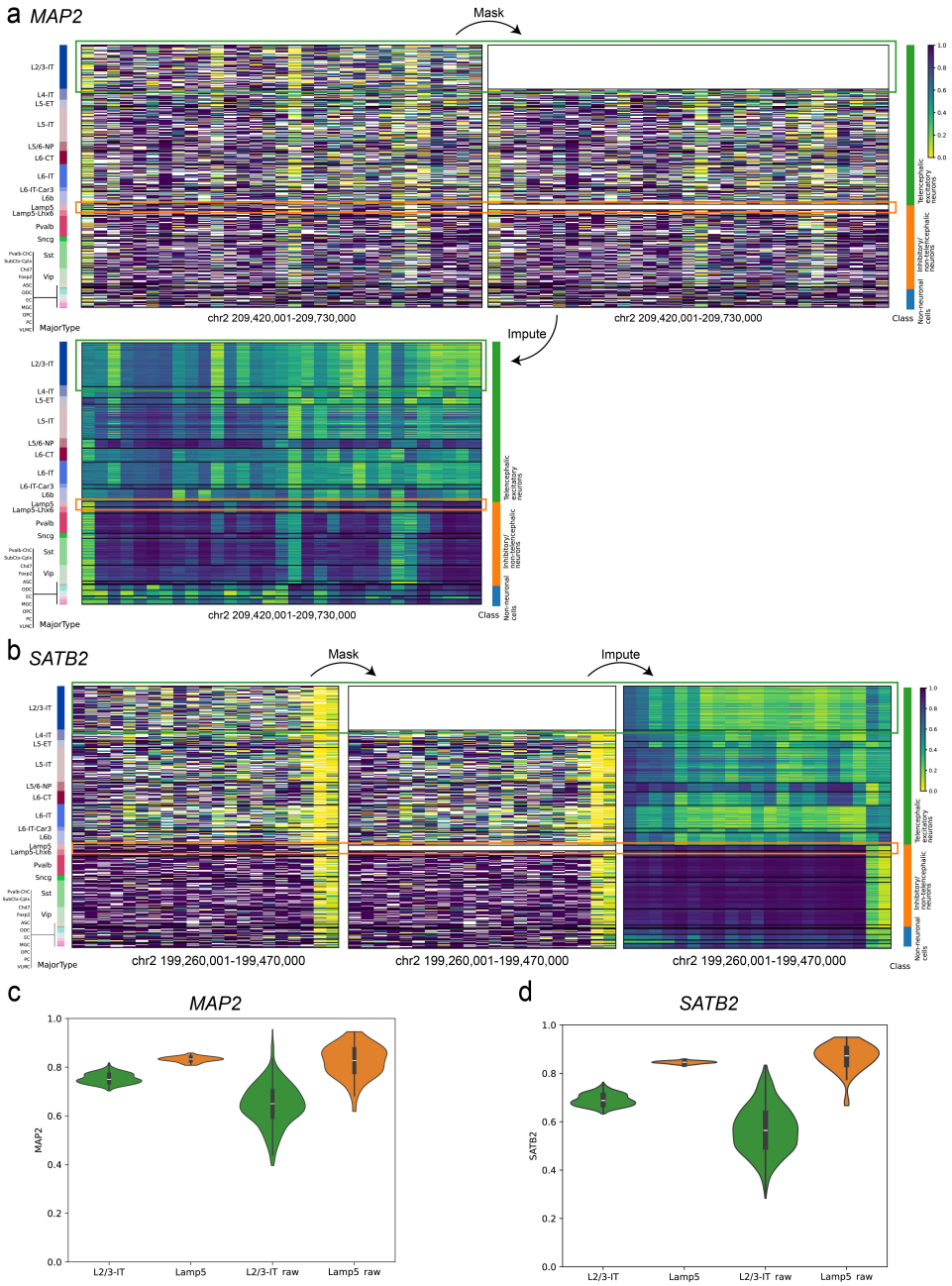


Fig. R7 scMethCraft effectively recovers masked cell-type-specific regions. a-b, Heatmaps of **(a)** *MAP2* and **(b)** *SATB2* gene-overlapping regions. For each panel, the three heatmaps from left to right correspond to the raw data before masking, the raw data after masking, and the imputed results obtained by applying scMethCraft to the masked data. Green boxes indicate L2/3-IT cells, and orange boxes indicate Lamp5 cells. **c-d**, Violin plots showing the average methylation levels of L2/3-IT and Lamp5 cells in the **(c)** *MAP2* and **(d)** *SATB2* gene overlapping regions before masking and after imputation. The maximum width of each violin is identical, and kernel density estimates have been truncated to the data range.

cell-type-specific regions exhibit extensive missingness, scMethCraft is able to accurately recover methylation values that match known cell-type-specific patterns by leveraging intercellular similarity. We further visualized the methylation levels of the two cell types in the corresponding gene regions using violin plots (Figs. R6c,d and R7c,d) before masking and after

imputation. Although the average methylation levels before masking and after imputation are not identical, scMethCraft clearly preserves and reflects the underlying cellular heterogeneity between these cell types.

In summary, scMethCraft provides accurate and biologically faithful imputation of scDNAm data, as demonstrated by quantitative correlation-based evaluation, recovery of known cell-type-specific methylation patterns, and targeted masking experiments on biologically meaningful regions. We again apologize for any confusion caused by the original presentation and have revised both the main text and the Supplementary File (Supplementary Figs. S15, S21-S22) to make these validation results more explicit and accessible.

COMMENT 3

3. Line 148, the statement that "no existing method has been specifically designed for the dimensionality reduction of scDNAm matrix" seems inaccurate. There have been quite a few single cell methylation data generation / analysis published in recent years, including Tian et al. 2023, which provided the human brain data used by the current manuscript. What method did those publications use?

RESPONSE TO COMMENT 3

We thank the reviewer for pointing out this potential ambiguity. We apologize that the original wording was not sufficiently clear. Our intention was not to suggest that scDNAm data cannot be analyzed or that no prior methods exist for single-cell DNA methylation analysis. Rather, we intended to emphasize that there is a lack of computational methods that are specifically designed for dimensionality reduction of binned scDNAm matrices while explicitly accounting for their unique data characteristics, such as unique data distribution and explicit NA values.

As stated in the original manuscript, a common and practical strategy in scDNAm data analysis is to aggregate methylation signals across genomic regions through genome binning, for example, by dividing the genome into fixed-length tiles and computing average CpG methylation levels per cell per region[1-4]. This procedure yields a cell-by-region scDNAm matrix that reduces data dimensionality, improves signal-to-noise ratio, and enhances interpretability. In practice, the downstream analysis of such scDNAm matrices often relies on pipelines originally developed for scRNA-seq or scATAC-seq data[3, 15]. However, due to substantial differences in data structure between scDNAm matrices and other single-cell modalities, including explicit missing values and distinct sparsity patterns, these conventional pipelines are not optimal when directly applied to scDNAm matrices. It is in this specific context that we intended to state that most existing approaches are not specifically tailored for dimensionality reduction tailored to the characteristics of scDNAm matrices.

To avoid potential misunderstanding, we have revised the wording in the manuscript. In addition, we provide further discussion and benchmarking of the scDNAm data analysis strategy used in Tian et al.[1] in Response #4 to Reviewer #1.

COMMENT 4

4. The choice to compare to SnapATAC2 seems unusual and needs to be justified. Tian et al. 2023 used a simple method to quantify methylation level across 100 kb bins. Why not compare to the method used by Tian et al., which is a much more commonly used method in existing publications?

RESPONSE TO COMMENT 4

We thank the reviewer for this thoughtful comment. We compared scMethCraft to SnapATAC2 because SnapATAC2 explicitly claims general applicability to multiple single-cell omics modalities, including single-cell DNA methylation data[16].

In contrast, the method used in Tian et al.[1] follows a pragmatic and widely used analysis pipeline proposed by Liu et al.[2], rather than a standalone algorithm specifically developed for scDNAm dimensionality reduction. In brief, this approach quantifies methylation levels across fixed genomic bins by first estimating cell-specific beta-binomial priors for methylation rates from raw methylcytosine and coverage counts, and then computing posterior methylation ratios that are normalized by each cell's global mean methylation level. This procedure effectively removes missing values and stabilizes estimates in low-coverage regions, resulting in imputed entries that are the global average methylation level of each cell, after which PCA is applied for dimensionality reduction and downstream analysis. This strategy primarily imputes all NA values in each cell using an estimate of a global distributional expectation and then applies PCA for dimensionality reduction. Moreover, it was not originally introduced as an independent method paper. Thus, we did not include it as a separate baseline in the initial comparison. We apologize for this omission. Following the reviewer's suggestion, we have now added comparisons to this pipeline (as implemented in Liu et al.[2]) as well as to MethSCAn[3], a recently published method for scDNAm data analysis, as additional baselines.

The expanded benchmarking results demonstrate that scMethCraft consistently achieves superior performance in both dimensionality reduction and imputation tasks across multiple datasets, including comparisons with the newly added baseline methods (Fig. R8). We have revised the manuscript accordingly to incorporate these additional comparisons and to more clearly highlight the advantages of scMethCraft (Figs. 2, 4, Supplementary Figs. S1-S5, Supplementary Notes S2, S6).

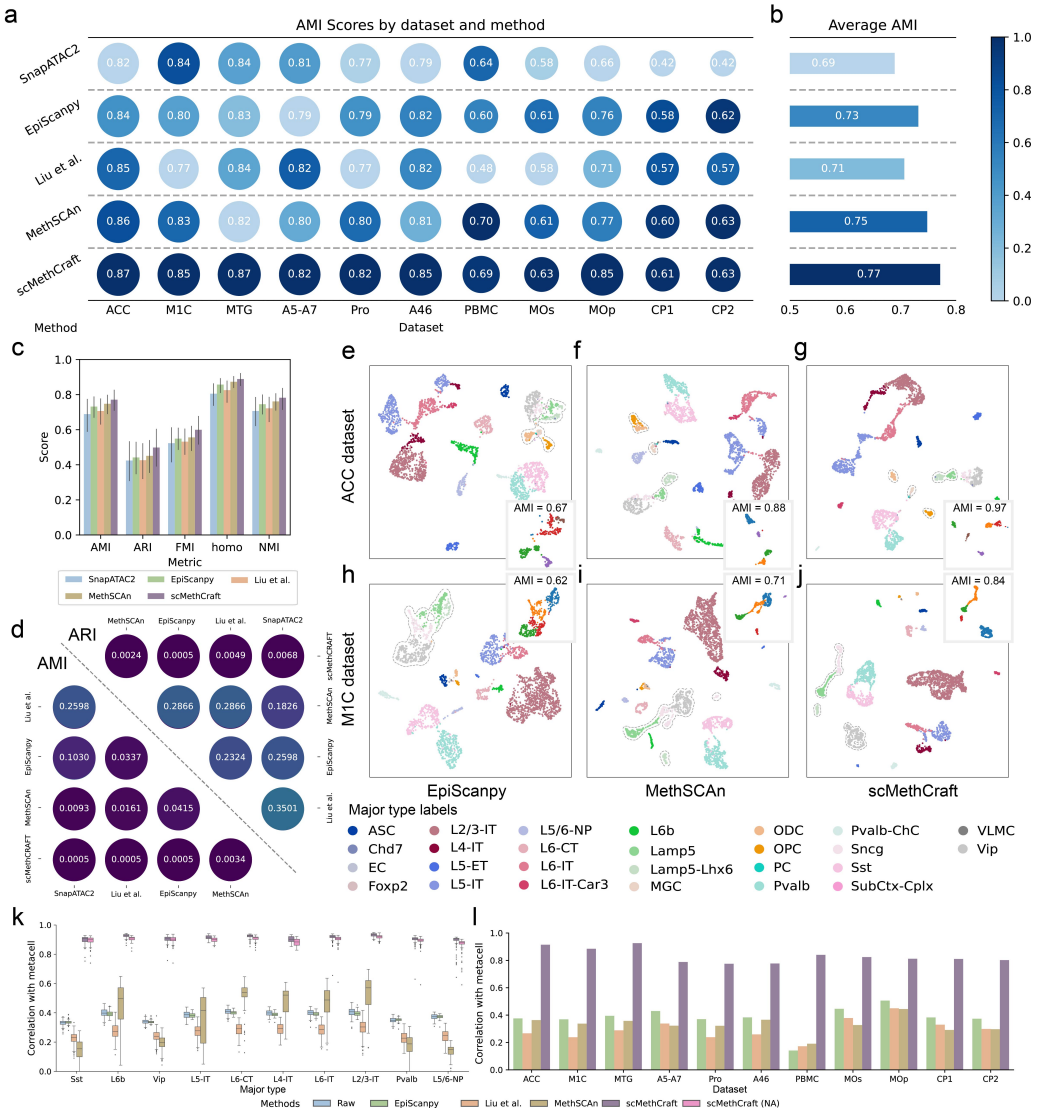


Fig. R8 scMethCraft consistently achieves superior performance in both dimensionality reduction and imputation tasks. a, Cell clustering performance assessed by AMI scores with respect to major cell types across different dimensionality reduction methods. Colors represent min-max normalized AMI scores within each dataset, where lighter blue indicates the lowest-performing method and darker blue indicates the highest-performing method. **b,** Average AMI scores with respect to major cell types, based on embeddings provided by different methods across 11 independent datasets. **c,** Cell clustering performance of embeddings generated by different methods was evaluated using five metrics across 11 datasets. **d,** One-sided paired Wilcoxon signed-rank tests were used to compare performance between methods. Each cell shows the P-value for the hypothesis that the method in the column outperforms the method in the row in terms of AMI or ARI across the 11 datasets. **e-j,** UMAP visualization of cell embeddings provided by different methods on the ACC and M1C datasets. Main panels are colored by major cell types, while inset panels are colored by cell clusters. The AMI values shown were computed only using the cell types highlighted by the boxes in the main panels. **k,** Spearman correlation between the average methylation level of each single cell and its corresponding meta-cell, computed for the raw matrix and for enhanced matrices generated by different methods across cell types in the ACC dataset. scMethCraft (NA) denotes correlations computed using only regions originally annotated as NA in the raw matrix. **l,** Spearman correlation between enhanced datasets and their corresponding meta-cells across 11 datasets processed by different methods.

COMMENT 5

5. The design of the batch correction benchmarking is questionable. Although batch effect may indeed exist between MTG and Pro datasets, the expectation that L2/3-IT cells from the two brain regions (MTG and Pro) should cocluster and overlap on a UMAP is not justified. Cortical regions are known to be somewhat different so L2/3-IT cells from the two regions may have meaningful biological differences. The authors should focus on MGE and CGE-derived interneurons or non-neuronal cells to benchmark the performance of batch correction. Interneurons and non-neuronal cells are less different between brain regions.

RESPONSE TO COMMENT 5

We thank the reviewer for these insightful comments, which prompted us to re-examine the biological interpretation of batch integration results across cortical regions. First, we thank the reviewer for pointing out that the expectation for L2/3-IT cells from two distinct cortical regions to cocluster and overlap in UMAP space is not biologically well justified. Cortical regions exhibit regional specialization, and thus L2/3-IT neurons from different regions may indeed harbor meaningful biological differences. We acknowledge that this assumption was previously overlooked. In the original manuscript, we have noted that L2/3-IT cells from MTG and Pro do not fully overlap in integrated UMAP space; however, this phenomenon was not explicitly attributed to functional and biological differences between neurons from distinct cortical regions. We have therefore revised the relevant section to more clearly emphasize scMethCraft's ability to preserve biologically meaningful heterogeneity between neurons from different cortical regions during batch integration, thereby avoiding overcorrection. Second, we appreciate the reviewer's suggestion to focus batch correction benchmarking on cell populations with more conserved properties across cortical regions, such as MGE- and CGE-derived interneurons as well as non-neuronal cells. Following this recommendation, we have performed additional batch correction analyses on these cell populations. The results of these analyses provide a more appropriate and biologically grounded evaluation of batch correction performance. We describe these additional analyses and results in detail below.

In the original manuscript, we have observed that most L2/3-IT cells from MTG and Pro exhibited substantial overlap in UMAP space after batch integration, while a subset of L2/3-IT cells from the Pro region remained separated (Fig. R9a,b). To further investigate this observation, we examined the original subtype annotations provided by the authors of the source study. In the original publication[1], experienced neuroscientists manually annotated subtypes based on differentially methylated genes (DMGs), with each annotated subtype differing from others by at least five DMGs. This annotation strategy indicates that cells from different subtypes harbor meaningful biological differences, whereas cells within the same

subtype are expected to be more similar. Notably, the non-overlapping L2/3-IT cells from the Pro region predominantly belonged to subtype 7 and subtype 9 (Fig. R9c), which were largely absent in the MTG region (Fig. R9d). In contrast, L2/3-IT subtypes shared between MTG and Pro were well integrated by scMethCraft and remained proximal in UMAP space, preserving subtype-specific coherence.

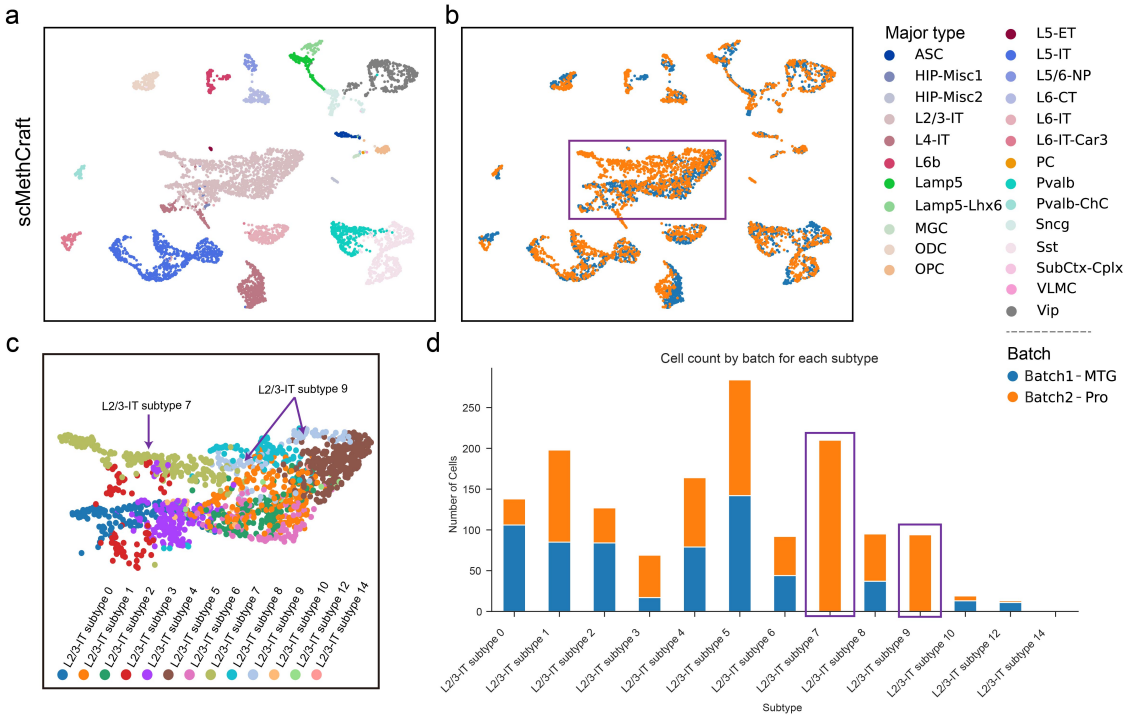


Fig. R9 scMethCraft preserves L2/3-IT subtypes during batch integration. a-b, UMAP visualization of the integrated embedding provided by scMethCraft on the MTG-Pro dataset colored by **(a)** Major type and **(b)** batch. The purple box indicates L2/3-IT cells. **c,** The L2/3-IT cells of scMethCraft's integrated embedding, colored by subtype. **d,** Subtype-specific batch composition of L2/3-IT cells.

In the revised manuscript, we further extended this analysis to other excitatory neuronal populations, including L4-IT and L5-IT neurons. We observed that L4-IT cells (Fig. R10a) from the two cortical regions showed minimal overlap in UMAP space, consistent with the fact that they belong to distinct biologically divergent subtypes (Fig. R10b). In contrast, L5-IT cells, which exhibit lower biological divergence across batches, were more effectively integrated (Fig. R10c,d). These additional observations further support the interpretation that the degree of integration achieved by scMethCraft is aligned with underlying biological similarity. These results indicate that while regional differences do exist among excitatory neurons, scMethCraft effectively integrates cells with limited biological divergence across cortical regions while preserving heterogeneity among functionally distinct neuronal subtypes. This demonstrates that scMethCraft achieves balanced batch integration without overcorrection, thereby retaining biologically meaningful heterogeneity across datasets.

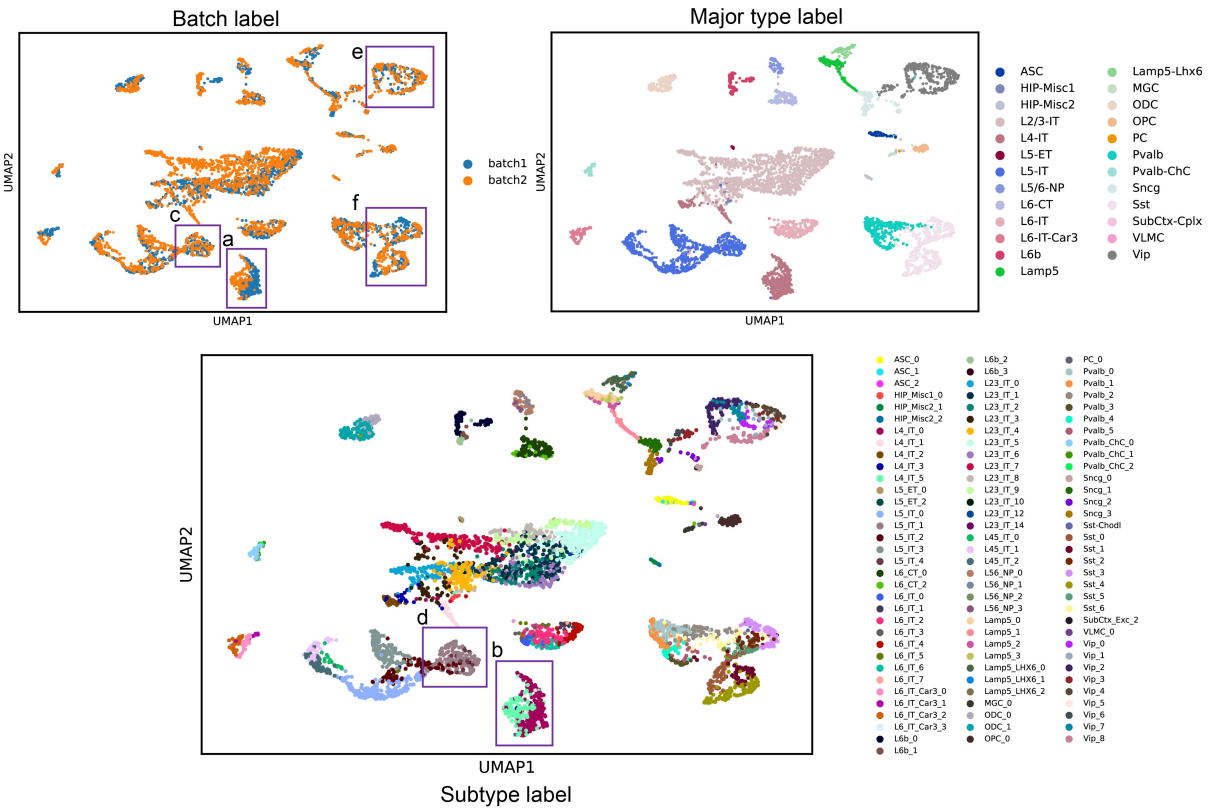


Fig. R10 scMethCraft preserves biological heterogeneity while integrating cell populations across batches. UMAP visualization of the integrated embeddings provided by scMethCraft on the MTG-Pro dataset colored by batch, major type and subtype.

Second, we examined MGE-derived SST cells and CGE-derived VIP cells, both of which exhibited pronounced batch effects before integration. Because interneurons are generally less divergent across cortical regions compared with excitatory neurons, effective removal of batch effects in these populations is particularly important for evaluating batch correction performance. Notably, scMethCraft achieved robust batch integration for both SST (Fig. R10e) and VIP cells (Fig. R10f), resulting in clear coclustering and substantial overlap between cells from different regions in the integrated space.

In summary, we have reanalyzed the batch integration performance of scMethCraft across neuronal populations from different cortical regions, with a particular focus on the balance between effective batch effect removal and the preservation of biologically meaningful heterogeneity. Using interneuron populations as representative examples, we further validated the batch correction capability of scMethCraft under conditions where minimal biological divergence is expected across regions, thereby avoiding the introduction of biological differences. We again thank the reviewer for their insightful comments and expertise in neuroscience, which greatly contributed to improving the clarity and biological rigor of the manuscript. The corresponding revisions and discussions have been incorporated into the main text, Fig. 3, and Supplementary Figs. S8-S9 of the revised manuscript.

COMMENT 6

6. The statistical method for DMR calling (rank sum test) is not rigorous. It is known to be overly lenient because it treats correlated observations (single cells from the same sample) as independent observations. Most DMR calling method uses count-based statistics, such as the beta binomial test. The authors may be constrained by a lack of biological replicates. Nevertheless, a more rigorous DMR calling approach should be used.

RESPONSE TO COMMENT 6

We thank the reviewer for raising these important concerns regarding the statistical rigor of DMR identification and the assumptions underlying differential testing in single-cell DNA methylation data. We agree that treating single cells as independent observations can lead to overly lenient significance estimates due to correlation among cells from the same sample. Below, we clarify the rationale behind our choice of statistical test and describe the additional corrections we have implemented to improve rigor and control false discoveries.

First, negative binomial-based methods (e.g., DESeq2[17]) are widely used for differential analysis of count-based data, such as scRNA-seq or bulk sequencing data. However, single-cell DNA methylation data are fundamentally different in structure: after regional aggregation, scDNAm data are naturally represented as continuous methylation ratios bounded between 0 and 1, rather than as discrete counts. As a result, these data do not satisfy the distributional assumptions required by negative binomial models. Indeed, in practice, directly using scDNAm matrices as input to DESeq2 results in errors such as "some values in assay are not integers", reflecting this incompatibility.

In contrast, the Wilcoxon rank-sum test is a non-parametric method that does not assume a specific underlying distribution and is therefore well-suited for continuous-valued measurements. It has been extensively adopted for differential feature identification in single-cell analyses, including in widely used toolkits such as Seurat[5] (default test for FindMarkers), Scanpy[6] (sc.tl.rank_genes_groups), and EpiScanpy[15] (epi.tl.rank_features), particularly when distributional assumptions are difficult to justify. Previous single-cell studies have also employed Wilcoxon-based tests for identifying differentially methylated or accessible regions in sparse, noisy single-cell epigenomic data[18, 19]. Taken together, we believe that the Wilcoxon rank-sum test represents a reasonable and commonly accepted choice for region-level scDNAm differential analysis under current data constraints.

Second, regarding the concern that treating single cells as independent observations may lead to overly lenient significance estimates. We agree with the reviewer that this assumption can increase the risk of false positives. To explicitly address the reviewer's concern regarding

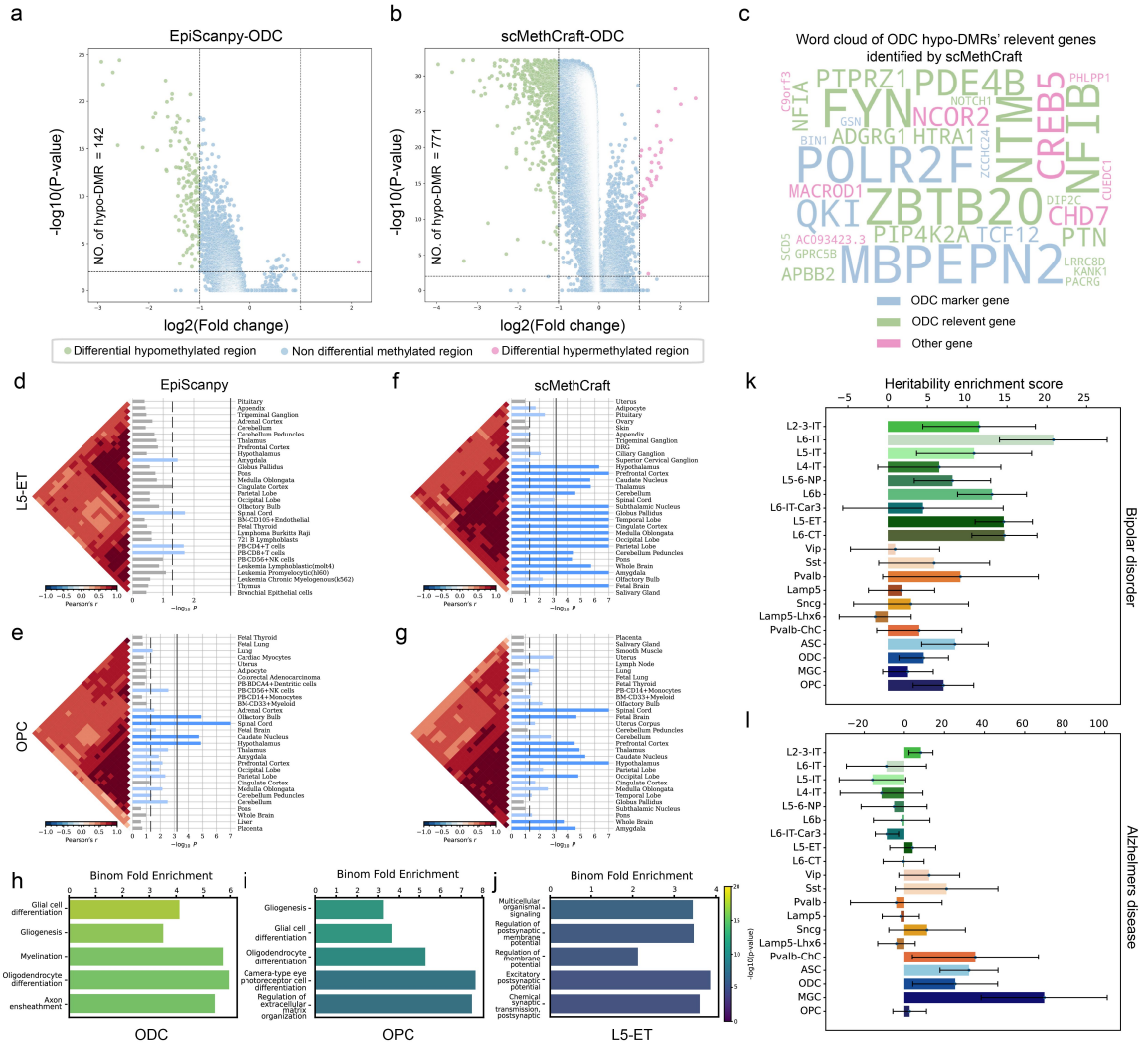


Fig. R11 scMethCraft reveals cell-type-specific biological insights. a-b, Volcano plots of ODC-specific DMRs identified by EpiScanpy and scMethCraft. **c,** Word cloud for 36 genes associated with ODC hypo-DMRs identified by scMethCraft, appearing in at least three instances. Gene font size is positively correlated with the number of associations. **d-g,** Cell-type-specific regions and background regions were subjected to SNPsea analysis to determine the top 30 significantly enriched tissues. Bars represent the empirical P-values calculated by the SNPsea algorithm. The significance of gene enrichment for a specific annotation was assessed using one-sided P-value cutoffs at the 0.05 level, represented by vertical dashed and solid lines. The unadjusted P-value cutoff is indicated by the dashed line, while the Bonferroni-corrected P-value cutoff is shown by the solid line. The heatmaps display the Pearson correlation coefficients, which measure the similarity between expression profiles. Expression profiles were ordered using hierarchical clustering with the unweighted pair-group method with arithmetic means (UPGMA). **h-j,** Top five enriched GO terms associated with major type-specific hypo-DMRs identified by scMethCraft, along with binomial test values (two-sided, not adjusted for multiple comparisons) and enrichment, as reported by GREAT. **k-l,** Heritability enrichment estimated by stratified LDSC of the SNPs within the major type-specific hypo-DMRs identified by scMethCraft, for two brain-related diseases. The error bars and centers of error bars represent the standard errors and average values of 200 equally sized jackknife blocks of adjacent SNPs about the estimates of enrichment.

inflated significance due to non-independence among single-cell observations, we applied a strict multiple testing correction to all DMR analyses. Specifically, we adopted the Bonferroni correction to control the family-wise error rate. As a highly conservative procedure, Bonferroni correction substantially reduces the risk of false positives and helps mitigate inflated significance arising from correlated observations. In the revised manuscript, Bonferroni-adjusted P-values are used as the default criterion for DMR identification in scMethCraft, and all downstream analyses have been re-evaluated accordingly (Fig. R11).

Importantly, re-running the DMR analyses using Bonferroni correction did not alter the main biological conclusions of the study, indicating that our findings are robust and not driven by overly permissive statistical testing (Fig. 5 and Supplementary Fig. S17).

REVIEWER #2 (REMARKS TO THE AUTHOR)

The authors present scMethCraft, a computational method designed for single cell methylation analysis that accounts for the missingness patterns intrinsic to this modality and the characteristics of this complicated data type. Their method offers a latent representation of cells that supports downstream analysis and appears robust to batch effects. Overall, this is an interesting and potentially valuable contribution to the single-cell community.

RESPONSE

We express our sincere gratitude to the reviewer for the affirmation of our work and the enthusiastic assessment, as well as constructive comments, which helped us to enhance the manuscript. We have carefully considered and answered every question raised by the reviewer and made revisions in the corresponding section of the revised manuscript.

COMMENT 1

While the algorithm reports gains against algorithms benchmarked against, the benchmarking is incomplete. The imputation capabilities of the network are useful, coupled with identification of DRM detection. This method needs to demonstrate its performance against other published methods and justify the sophisticated architecture.

This should include a comparison of modelling assumptions, scalability and outputs relative to scMethCraft. In particular, MethSCAn should be included in the comparisons (it feels like a natural comparison) and, given the apparent inspiration, scBasset. The latter is due to its influence on the genomic sequence module within scMethCraft.

Also [20]as another tool: https://doi.org/10.1186/s13059-019-1665-8

Kremer, L.P.M., Braun, M.M., Ovchinnikova, S. et al. Analyzing single-cell bisulfite sequencing data with MethSCAn. Nat Methods 21, 1616–1623 (2024). https://doi.org/10.1038/s41592-024-02347-x

Kapourani, CA., Argelaguet, R., Sanguinetti, G. et al. scMET: Bayesian modeling of DNA methylation heterogeneity at single-cell resolution. Genome Biol 22, 114 (2021). https://doi.org/10.1186/s13059-021-02329-8

Kapourani, CA., Sanguinetti, G. Melissa: Bayesian clustering and imputation of single-cell methylomes. Genome Biol 20, 61 (2019). https://doi.org/10.1186/s13059-019-1665-8

RESPONSE TO COMMENT 1

We thank the reviewer for these thoughtful suggestions regarding the completeness of the benchmarking and for highlighting several relevant methods in the field. In response, we have

substantially expanded the comparative analysis and clarified the scope and positioning of scMethCraft relative to existing approaches.

scBasset: We agree that scBasset[21] is relevant to discuss due to its conceptual influence on the genomic sequence module in scMethCraft. In the initial manuscript preparation, we evaluated scBasset on scDNAm data. However, scBasset produced poor results (Fig. R12). This behavior is not due to incorrect implementation, as we verified its strong performance on scATAC-seq data in line with prior work[8]. Instead, the poor performance on scDNAm data likely arises from fundamental differences in data structure: scBasset does not account for explicit missing values (which are pervasive in scDNAm matrices), nor for fixed-length genomic bins (scATAC-seq peaks are typically shorter and variable in length). While the one-hot sequence module in scMethCraft is inspired by scBasset, it has been substantially re-engineered to accommodate scDNAm-specific properties, including binning strategy, loss function design, and training procedure, resulting in markedly improved performance on scDNAm data. Thus, rather than treating scBasset as a direct baseline, we discuss its relationship to scMethCraft through a dedicated ablation analysis.

MethSCAn: MethSCAn[3] is an effective and well-designed method for scDNAm data analysis. Its primary contribution lies in identifying informative, highly variable methylation regions from whole-genome single-cell bisulfite sequencing data, thereby constructing a more suitable scDNAm feature matrix. In contrast, scMethCraft is designed to operate on a pre-defined, fixed-length cell-by-region methylation matrix and focuses on downstream tasks such

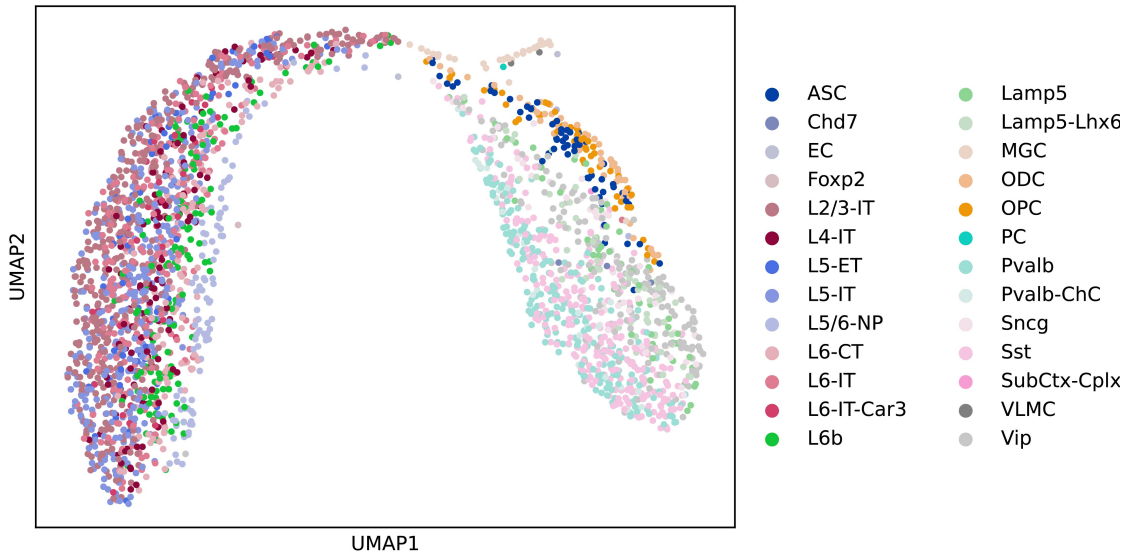


Fig. R12 UMAP visualization of the cell embedding created by scBasset on the ACC dataset.

as dimensionality reduction, imputation, and batch integration. As a result, the two methods differ in both input representation and core analytical objectives, and they do not fully overlap in terms of task definition. This mismatch was the main reason MethSCAn was not included in our initial comparison in the original manuscript. While MethSCAn was originally developed to identify informative methylation regions and construct feature matrices from whole-genome single-cell bisulfite sequencing data, it also provides an impute-PCA-based strategy for handling missing values and dimensionality reduction. Following the reviewer's suggestion, we have now included MethSCAn as a baseline for comparison in the revised manuscript. In addition, as discussed in Response #4 to Reviewer #1, we have also added comparisons to the bin-based pipeline following Liu et al.[2], which is widely adopted in existing scDNAm studies. The expanded benchmarking results show that scMethCraft consistently achieves superior performance across multiple datasets in both dimensionality reduction and imputation tasks, including comparisons with the newly added baseline methods (Fig R13). We have revised the manuscript accordingly to incorporate these additional comparisons and to better highlight these strengths.

Melissa and scMET: Melissa[20] and scMET[22] are pioneering methods for scDNAm data analysis and have made important conceptual contributions to the field. However, with the rapid development of scDNAm sequencing technologies, their applicability to large-scale datasets has become limited. In the datasets used in our manuscript, which often contain thousands of cells, we were unable to successfully run Melissa or scMET within practical memory (1 TB) or time (168 hours) constraints. This limitation has also been explicitly reported in the MethSCAn study[3]. In addition, these methods primarily operate at single-CpG resolution and aim to infer methylation states at individual sites. As dataset sizes increase, region-based binning approaches have become a more practical and commonly adopted strategy for scDNAm data analysis. scMethCraft is specifically designed for binned scDNAm matrices and addresses multiple analytical task. Consequently, differences in input representation, modeling objectives, and scalability prevent a fair and direct comparison. Notably, the peer-review discussion of MethSCAn provides a detailed analysis of why several earlier scDNAm methods are no longer suitable as baselines for large-scale studies (https://static-content.springer.com/esm/art%3A10.1038%2Fs41592-024-02347-x/MediaObjects/41592_2024_2347_MOESM2_ESM.pdf). We do cite Melissa and scMET now as an early prior work in the main text.



Fig. R13 scMethCraft consistently achieves superior performance in both dimensionality reduction and imputation tasks. a, Cell clustering performance assessed by AMI scores with respect to major cell types across different dimensionality reduction methods. Colors represent min-max normalized AMI scores within each dataset, where lighter blue indicates the lowest-performing method and darker blue indicates the highest-performing method. **b,** Average AMI scores with respect to major cell types, based on embeddings provided by different methods across 11 independent datasets. **c,** Cell clustering performance of embeddings generated by different methods was evaluated using five metrics across 11 datasets. **d,** One-sided paired Wilcoxon signed-rank tests were used to compare performance between methods. Each cell shows the P-value for the hypothesis that the method in the column outperforms the method in the row in terms of AMI or ARI across the 11 datasets. **e-j,** UMAP visualization of cell embeddings provided by different methods on the ACC and M1C datasets. Main panels are colored by major cell types, while inset panels are colored by cell clusters. The AMI values shown were computed only using the cell types highlighted by the boxes in the main panels. **k,** Spearman correlation between the average methylation level of each single cell and its corresponding meta-cell, computed for the raw matrix and for enhanced matrices generated by different methods across cell types in the ACC dataset. scMethCraft (NA) denotes correlations computed using only regions originally annotated as NA in the raw matrix. **l,** Spearman correlation between enhanced datasets and their corresponding meta-cells across 11 datasets processed by different methods.

In summary, we have incorporated MethSCAn and the bin-based pipeline following Liu et al. as additional baselines in response to the reviewer's suggestions (Figs. 2, 4, Supplementary Figs. S1-S5, Supplementary Notes S2, S6). We explicitly discuss the relationship between scMethCraft and scBasset in the ablation study, clarifying both the inspiration and the key engineering differences. Finally, due to scalability constraints and fundamentally different analytical objectives, Melissa and scMET are not directly comparable to scMethCraft on the datasets considered. Issues related to scalability are further discussed in Response #7 to Reviewer #2. We hope that the reviewer will find these revisions to improve the transparency, fairness, and completeness of the comparative evaluation.

COMMENT 2

In Figure 2, the performance of scMethCraft is satisfactory but the gains over competing methods, especially EpiScanpy is not dramatic. Considering that scMethCraft has a complex deep learning architecture, I would be curious to find out the run times of scMethCraft and competing methods. Furthermore, it is not clear to me, from Fig2EFG panels, how well OPCs, ODCs, and MGCs cells are better separated by scMethCraft. Either the color scale needs to be adjusted or a separate supplemental Figure can display this.

RESPONSE TO COMMENT 2

We thank the reviewer for the constructive comments and for acknowledging the performance of scMethCraft. Regarding the concern about computational efficiency and runtime, we have provided a detailed response in Response to Comment #7 of Reviewer #2, where we systematically report and compare the running times and memory usage of scMethCraft and competing methods.

With respect to the magnitude of performance gains, we note that overall clustering improvements can be partially masked by abundant and easily separable cell types. As a result, global clustering metrics may underestimate improvements in more challenging scenarios. When focusing on highly confusable cell populations, scMethCraft demonstrates substantially stronger advantages. For example, in the ACC dataset (Fig. R13e-g), scMethCraft achieves a 45% AMI improvement over EpiScanpy for five particularly difficult-to-distinguish cell types. Similarly, in the M1C dataset (Fig. R13h-j), scMethCraft attains a 35% AMI improvement for four closely related cell types. These results highlight scMethCraft's strength in resolving subtle but biologically meaningful differences, demonstrating its advantage in challenging and biologically relevant settings.

We also apologize for any confusion caused by Fig. 2e-g in the original presentation. To address this issue, we have made several revisions. First, we have refined the description in the main text to more clearly articulate how scMethCraft improves the separation of closely related cell types. Second, we have updated Fig. 2 (now Fig. R13) with a clearer and more informative scheme. In addition, we now provide an additional supplementary figure (Supplementary Fig. S3) in which multiple coloring strategies are applied to the embeddings, specifically to better highlight the separation among OPCs, ODCs, and MGCs. These revisions are intended to more clearly demonstrate the potential of scMethCraft in capturing heterogeneity among closely related cell populations.

We appreciate the reviewer's suggestion, which has helped us improve both the clarity and completeness of the manuscript.

COMMENT 3

scMethCraft's performance on batch correction is impressive, and a clear improvement.

RESPONSE TO COMMENT 3

We thank the reviewer for this positive assessment and are pleased that the improvement of scMethCraft in batch correction is recognized.

COMMENT 4

The manuscript states:

"The input of scMethCraft is a single-cell DNA methylation cell-by-region matrix, in which each read count represents the average methylation level across CpG sites within a genomic tile for each individual cell."

This could be clearer.

- **Is the model input percentage methylation per region per cell (#sites methylated / #sites measured) or the read count as written?**
- **If it is the latter, does the count represent the number of methylated reads or the total measured reads?**
- **Is there a minimum coverage threshold / number of sites measured below which data are excluded for a bin within a cell? (in contrast to across cell filtering which I interpret the 80% missingness threshold to refer to). Was model sensitivity to this choice explored?**

RESPONSE TO COMMENT 4

We thank the reviewer for pointing out this lack of clarity and apologize for the ambiguous use of the term “read count” in the original manuscript. Specifically, the input to scMethCraft is the percentage methylation level per genomic region per cell, computed as the fraction of methylated CpG sites over the total number of CpG sites measured within that region for a given cell. Thus, the model input represents a continuous value between 0 and 1, rather than raw read counts. We have revised the manuscript accordingly to clarify this definition.

With respect to filtering, regions with no CpG sites covered in a given cell are treated as missing (NA) values. We did not impose an explicit minimum coverage threshold at the individual cell-region level, in order to retain as much information as possible given the inherent sparsity of single-cell DNA methylation data. Instead, we applied region-wise filtering by removing genomic regions that are missing in more than 80% of cells, such that only regions with sufficient coverage across the dataset are included as model input. This strategy is conceptually analogous to filtering lowly expressed genes in scRNA-seq analyses[6].

We next explain the rationale for choosing the 80% missingness threshold. Clearly, if a genomic region has no methylation measurements in any cell, including such a region in model training would not contribute any informative signal for parameter updates. By examining the histogram of the region-wise distribution of missing (NA) rates in the ACC dataset, we

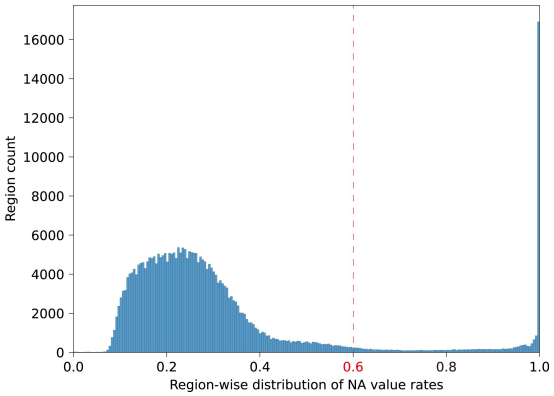


Fig. R14 The region-wise proportion of explicit NA values of the ACC dataset.

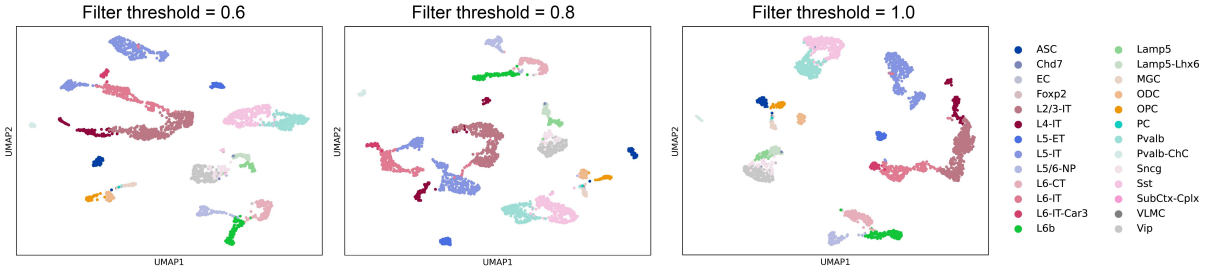


Fig. R15 UMAP visualization of the embedding created by scMethCraft with varying filter thresholds.

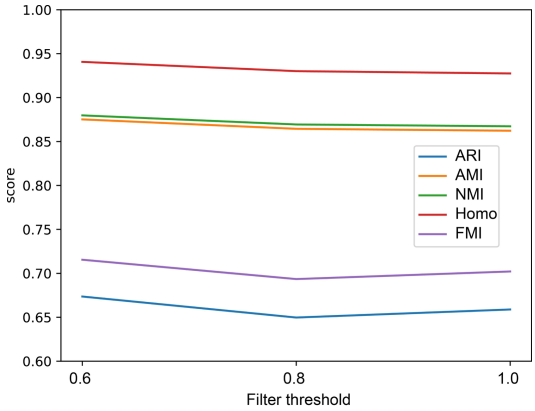


Fig. R16 Cell clustering performance of various filter thresholds.

observed that the majority of regions exhibit moderate missingness, with NA rates concentrated around 0.1-0.4 (Fig. R14). In contrast, a second prominent peak appears near 1.0, indicating that a non-negligible number of regions have extremely sparse or no methylation coverage across cells. Excluding these regions from the model input is therefore a natural and data-driven choice. Moreover, the density of regions with NA rates between 0.6 and 1.0 is very low, suggesting that the precise choice of a cutoff within this range should not critically affect model behavior. Based on this observation, scMethCraft is unlikely to be sensitive to the exact threshold chosen around 0.8. To empirically validate this expectation, we further evaluated the sensitivity of scMethCraft to alternative coverage thresholds by filtering out regions that are missing in all cells, in more than 80% of cells, or in more than 60% of cells, respectively. The resulting analyses were consistent with our expectation, showing that scMethCraft is not sensitive to these alternative filtering strategies (Figs. R15, R16).

The coverage filtering threshold is implemented as a configurable parameter in scMethCraft. We have added a detailed discussion of the preprocessing steps to the revised Supplementary Note S4.

COMMENT 5

Lines 116-118 state that a 50-dimensional embedding is used for each region, how was this dimensionality chosen, and was model performance explored as a function of this choice?

RESPONSE TO COMMENT 5

We appreciate the reviewer for this insightful comment. The choice of 50 represents a relatively moderate setting. For example, Scanpy[6] maps cells into a 50-dimensional space by default using PCA. We empirically selected 50 as the dimensionality of the feature vector, which provides a reasonable balance between representational capacity and computational efficiency.

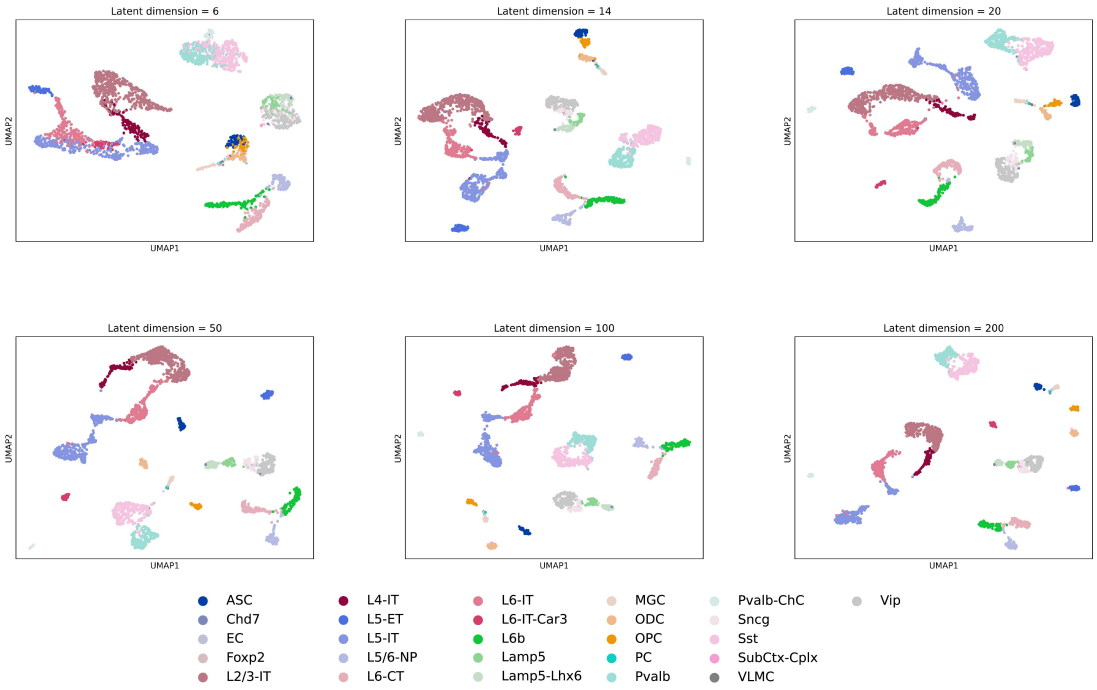


Fig. R17 UMAP visualization of the embedding created by scMethCraft with varying latent dimensions.

Motivated by the reviewer's suggestion, we further examined the impact of different feature vector dimensionalities on model performance. On the ACC dataset, we evaluated scMethCraft with embedding dimensions ranging from 6, 14, 20, 50, 100, to 200, followed by visualization and cell clustering based on the resulting cell embeddings (Figs. R17, R18). The results indicate that when the dimensionality is low, the model is unable to sufficiently capture DNA sequence information. However, once the dimensionality reaches approximately 20, the embeddings are generally sufficient to encode the relevant sequence information, and further increasing the dimensionality does not lead to a clear improvement in model performance.

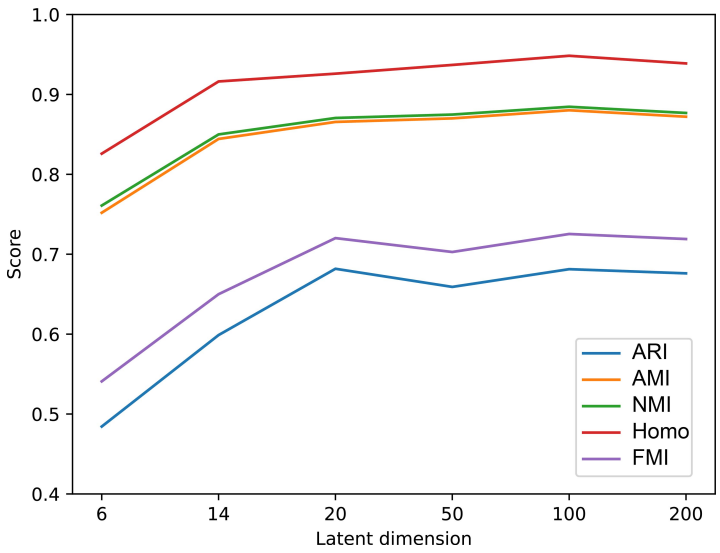


Fig. R18 Clustering performance of various latent dimensions.

Following the reviewer's suggestion, we have therefore treated the embedding dimensionality as an optional hyperparameter in scMethCraft and added relevant discussion in Supplementary Note S4 to accommodate the needs of different users.

COMMENT 6

The phrase "Upon completion of training, the similarity matrix S undergoes normalization as Snorm" should specify the normalization used (e.g., z-score normalization, min-max scaling) to aid reproducibility and clarity.

RESPONSE TO COMMENT 6

Thank you for the valuable comment that contributes to refining the details in our manuscript. In fact, the normalization strategy used in scMethCraft was already described in the Methods section of the original manuscript under "Similarity matrix normalization and batch integration implementation". We apologize for not explicitly referencing this description in the main text, which may have confused readers. In the revised manuscript, we have added the corresponding reference to the Methods section at the appropriate location. Here, we restate the normalization procedure and clarify its advantages.

We adopt a symmetric normalization scheme[23]. The symmetrically normalized similarity matrix $\mathbf{S}' \in R^{n\times n}$ is calculated as:

$$\mathbf{S}' = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}.$$

Next, we define the adjusted similarity matrix $\mathbf{S_{norm}} \in R^{n\times n}$ as:

$$\mathbf{S_{norm}} = \mathbf{S}' + \alpha\mathbf{E},$$

where α is a hyperparameter, and $\mathbf{E}$ denotes the identity matrix. The resulting matrix $\mathbf{S_{norm}}$ is referred to as the normalized similarity matrix.

This normalization strategy strictly preserves the non-negativity of the signal, ensuring that similarity values consistently represent positive biological associations and avoiding the loss of physical interpretability that can arise from z-score normalization due to the introduction of artificial negative values. In addition, it precisely regulates the relative contributions of endogenous information from each cell itself and exogenous information from similar cells, thereby enabling a high-fidelity representation of intercellular heterogeneity.

COMMENT 7

The model architecture suggests potentially high computational costs, particularly since the cell-cell similarity matrix becomes O(N2).

- **Please provide details on the computational resources used (number and type of GPUs, runtime, memory usage)**
- **Please include these quantities as a function of the number of cells and genomic bins used**
- **Comparing these costs to competitors would further strengthen this work and encourage users**

RESPONSE TO COMMENT 7

We thank the reviewer for this constructive comment, which highlights the importance of computational efficiency and scalability for practical applications. We fully agree with the reviewer that runtime and memory usage are key factors affecting the usability of computational methods. Below, we respond to the reviewer's three concerns in turn, focusing on computational resource usage, theoretical complexity, and comparison with existing methods.

First, we report the practical computational resource requirements of scMethCraft. All experiments in this study were conducted using a single NVIDIA RTX 3090 GPU (24 GB memory). The server configuration was as follows: Intel(R) Xeon(R) Platinum 8375C CPU @ 2.90 GHz (128 cores), Ubuntu 22.04.3 LTS, and 1 TB system memory. We ran scMethCraft with a batch size of 128 on the Mop (296 cells), ACC (2,698 cells), and Mixed (~16,000 cells) datasets, and systematically recorded GPU memory usage and runtime (Fig. R19). As the number of cells increased, GPU memory usage grew from approximately 13 GB to approximately 20 GB (Fig. R19a), remaining well within the capacity of the RTX 3090. Importantly, as a deep learning-based method, scMethCraft allows GPU memory usage to be substantially reduced by decreasing the batch size, enabling its use on GPUs with more limited memory. Across all three datasets, scMethCraft completed the full analysis pipeline within

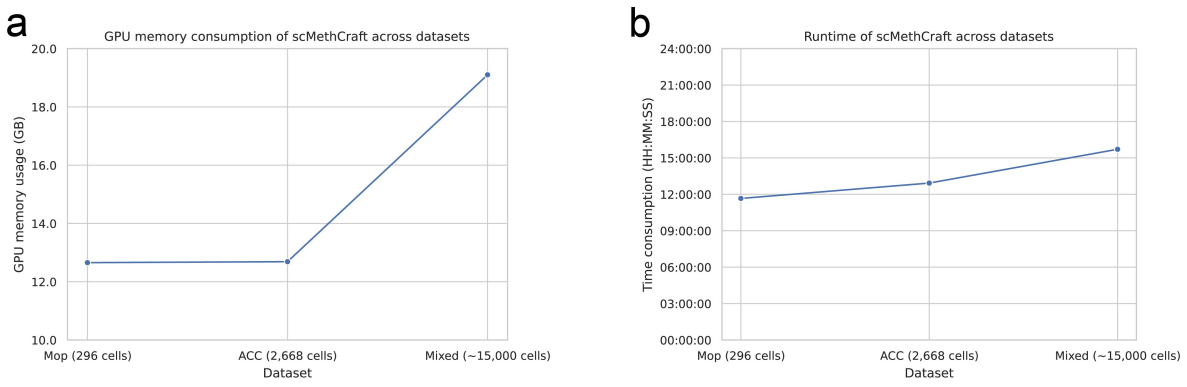


Fig. R19 GPU memory consumption and runtime of scMethCraft across datasets of increasing scale. a, Peak GPU memory usage during model training. **b,** Runtime of model training.

approximately half a day (Fig. R19b). Therefore, for scDNAm datasets of commonly encountered sizes, the runtime and memory consumption of scMethCraft are well within a range acceptable to biologists.

Second, we analyze the theoretical computational complexity of scMethCraft and its scaling behavior with respect to the number of cells and genomic bins. scMethCraft mainly consists of two computational modules: the sequence feature extraction module and the similarity weighting module. The former depends on the number and size of genomic bins and, in each training epoch, effectively traverses the entire reference genome. Because the total length of the genome is finite, using finer genomic binning increases the number of bins while proportionally decreasing their lengths; as a result, the overall computational burden of this module can be regarded as approximately constant. With a batch size of 128, training this module requires approximately 12 GB of GPU memory. The similarity weighting module explicitly models similarities among cells based on their methylation patterns. As noted by the reviewer, in the original implementation, this module has a computational complexity of $O(n^2)$. However, for scDNAm datasets of currently common sizes (on the order of several thousand cells), this computational cost is manageable in practice and sufficient to cover the vast majority of published and actively used datasets. To address the reviewer's concern regarding scalability to larger datasets, we further introduced an extended version based on a low-rank approximation of the similarity matrix in this revision. Specifically, the explicitly stored similarity matrix $\mathbf{S}$ is factorized into the product of two low-dimensional matrices, thereby avoiding the construction and storage of the full $n \times n$ matrix. Assuming there is a total of n cells, we denote the similarity matrix obtained through iterative training as $\mathbf{S} \in R^{n \times n}$. Let $\mathbf{P} \in R^{n \times k}$ denote the trainable parameters of the similarity weighting module, where k is the hyperparameter (default 256). To enforce the non-negativity, symmetry, and a zero-valued main diagonal property of the similarity matrix $\mathbf{S}$, we define it as:

$$\mathbf{S} = ((|\mathbf{P}||\mathbf{P}^{\mathbf{T}}|) - \mathrm{diag}(\mathbf{P}\mathbf{P}^{\mathbf{T}}))$$

By factorizing the similarity matrix, all operations involving the similarity matrix are implemented through its factorized form, and the full $n \times n$ matrix S is never explicitly constructed or stored in memory. Under this approximation, the computational complexity of the similarity weighting module can be reduced from $O(n^2)$ to approximately $O(n)$ (Fig. R20a,b). We emphasize that this modification introduces an approximation only at the computational level and does not alter the biological assumptions or modeling objectives of the method. We validated this extended version and found that it continues to effectively capture

intercellular similarity relationships while maintaining performance trends consistent with the original implementation (Fig. R20c-f).

Finally, we discuss the runtime and memory usage of scMethCraft in comparison with competing methods. We acknowledge that, compared with some methods based on linear dimensionality reduction or lightweight statistical models, scMethCraft incurs a higher runtime. This difference primarily arises from its deep learning architecture, which requires sufficient training even on relatively small datasets.

Moreover, we believe the computational cost of scMethCraft is reasonable and acceptable. First, in practice, the observed runtimes fall well within time scales that are acceptable for routine biological data analysis applied to contemporary scDNAm datasets containing thousands to tens of thousands of cells, a setting that increasingly reflects current experimental scales. Second, scMethCraft demonstrates consistent performance improvements across multiple datasets and evaluation metrics, while substantially enhancing biological interpretability. Third, scMethCraft integrates multiple key analytical tasks, including dimensionality reduction, imputation, batch correction, and differential region identification, within a single framework, allowing users to perform multiple downstream analyses with a single model training, thereby simplifying the overall analysis workflow. Fourth, after architectural optimization, its overall computational complexity scales approximately linearly with the number of cells, preventing exponential growth in computational cost for larger datasets. In contrast, many existing scDNAm data analysis methods were originally developed for smaller datasets and are therefore primarily applicable in limited-scale scenarios. For example, methods such as Camelia[24] and DeepCpG[25] encounter practical memory limitations as dataset sizes increase, while Epiclomal[26] and MambaCpG[27] typically require substantially longer runtimes and are primarily suited to datasets containing on the order of hundreds of cells.

In summary, although scMethCraft employs a more expressive model architecture, its computational cost remains controllable and scalable in practical applications. By introducing low-rank approximation-based extensions, we further improve its applicability to large-scale datasets. In the revised manuscript, we have also added a detailed discussion of the computational resources required for model training and the corresponding fast implementation (Methods, Supplementary Figs. S18, S20). We hope the reviewer will find that these improvements effectively address the concerns regarding computational efficiency and scalability, and further enhance the practical utility of scMethCraft for biological research.

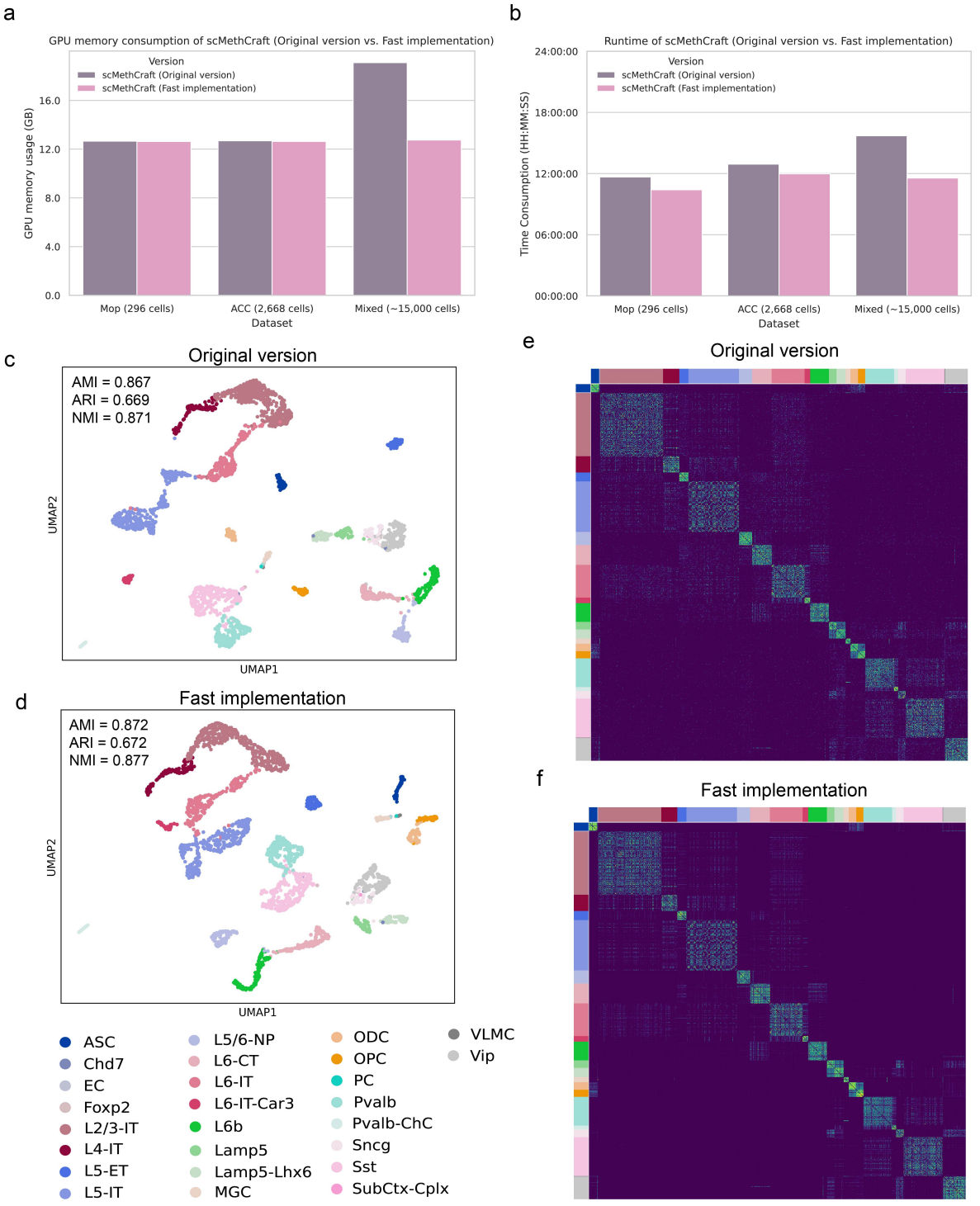


Fig. R20 Performance comparison of scMethCraft between the original version and the fast implementation. a, GPU memory consumption, and **b,** runtime of scMethCraft across datasets of increasing scale, contrasting the original version with the fast implementation. **c-d,** UMAP visualization and clustering performance of cell embeddings produced by **(c)** the original version and **(d)** the fast implementation. **e-f,** The heatmap of the cell-to-cell similarity matrix created by **(e)** the original version and **(f)** the fast implementation, where cells were sorted according to their major cell types. All similarity values are linearly scaled to the range 0-1.

COMMENT 8

The reported performance differences under clustering metrics (e.g., ARI, AMI) appear modest. However, these are paired comparisons, so relative improvements may be obscured by summary statistics (as the significance of the test statistic that accounts for this shows). Consider presenting visualizations of the relative changes (i.e., relative disimprovement compared to scMethCraft) for the key metrics.

RESPONSE TO COMMENT 8

We thank the reviewer for this accurate and insightful observation. Following the reviewer's suggestion, we have added visualizations of relative performance changes in the revised manuscript and clarified this point in the main text. First, we replotted Fig. 2, in which colors represent min-max normalized AMI scores within each dataset, with lighter blue indicating the lowest-performing method and darker blue indicating the highest-performing method (Fig. R13a,b). This visualization highlights the consistent and robust advantage of scMethCraft across datasets.

Second, we note that improvements on difficult or highly confusable cell populations, which often constitute a relatively small fraction of the data, can be diluted by abundant and easily separable cell types in global summary statistics. Thus, we additionally quantified performance gains on selected confusable cell types, thereby more directly revealing scMethCraft's ability to capture subtle but meaningful heterogeneity (Fig. R13e-j).

Finally, we emphasize the importance of the paired statistical tests. Although absolute differences in some clustering metrics may appear small, the paired Wilcoxon signed-rank tests consistently indicate statistically significant and positive relative gains achieved by scMethCraft across datasets (Fig. R13d). Together, these analyses provide a more faithful representation of scMethCraft's performance improvements under a paired comparison setting.

COMMENT 9

The manuscript briefly notes the intra-batch weighting parameter, β, is set to 0.2.

- **Was sensitivity to this choice explored by the authors? I would like to see this.**
- **Are there dataset characteristics that might inform choices of the magnitude? E.g., batch sizes, coverage.**

I think a short discussion for estimating β from data, or future directions such as making β adaptive, would enhance the paper.

RESPONSE TO COMMENT 9

We thank the reviewer for raising this important question regarding the intra-batch weighting parameter β. We fully agree that a more thorough discussion of this parameter can substantially strengthen the understanding and robustness of scMethCraft for batch integration. To this end, we conducted a comprehensive set of experiments and analyses from three complementary perspectives: (i) a sensitivity analysis across multiple datasets, (ii) a theoretically motivated adaptive parameter selection strategy, and (iii) practical considerations for parameter choice in real-world applications.

First, we systematically evaluated the effect of β on batch integration performance across four datasets which were used in the original manuscript, by varying β over the range [0, 1]. For the MTG-Pro dataset, we visualized the integration results using UMAP visualization under different values of β (Fig. R21). These visualizations show that when β is close to 0, the two batches are poorly integrated. As β increases, batch mixing gradually improves and

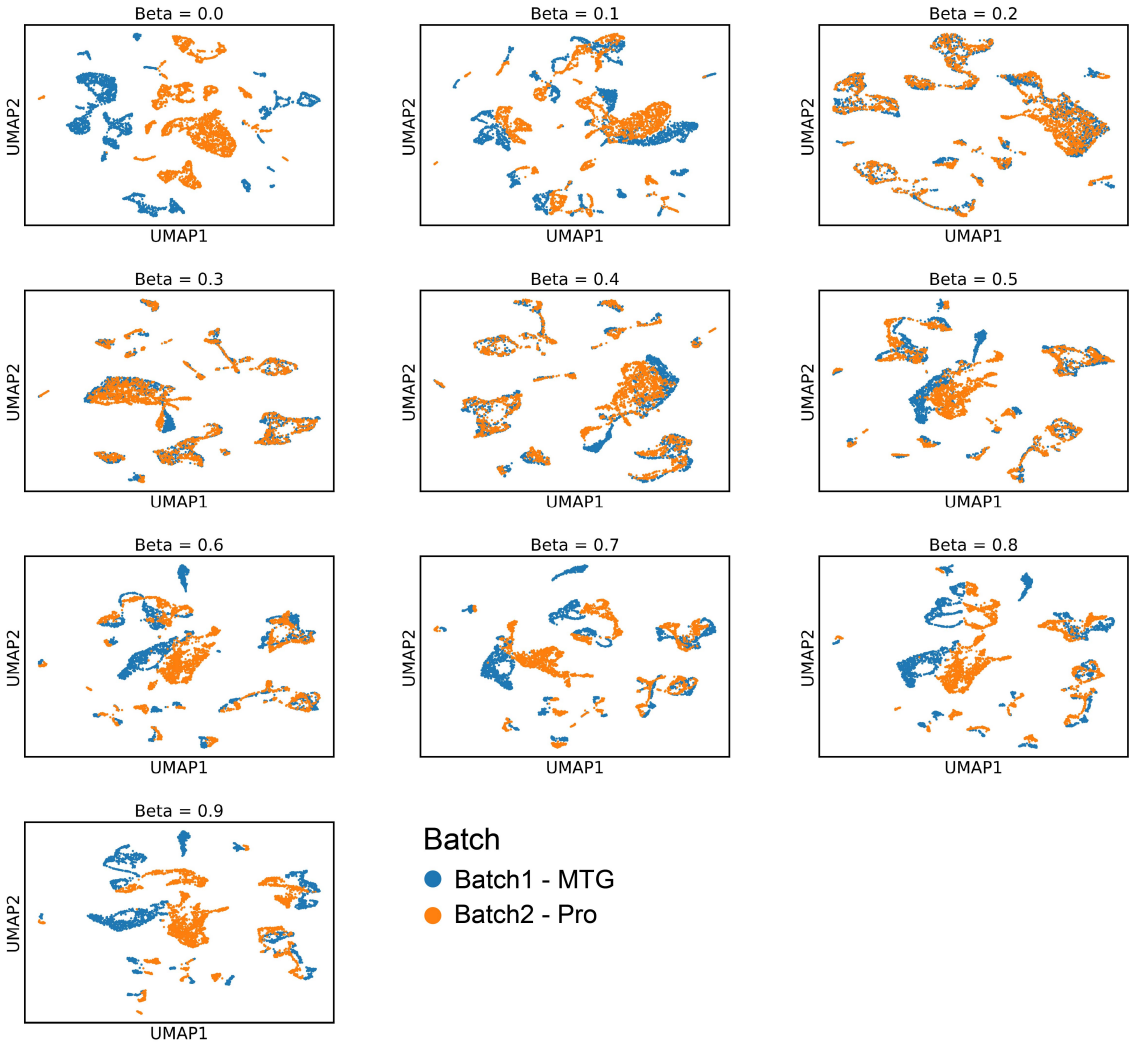


Fig. R21 UMAP visualization under different values of β.

reaches an optimal level around $\beta \approx 0.2$. In addition to visualization, we evaluated integration quality using the graph iLISI metric, which does not rely on cell-type annotations and therefore provides a more direct assessment of batch integration (Fig. R22). The results consistently show that performance increases from near zero at small values of β, reaches a plateau or optimum around $\beta \approx 0.2$, and then gradually decreases as β approaches 1. This trend is consistently observed across all four datasets. Based on these observations, we empirically selected β with 0.2 as the default value in our original experiments.

Second, the design of the batch integration strategy in scMethCraft is inspired by scCASE[8], which aims to balance intra-batch similarity and inter-batch similarity for the same cell types. Conceptually, the goal is that, after integration, similarities between cells of the same type should be comparable within and across batches. Motivated by this principle, we developed an adaptive strategy for selecting β based on batch-level statistics, including the average intra-batch similarity, average inter-batch similarity, and the number of cells in each batch:

We define a binary indicator matrix:

$$\mathbf{M} \in \{0,1\}^{N \times N},$$

which encodes whether a pair of cells belongs to the same batch:

$$\mathbf{M} = \begin{cases} 1, & if\ B_i = B_j, \\ 0, & if\ B_i \neq B_j. \end{cases}$$

We then estimate an adaptive scaling factor β, which reflects the relative weighting between inter-batch similarity and intra-batch similarity:

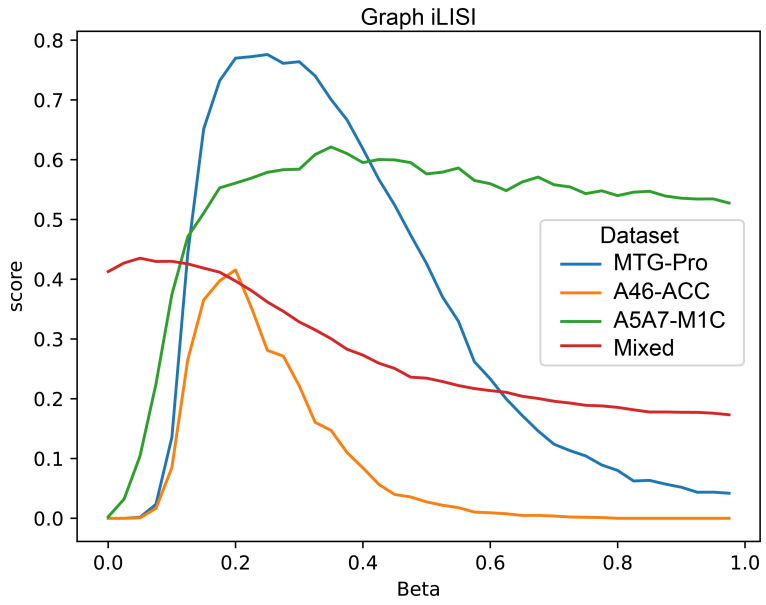


Fig. R22 Graph iLISI scores across different β values on four datasets.

$$\beta = \frac{\sum \mathbf{M}}{\sum(1-\mathbf{M})} \cdot \frac{\text{mean}\left(\mathbf{S_{norm}} \odot (1-\mathbf{M})\right)}{\text{mean}(\mathbf{S_{norm}} \odot \mathbf{M})},$$

where $\mathbf{S_{norm}}$ denotes the input cell-cell similarity matrix, and $\odot$ represents the Hadamard product. The term $\sum \mathbf{M}$ corresponds to the total number of within-batch cell pairs, while $\sum(1-\mathbf{M})$ represents the total number of cross-batch cell pairs.

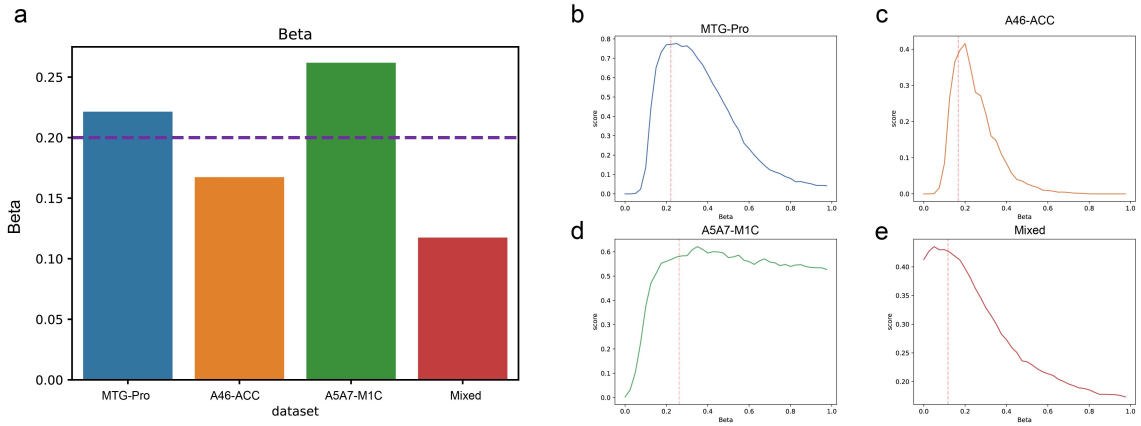


Fig. R23 Adaptive parameter selection strategy β values across four datasets. a, Adaptively selected β values for each dataset. **b-e,** Graph cLISI scores evaluated across different β values for each dataset. The dashed vertical line indicates the adaptively selected β value.

We applied this adaptive parameter selection strategy to the four datasets described above. In all cases, the adaptively selected β values were relatively small and consistently close to 0.2 (Fig. R23a). The Figs. R23b-e and R24 show the resulting Graph cLISI scores and UMAP visualization obtained using the adaptive β values, demonstrating that the adaptive strategy achieves strong batch integration performance, indicating that the adaptive strategy recovers integration performance comparable to the empirically chosen default. Finally, although scMethCraft is sensitive to the full range of β values between 0 and 1, this behavior is expected and reasonable, as the two extremes correspond to completely ignoring intra-batch information and treating intra-batch and inter-batch information equally, respectively, the latter effectively disabling batch integration. Importantly, the mathematical interpretation of β enables principled, data-driven parameter selection, ensuring its effectiveness across different datasets. Moreover, the choice of β is decoupled from model training, which allows users to explore different values after training. In practice, users can test alternative integration strengths within minutes and select an appropriate value based on visual inspection, quantitative metrics, or specific application needs.

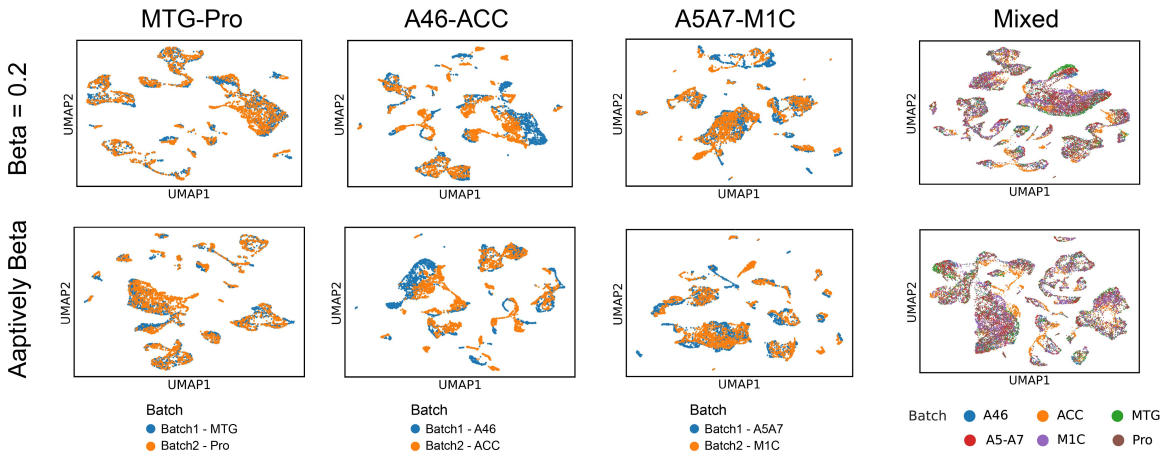


Fig. R24 UMAP visualization of adaptive parameter selection strategy β values across four datasets.

In summary, our results demonstrate that a default choice of $\beta = 0.2$ consistently yields strong batch integration performance across multiple datasets with varying characteristics. Moreover, the parameter β has a clear and interpretable mathematical meaning, which enables adaptive selection strategies based on dataset-specific statistics. Finally, because the choice of β is decoupled from model training, it can be efficiently adjusted post hoc, allowing users to rapidly explore alternative integration strengths and tailor batch correction to different data scenarios. We have added a detailed discussion of parameter selection to the Methods section and Supplementary Note S5, and we present the adaptive strategy as the recommended default option in the tutorial. We thank the reviewer again for this insightful suggestion.

COMMENT 10

The imputation capabilities of scMethCraft looks promising, as cell-to-cell similarity module in the network presumably should aid in imputation. There is one clear challenge with the imputation. It seems the average methylation level's gradient within specific clusters in Fig4 EFGH become much more pronounced after imputation. These gradients are much less pronounced in raw data. Furthermore, the cells that are labeled green (TENs) seemingly get higher in average methylation after imputation (panel H), whereas blue (non-neuronal cells) get lower average methylation after imputation. This can be a by product of imputation, naturally, as cells share more information across other cells. Authors should quantify this effect by looking at average methylation per cell per cell type before and after imputation to ensure imputation does not overemphasize certain patterns in specific cell types.

RESPONSE TO COMMENT 10

We thank the reviewer for this careful and insightful evaluation of the imputation results. We agree that imputation based on cell-cell similarity may, in principle, alter average methylation

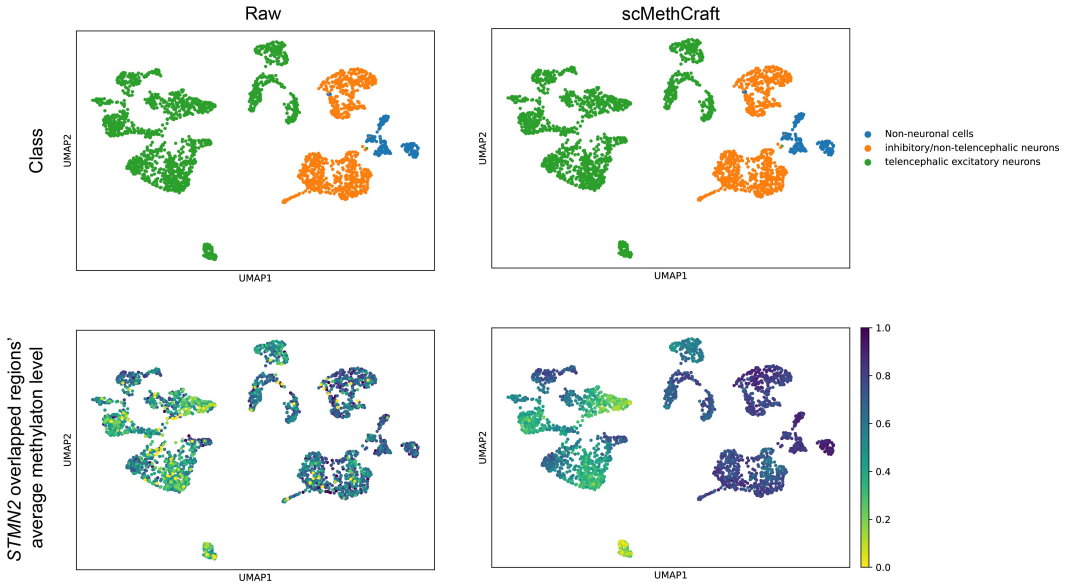


Fig. R25 UMAP visualization of ACC datasets colored by cell class and average methylation levels of *STMN2*-overlapping regions.

levels and potentially exaggerate cell-type-specific patterns, and that this effect should be quantitatively assessed.

First, the pronounced gradients observed in Fig. 4E-H, as well as the changes in average methylation levels for some types before and after imputation, primarily reflect reduced technical noise and strengthened methylation signal recovery, rather than an overemphasis of cell-type-specific patterns. To further support this interpretation, we recolored the UMAP embedding generated by EpiScanpy using average methylation levels of *STMN2*-overlapping regions after scMethCraft imputation (Fig. R25). This visualization shows a high degree of consistency between the pre- and post-imputation patterns, suggesting that the imputation preserves the relative relationships among cells.

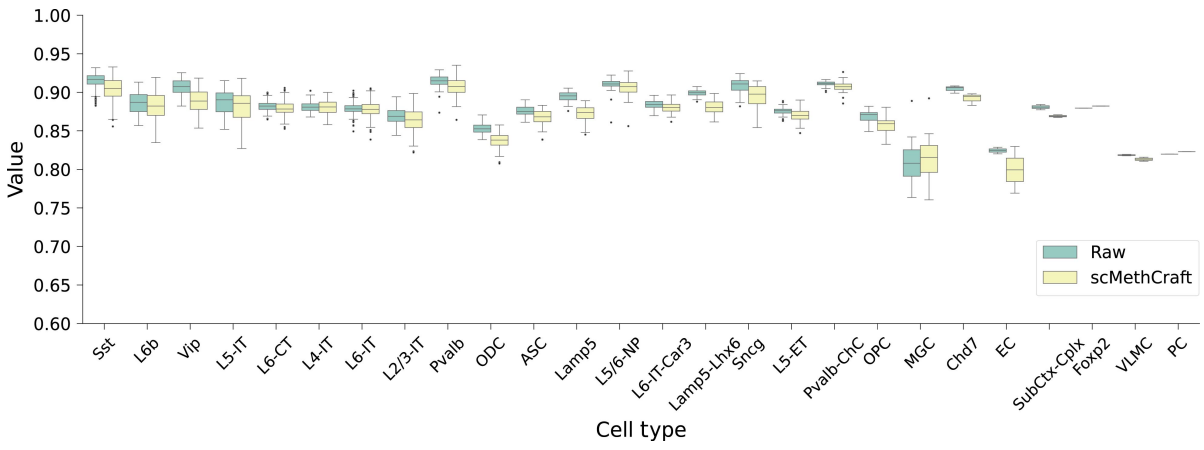


Fig. R26 Distribution of average DNA methylation levels within each cell type.

Second, to directly address the reviewer's concern, we quantitatively evaluated the impact of imputation on average methylation levels at the single-cell level. Specifically, for each cell, we compared the average methylation level before imputation and after scMethCraft imputation. We then summarized these values across different cell types (Fig. R26). While some degree of change in per-cell average methylation is unavoidable after imputation due to information sharing across cells, our results indicate that the overall distribution of average methylation levels within each cell type remains largely consistent before and after scMethCraft imputation, without overemphasizing particular patterns in specific cell types.

Taken together, these results demonstrate that scMethCraft enhances methylation signal without introducing substantial cell-type-specific bias or artificial pattern amplification, as further illustrated in the revised Fig. 4 and Supplementary Fig. S16.

COMMENT 11

Can scMethCraft also identify hypermethylated DMRs? The results focus on hypomethylation.

RESPONSE TO COMMENT 11

We thank the reviewer for this helpful question. scMethCraft is capable of identifying both hypomethylated and hypermethylated DMRs. The underlying statistical framework is symmetric with respect to methylation changes, and hypermethylated DMRs can be identified using the same statistical test by reversing the fold-change criterion (Fig. R27).

In this study, we focused on hypomethylated DMRs because they are more straightforward to interpret in the context of gene regulation and cell-type-specific regulatory programs. As

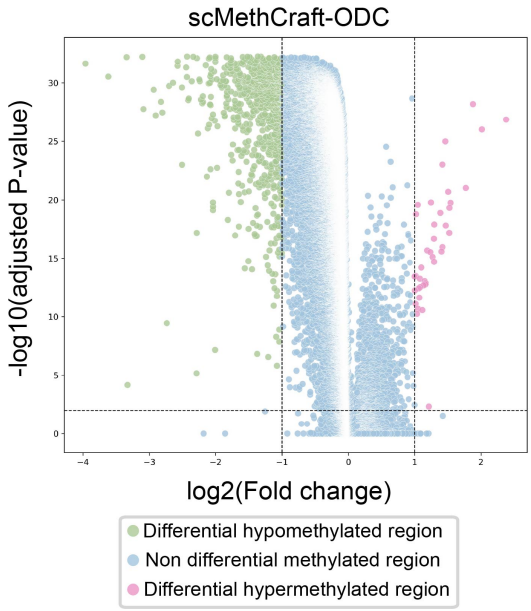


Fig. R27 ODC-specific DMRs identified by scMethCraft on the ACC dataset.

noted in previous studies, lower DNA methylation levels are often associated with higher gene expression, and many publicly available marker gene databases primarily catalog genes with high expression levels. As a result, hypomethylated DMRs are more likely to overlap with well-characterized marker genes and are supported by independent lines of evidence. Consistent with this, prior single-cell DNA methylation studies have predominantly focused on hypomethylated DMRs for functional interpretation[1-3].

COMMENT 12

I appreciate the fact that authors broke their network down into different modules and show the performance. It is somewhat counterintuitive yet interesting that it is the one-hot-encoded DNA sequence modality that yields the highest improvement in clustering, whereas my expectation would be that the improvement comes from cell-to-cell similarity module, as that allows improve data sparsity. Furthermore, the DNA sequence elements exhibit no variation across cell types; thus, it is surprising to see that that information can help identify better classification. I think this is worth touching upon in the discussion further, especially if the authors have clear insights into this, as the discussion section is relatively short.

RESPONSE TO COMMENT 12

We thank the reviewer for carefully examining the modular design of scMethCraft. Below, we address the two key points raised by the reviewer.

First, we clarify the functional relationship between the sequence feature extraction module and the similarity weighting module. scMethCraft is fundamentally a sequence-based model, in which the sequence feature extraction module provides the primary basis for methylation prediction. The similarity weighting module is therefore built on top of the sequence feature extraction module and serves as a refinement mechanism rather than a standalone driver of performance. Its primary role is to reweight and smooth sequence-based predictions by leveraging learned intercellular correlations. In this sense, the similarity weighting module complements the sequence module.

Second, we emphasize that genomic sequence information is not treated as a cell-level feature in scMethCraft. Instead, DNA sequences of genomic regions are modeled as inputs, while measured methylation levels are used as supervision. Although the underlying DNA sequence is shared across cells, scMethCraft learns how sequence-derived features interact with cell-specific methylation levels, for example, which genomic region features are associated with low or high methylation levels in particular cells. Thus, scMethCraft learns a low-dimensional representation that encodes how genomic regions give rise to distinct

methylation patterns across different cell populations. As a result, the learned representations become informative for distinguishing cell types, even though the raw DNA sequence itself does not vary between cells.

Taken together, the sequence feature extraction module captures the association between genome regions and cell-specific methylation levels, while the similarity weighting module further improves performance by smoothing predictions using correlated methylation information across cells. We have expanded the Discussion section in the revised manuscript to explicitly address this point and clarify the complementary roles of these two modules.

COMMENT 13

I could not find clearly where the labels for performance evaluation are obtained from, especially in for Figure 2. Did the authors generate these labels or are they obtained from the original studies?

RESPONSE TO COMMENT 13

We apologize for the lack of clarity regarding the source of the labels used for performance evaluation. For Brain datasets, the cell type labels were not generated by us but were obtained directly from the original studies associated with each dataset (Supplementary Table S1). In those original works, cell types were annotated based on low-methylated gene markers. For the newly added PBMC dataset (https://ngdc.cncb.ac.cn/gsa-human/s/Trk7H9dU), where paired scRNA-seq data are available, we performed cell type annotation based on well-established marker genes in the scRNA-seq modality and transferred these labels to the scDNAm data (Supplementary Fig. S19). We have now explicitly clarified the origin of these labels in the Methods section of the revised manuscript to avoid any ambiguity.

COMMENT 14

The authors have created a clean and user-friendly github repository for this method. However, in my review I noticed that scBasset has strongly informed the genomic sequence module. I feel like this should be acknowledged explicitly within the paper given this module is a central contributor to the success of the method (based on Figure 6). Proper attribution would improve transparency and clarify novelty relative to prior work.

RESPONSE TO COMMENT 14

We thank the reviewer for this thoughtful comment and for highlighting the importance of proper attribution and transparency. We agree that the genomic sequence module in

scMethCraft is conceptually informed by prior work, particularly scBasset, and that this relationship should be stated more explicitly in the manuscript.

In scMethCraft, we adopt a sequence-based modeling strategy inspired by scBasset to extract informative features from genomic sequences. To accommodate the distinct characteristics of single-cell DNA methylation data, including region-based binning and explicit missing values, we systematically modified multiple components of the model, including the input representation, loss function, and training strategy, to better suit scDNAm data (see Response #1 to Reviewer #2).

Beyond these adaptations, scMethCraft further extends the sequence modeling component by jointly integrating multiple sequence encoding perspectives and explicitly incorporating positional information of methylated regions through a hybrid neural network architecture, enabling more effective capture of sequence-derived regulatory features. Moreover, scMethCraft explicitly models cell-to-cell similarity and introduces an iteratively updated similarity-weighted module. This module dynamically computes intercellular similarity during training and incorporates methylation information from similar cells into the prediction process, thereby smoothing imputed methylation levels using observed signals. This design reduces single-cell-level prediction bias, improves cell embeddings, and provides a more reliable foundation for downstream tasks such as batch integration and cell type annotation.

Overall, while the sequence modeling strategy in scMethCraft is inspired by scBasset, scMethCraft substantially extends prior work both methodologically and architecturally, resulting in a comprehensive framework specifically designed for the processing and analysis of single-cell DNA methylation data. Following the reviewer's suggestion, we have revised the main text and GitHub documentation to explicitly acknowledge the influence of scBasset on the design of the genomic sequence module and to more clearly delineate the novelty of scMethCraft relative to prior work. In addition, we have further improved the usability of the code, making it more user-friendly.

COMMENT 15

The size of the genomic windows used in the algorithm is barely mentioned in main text. No reason why that number has to be only in the methods section, as this is an important parameter of the model.

RESPONSE TO COMMENT 15

We thank the reviewer for pointing out that the genomic window (bin) size was insufficiently highlighted in the main text. We have now added a clear description of the default bin size used

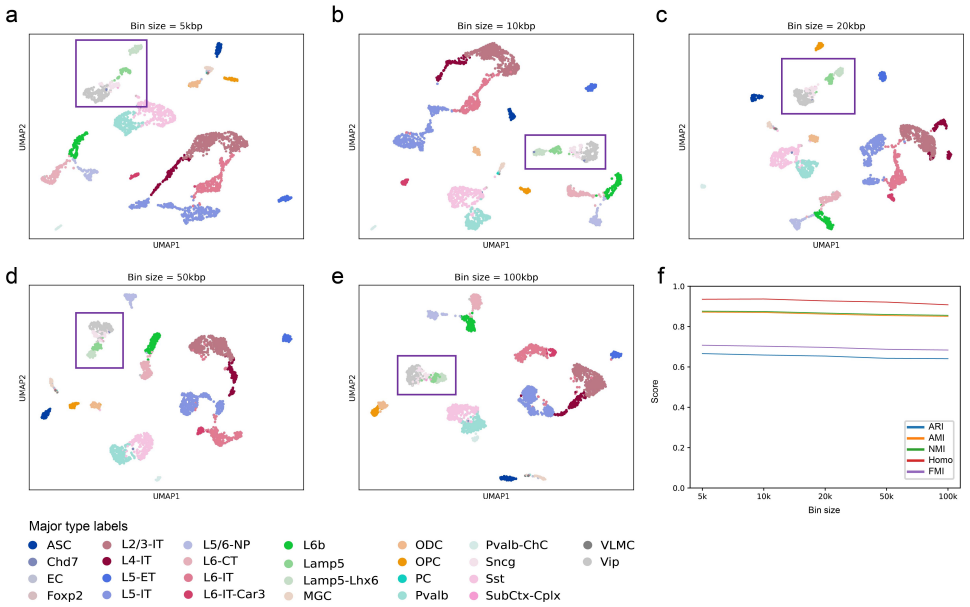


Fig. R28 Performance of scMethCraft with varying bin size. a-e, UMAP visualization of the embedding created by scMethCraft with varying bin size. f, Cell clustering performance of various bin sizes.

by scMethCraft and the corresponding baselines in the main text. We also emphasize that scMethCraft's design allows users to easily adjust the bin size to accommodate different datasets or analysis requirements.

As discussed in our Response to Comment #1 of Reviewer #1, for example, in the ACC dataset, we partitioned the data using bin sizes of 5 kbp, 10 kbp, 20 kbp, 50 kbp, and 100 kbp. We observed that bin sizes below 100 kbp have a limited impact on data visualization and clustering results (Fig. R28). Nevertheless, some closely related cell types, such as Vip, Sncg, Lamp5, and Lamp5-Lhx6, may become more difficult to distinguish with larger bin sizes, as indicated by the purple boxes. We have incorporated bin size as an optional parameter in scMethCraft and added relevant discussion in Supplementary Note S4 to accommodate the needs of different users.

COMMENT 16

In the intro, authors claim " 57 observed in scRNA-seq or scATAC-seq data4, 11. Finally, gene expression often inversely 58 correlates with methylation levels." and argue after that other algorithms for sc-RNA and sc-ATAC may not perform well because they are not capturing an inverse correlation. The actual concern is that methylation is typically represented by percentage of GCs methylated rather than a count based paradigm, as in RNA and ATAC. If methylation is also represented as counts, there is not reason the anti-correlation of methylation to gene expression (which is not always the case) should make standard dimensionality reduction methods less applicable.

RESPONSE TO COMMENT 16

We thank the reviewer for the careful reading of our manuscript and apologize for any confusion caused by our wording. Our intention in the Introduction was not to argue that the inverse correlation alone renders existing methods inapplicable, but rather to emphasize the fundamental differences in data characteristics between single-cell DNA methylation and other single-cell modalities.

First, we would like to clarify the motivation behind our original statement. In scRNA-seq and scATAC-seq data, higher counts typically correspond to stronger biological signals, whereas lower counts often indicate weak or absent signals and tend to be noisier. As a result, many existing methods are implicitly or explicitly designed under the assumption that larger values carry stronger biological meaning. In contrast, single-cell DNA methylation data are commonly represented as regional methylation ratios, in which higher values do not inherently indicate stronger signals. This mismatch between numerical scale and biological semantics can lead to suboptimal performance when such methods are directly applied. Using snapATAC2 on the ACC dataset as an example, we found that directly using methylation percentages as input leads to confusing low dimensional representations (Fig. R29a). Simply transforming the data to 1 minus methylation percentage substantially improves the dimensionality reduction performance (Fig. R29b), as also adopted in our baseline evaluations. We agree with the reviewer that this effect is not necessarily attributable to the inverse relationship with gene expression itself, but rather to this scale-semantics mismatch inherent to scDNAm data, and we have revised the Introduction accordingly to clarify this point.

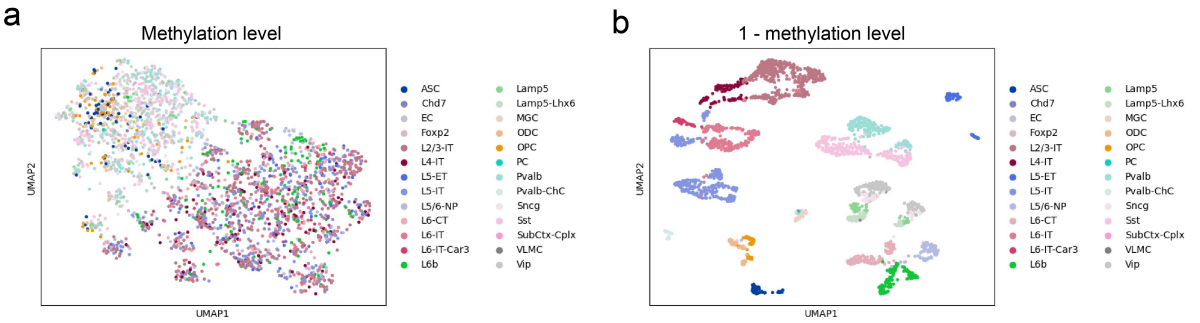


Fig. R29 Dimensionality reduction performance of SnapATAC2 using methylation level and 1 minus methylation level as input.

Second, while the reviewer suggests representing methylation data in a count based manner, which is a valuable perspective, we note that nearly all existing studies represent DNA methylation as regional methylation percentages. We explored a count based representation by aggregating methylated cytosines within 10 kbp bins. However, we found that such count based features fail to adequately capture cell specific methylation patterns, and the resulting counts show no strong correlation with methylation percentages, limiting their biological interpretability in this context (Fig. R30).

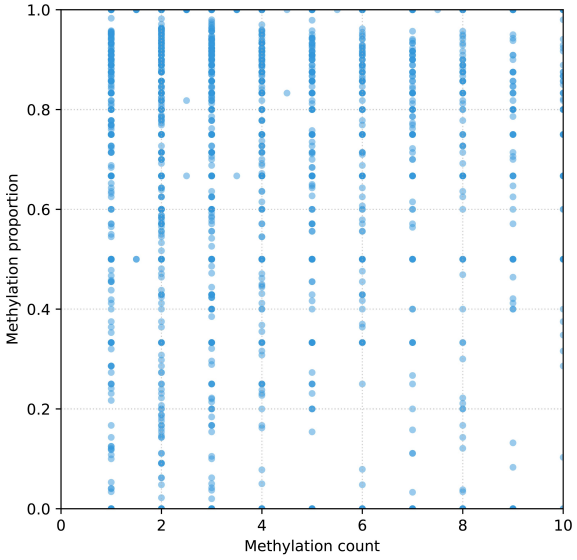


Fig. R30 Distribution of average DNA methylation levels within each cell type.

In summary, we have revised the Introduction to emphasize that the primary challenges motivating scMethCraft arise from the mismatch between numerical scale and biological semantics in scDNAm data, rather than from its inverse relationship with gene expression, thereby avoiding potential misunderstanding for readers.

COMMENT 17

Typo

148 knowledge, no exiting method have been specifically designed for the dimensionality reduction

RESPONSE TO COMMENT 17

Thank you for pointing out this typo. We have corrected the grammatical error in the revised manuscript and carefully rechecked the entire manuscript to avoid similar issues.

REVIEWER #3 (REMARKS TO THE AUTHOR)

Wang et al. introduce a new neural network-based model for the analysis of single cell DNA methylation data, named scMethCraft. Their model uses a hybrid neural network architecture to predict DNA methylation within genomic bins. The major novelty of this approach is the incorporation of DNA sequence information into the modeling of DNA methylation data, achieved through the hybrid network. Overall this fills an important need in the area of single-cell DNA methylation analysis, and the authors demonstrate versatile use-cases including cell clustering, batch correction, and data imputation. However, there are a few points that require further consideration, outlined below:

RESPONSE

We thank the reviewer for their positive evaluation of our study and for the insightful and constructive comments. These suggestions have been carefully considered and have led to substantial improvements in both the clarity and completeness of the revised manuscript. All points raised by the reviewer have been addressed in detail, with corresponding revisions made in the relevant sections.

COMMENT 1

1. There is no information given about the computational resources required to train the model. It is important that the authors can show the time required for training their model, how this scales with increasing numbers of cells, what GPU hardware is required, and how this compares with other competing methods.

RESPONSE TO COMMENT 1

We appreciate the reviewer for emphasizing the need to clearly document computational requirements, which are crucial for evaluating the real-world applicability of the proposed method. We concur that runtime, memory consumption, scalability with increasing data size, and comparisons with existing approaches are key factors for practical adoption. In the following, we address these points by detailing the computational resources used, analyzing the theoretical complexity, and discussing performance relative to other methods.

First, all experiments were performed using a single NVIDIA RTX 3090 GPU with 24 GB of memory. The computing environment consisted of an Intel(R) Xeon(R) Platinum 8375C CPU (128 cores), Ubuntu 22.04.3 LTS, and 1 TB of system memory. We evaluated scMethCraft on datasets of increasing size, including Mop (296 cells), ACC (2,698 cells), and Mixed (~16,000 cells), using a batch size of 128. For each dataset, we recorded both peak GPU

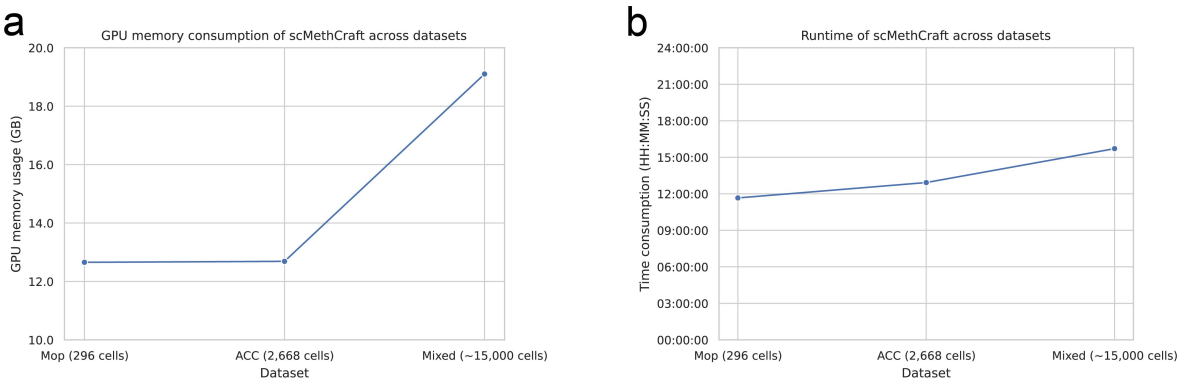


Fig. R31 GPU memory consumption and runtime of scMethCraft across datasets of increasing scale. a, Peak GPU memory usage during model training. **b,** Runtime of model training.

memory usage and training time (Fig. R31). As the dataset size increased, GPU memory usage rose from approximately 13 GB to 20 GB, remaining within the capacity of a single RTX 3090. Moreover, memory consumption can be further reduced by decreasing the batch size, allowing scMethCraft to run on GPUs with more limited resources. Across all tested datasets, the complete analysis pipeline finished within approximately half a day, which we consider practical for routine biological data analysis.

Secondly, from an algorithmic perspective, scMethCraft consists of two main components: a genomic sequence feature extraction module and a cell-cell similarity weighting module. The sequence module processes genomic bins and, in each training epoch, effectively scans the reference genome. Because the genome length is fixed, using finer binning increases the number of bins while proportionally reducing their size, resulting in an approximately constant overall computational burden. With a batch size of 128, this module requires around 12 GB of GPU memory during training. The similarity weighting module explicitly models relationships between cells based on methylation patterns. In its original form, this step has a complexity of $O(n^2)$. However, for dataset sizes commonly encountered in current scDNAm studies (thousands of cells), this cost remains tractable in practice. To further improve scalability, we introduced an extended implementation based on a low-rank approximation of the similarity matrix. Assuming there is a total of n cells, we denote the similarity matrix obtained through iterative training as $\mathbf{S} \in R^{n \times n}$. Let $\mathbf{P} \in R^{n \times k}$ denote the trainable parameters of the similarity weighting module, where k is the hyperparameter (default 256). To enforce the non-negativity, symmetry, and a zero-valued main diagonal property of the similarity matrix $\mathbf{S}$, we define it as:

$$\mathbf{S} = ((|\mathbf{P}||\mathbf{P}|^{\mathbf{T}}) - \mathrm{diag}(\mathbf{P}\mathbf{P}^{\mathbf{T}}))$$

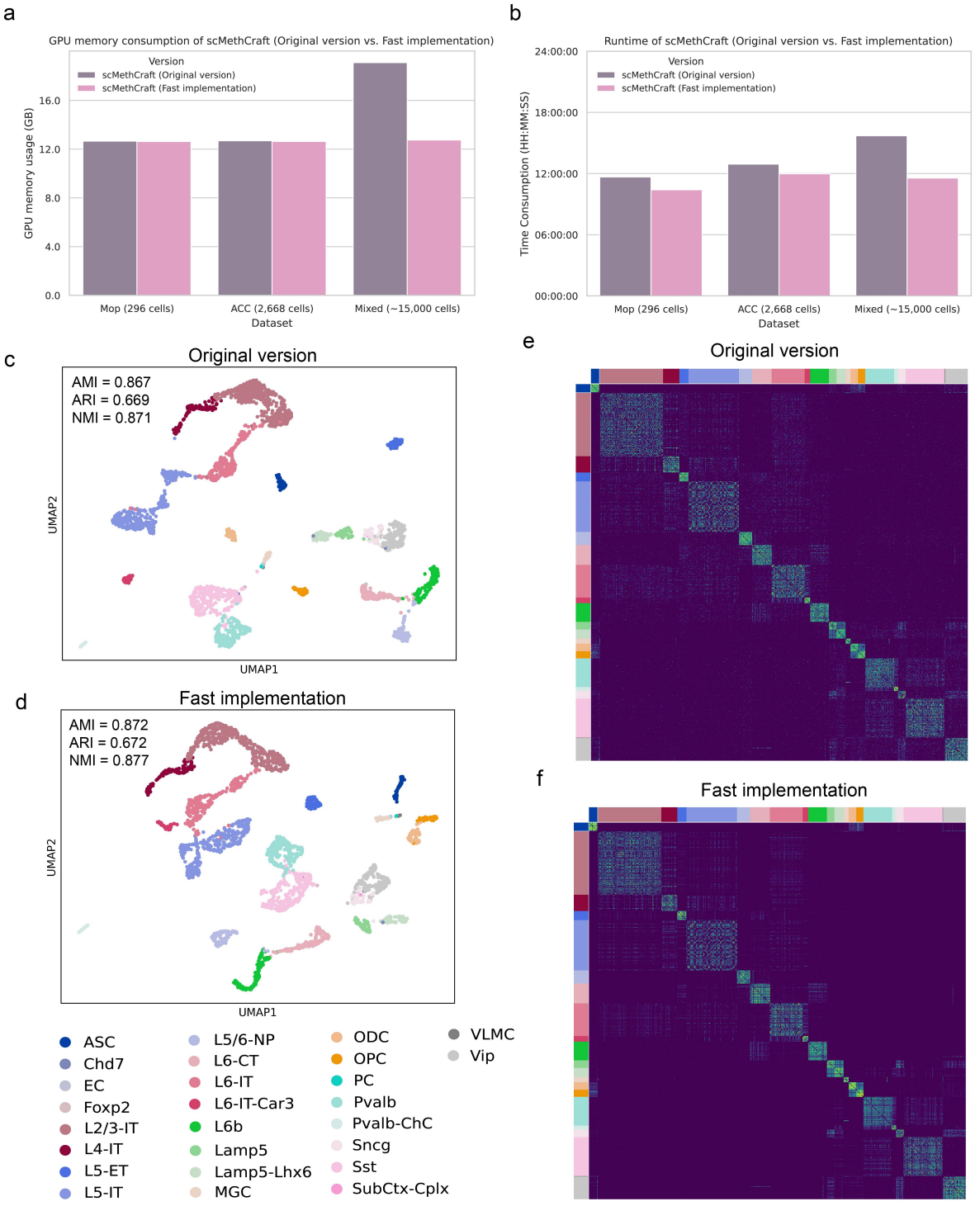


Fig. R32 Performance comparison of scMethCraft between the original version and the fast implementation. a, GPU memory consumption, and **b,** runtime of scMethCraft across datasets of increasing scale, contrasting the original version with the fast implementation. **c-d,** UMAP visualization and clustering performance of cell embeddings produced by **(c)** the original version and **(d)** the fast implementation. **e-f,** The heatmap of the cell-to-cell similarity matrix created by **(e)** the original version and **(f)** the fast implementation, where cells were sorted according to their major cell types. All similarity values are linearly scaled to the range 0-1.

By factorizing the similarity matrix, all operations involving the similarity matrix are implemented through its factorized form, and the full $n \times n$ matrix S is never explicitly constructed or stored in memory. This reduces the effective computational complexity from $O(n^2)$ to approximately $O(n)$ (Fig. R32a,b). Importantly, this approximation affects only the computational implementation and does not change the biological assumptions or modeling goals. Empirical evaluation shows that this extension preserves the ability to capture meaningful intercellular similarities and yields performance trends consistent with the original model (Fig. R32c-f).

Finally, we discuss the runtime and memory usage of scMethCraft in comparison with competing methods. We acknowledge that, compared with some methods based on linear dimensionality reduction or lightweight statistical models, scMethCraft incurs a higher runtime. This difference primarily arises from its deep learning architecture, which requires sufficient training even on relatively small datasets.

Moreover, we believe the computational cost of scMethCraft is reasonable and acceptable. First, in practice, the observed runtimes fall well within time scales that are acceptable for routine biological data analysis applied to contemporary scDNAm datasets containing thousands to tens of thousands of cells, a setting that increasingly reflects current experimental scales. Second, scMethCraft demonstrates consistent performance improvements across multiple datasets and evaluation metrics, while substantially enhancing biological interpretability. Third, scMethCraft integrates multiple key analytical tasks, including dimensionality reduction, imputation, batch correction, and differential region identification, within a single framework, allowing users to perform multiple downstream analyses with a single model training, thereby simplifying the overall analysis workflow. Fourth, after architectural optimization, its overall computational complexity scales approximately linearly with the number of cells, preventing exponential growth in computational cost for larger datasets. In contrast, many existing scDNAm data analysis methods were originally developed for smaller datasets and are therefore primarily applicable in limited-scale scenarios. For example, methods such as Camelia[24] and DeepCpG[25] encounter practical memory limitations as dataset sizes increase, while Epiclomal[26] and MambaCpG[27] typically require substantially longer runtimes and are primarily suited to datasets containing on the order of hundreds of cells.

In summary, while scMethCraft relies on a relatively expressive model architecture, our results show that its runtime and memory usage are well controlled and scale favorably in real-world settings. The incorporation of a low-rank approximation further improves scalability, enabling efficient analysis of larger datasets. In the revised manuscript, we have expanded the Methods section to include a detailed description of computational resource usage and the optimized implementation (Methods, Supplementary Figs. S18, S20). We hope the reviewer

will agree that these additions sufficiently address the concerns regarding computational efficiency and demonstrate the practical feasibility of scMethCraft for scDNAm data analysis.

COMMENT 2

2. The authors also present the identification of DMRs, and propose that their model is better able to identify these regions compared to other methods. However, these differential tests have been performed on the "enhanced" data from the model, meaning that the cells are no longer independent observations. This violates the assumptions of the statistical test and will result in inflated significance values and inflated false discovery rate. The methods also do not state that any multiple testing correction was performed on the p-values obtained.

RESPONSE TO COMMENT 2

We thank the reviewer for raising important concerns regarding the statistical validity of DMR identification when performed on model-enhanced single-cell data. We agree that using enhanced scDNAm profiles for DMR identification can further increase dependency among cells, thereby violating the assumption of independent observations and potentially inflating statistical significance. This is an important limitation that requires careful consideration.

To address this concern, we have applied a strict multiple testing correction to all P-values obtained from DMR identification. Specifically, we adopted the Bonferroni correction to control the family-wise error rate. As a highly conservative procedure, Bonferroni correction substantially reduces false positives and helps mitigate the inflation of significance arising from non-independence among single-cell observations. In the revised manuscript, Bonferroni-adjusted P-values are used as the default criterion for DMR identification in scMethCraft. All DMR results and downstream analyses have been re-evaluated accordingly (Fig. R33). Importantly, these revisions did not alter the main biological conclusions of the study.

In summary, we adopted a conservative statistical strategy for DMR identification by applying strict Bonferroni correction to account for the non-independence among enhanced single-cell profiles. We have clarified this choice in the revised Methods section and re-evaluated all downstream analyses using Bonferroni-adjusted P-values. Importantly, despite the increased stringency, the key downstream results and biological conclusions remain largely unchanged (Fig. 5 and Supplementary Fig. S17).

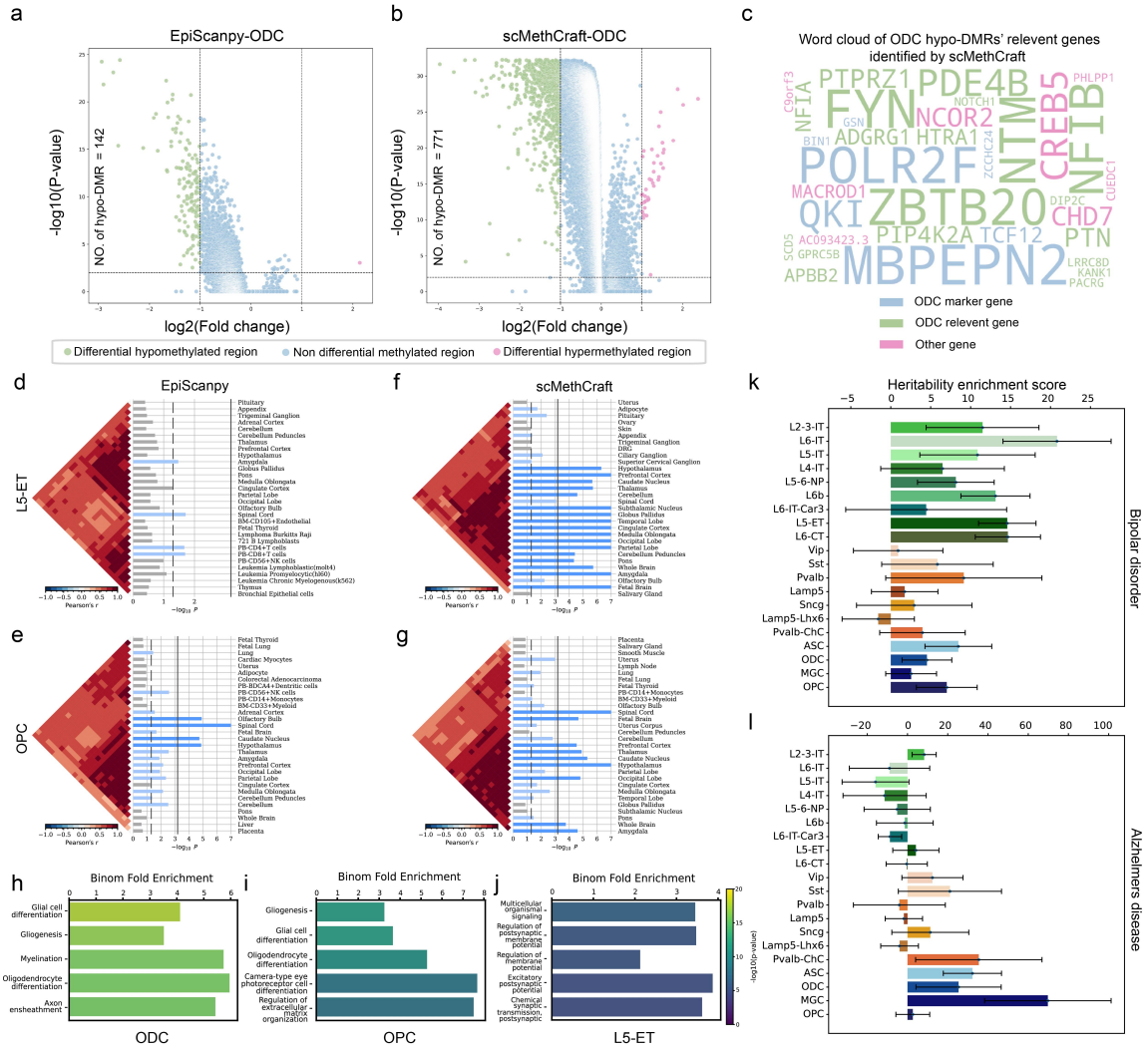


Fig. R33 scMethCraft reveals cell-type-specific biological insights. a-b, Volcano plots of ODC-specific DMRs identified by EpiScanpy and scMethCraft. **c,** Word cloud for 36 genes associated with ODC hypo-DMRs identified by scMethCraft, appearing in at least three instances. Gene font size is positively correlated with the number of associations. **d-g,** Cell-type-specific regions and background regions were subjected to SNPsea analysis to determine the top 30 significantly enriched tissues. Bars represent the empirical P-values calculated by the SNPsea algorithm. The significance of gene enrichment for a specific annotation was assessed using one-sided P-value cutoffs at the 0.05 level, represented by vertical dashed and solid lines. The unadjusted P-value cutoff is indicated by the dashed line, while the Bonferroni-corrected P-value cutoff is shown by the solid line. The heatmaps display the Pearson correlation coefficients, which measure the similarity between expression profiles. Expression profiles were ordered using hierarchical clustering with the unweighted pair-group method with arithmetic means (UPGMA). **h-j,** Top five enriched GO terms associated with major type-specific hypo-DMRs identified by scMethCraft, along with binomial test values (two-sided, not adjusted for multiple comparisons) and enrichment, as reported by GREAT. **k-l,** Heritability enrichment estimated by stratified LDSC of the SNPs within the major type-specific hypo-DMRs identified by scMethCraft, for two brain-related diseases. The error bars and centers of error bars represent the standard errors and average values of 200 equally sized jackknife blocks of adjacent SNPs about the estimates of enrichment.

COMMENT 3

3. The authors focus their analysis and evaluations solely on the brain; some examples in other tissues are required to demonstrate that the model can be applied to a diversity of tissues rather than only the brain.

RESPONSE TO COMMENT 3

We thank the reviewer for this important comment. Firstly, scDNAm studies to date have indeed been predominantly focused on the brain. This is largely because neuronal cell types exhibit exceptionally pronounced and stable DNA methylation differences, making methylation a particularly informative modality for resolving cell identity and cellular heterogeneity in brain tissue. Multiple studies have demonstrated that cell-type-specific methylation patterns are especially strong and reproducible in the mammalian brain, enabling reliable cell-type discrimination using methylation data alone[28-30]. As a result, several large-scale and high-quality scDNAm atlases have been generated for the human and mouse brain, providing both sufficient cell numbers and well-characterized biological annotations for systematic method evaluation[1, 2]. In contrast, scDNAm datasets for non-brain tissues remain

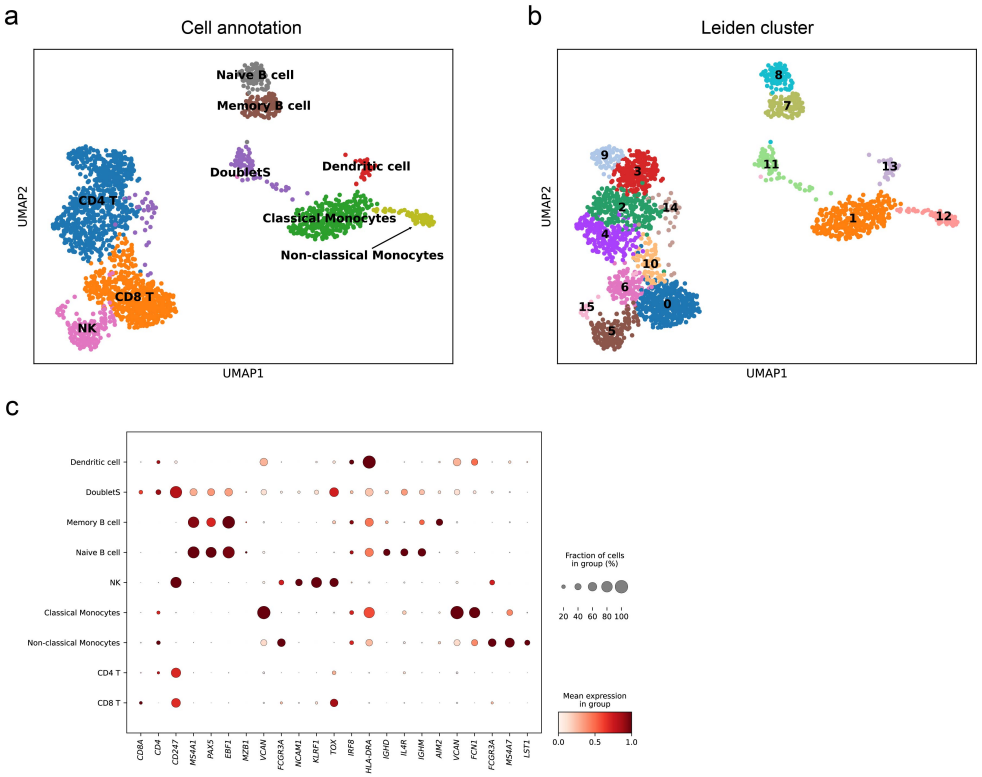


Fig. R34 Cell type annotation based on paired scRNA-seq data. a-b, UMAP visualization of scRNA-seq data colored by cell-type annotations and Leiden clusters, illustrating the correspondence between transcriptome-based clustering and manual cell-type annotation. **c**, Dot plot of the expression of canonical marker genes across annotated cell types based on paired scRNA-seq data.

relatively scarce and are typically characterized by smaller sample sizes and lower sequencing depth. Moreover, methylation patterns in many peripheral tissues tend to be less cell-type-specific, making it difficult to confidently assign cell identities based on methylation alone. These factors substantially limit the feasibility of rigorous benchmarking and biological interpretation when relying solely on scDNAm data from non-brain tissues.

Secondly, to directly address the reviewer's concern and to evaluate the applicability of scMethCraft beyond the brain, we additionally collected a human peripheral blood mononuclear cell (PBMC) single-cell multi-omics dataset from SeekGene BioSciences (https://ngdc.cncb.ac.cn/gsa-human/s/Trk7H9dU). This dataset contains matched single-cell transcriptomic and DNA methylation profiles from the same cells. We used the paired transcriptomic data to perform reliable cell-type annotation based on established marker genes (Fig. R34), and then applied scMethCraft to the corresponding scDNAm data. Comprehensive evaluations of performance demonstrate that scMethCraft remains effective on non-brain tissues and generalizes well across different sequencing technologies (Details of performance evaluation are provided in Response #7 to Reviewer #3).

Together, while brain tissue provides a uniquely informative and well-supported setting for scDNAm method development, scMethCraft is not limited to brain-specific applications and can be robustly applied to other tissues.

COMMENT 4

4. In the evaluation of data integration results, the authors have only used examples where the same cell types are present across all datasets. An important aspect of integration performance that is therefore overlooked is the ability of the model to perform accurate integration when some of the cell types are unshared across datasets.

RESPONSE TO COMMENT 4

We thank the reviewer for raising this important point. We agree that a critical aspect of data integration performance is the ability to correctly handle scenarios in which cell type compositions are only partially shared across datasets, particularly the ability to avoid incorrectly merging biologically distinct cell populations, a phenomenon commonly referred to as overcorrection.

To explicitly address this concern, we performed additional experiments to evaluate the batch integration behavior of scMethCraft when certain cell types are present in only a single batch. Specifically, we constructed a partially overlapping scenario by artificially removing

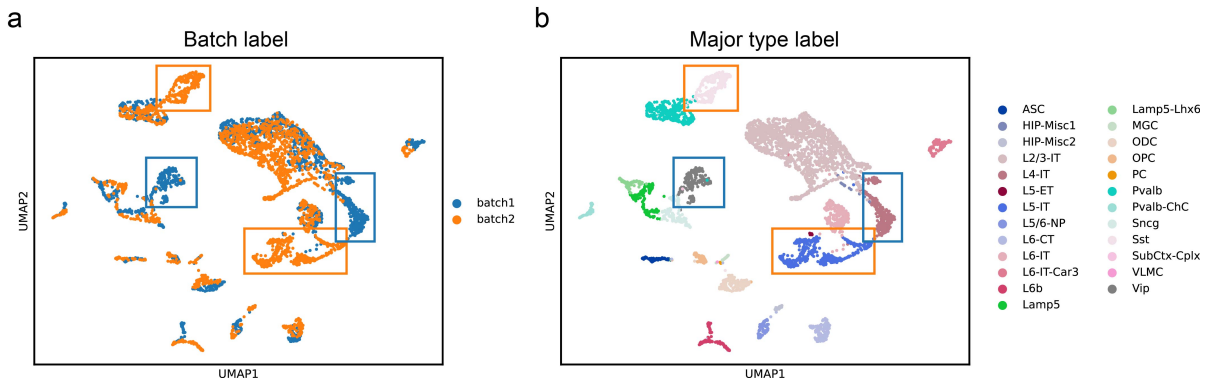


Fig. R35 Batch integration behavior of scMethCraft under partially overlapping cell type compositions. **a,** UMAP visualization colored by batch labels. **b,** UMAP visualization colored by cell-type annotations. Blue boxes indicate cell types present exclusively in batch 1, while orange boxes indicate cell types present exclusively in batch 2.

L5-IT and SST cells from batch 1 and removing L4-IT and VIP cells from batch 2. These cell types originally exhibited good batch integration, and excluding them allows us to assess scMethCraft's performance under a more challenging partially overlapping condition. We then applied scMethCraft to perform batch integration under these conditions. The results show that scMethCraft successfully aligns shared cell types across batches, while preserving batch-specific (unshared) cell types as distinct groups rather than forcing them to merge with unrelated populations (Fig. R35). This behavior demonstrates that scMethCraft is capable of achieving accurate integration under heterogeneous cell type compositions, while maintaining intrinsic cellular heterogeneity and avoiding overcorrection.

In addition, as discussed in Response #5 to Reviewer #1, we further examined integration performance in the presence of known and biologically meaningful differences between cell subtypes. In these analyses, scMethCraft preserves cell subpopulations with established biological differences while integrating populations that are biologically consistent, demonstrating a balanced ability to remove technical noise while retaining biologically meaningful heterogeneity.

For completeness and ease of review, we briefly reiterate and summarize those findings here. Analysis of L2/3-IT neurons shows that scMethCraft achieves effective batch integration while respecting biologically defined subtype structure. Most L2/3-IT cells from MTG and Pro overlap substantially in the integrated UMAP space, indicating successful alignment of shared populations across cortical regions (Fig. R36a,b). At the same time, multiple Pro-derived L2/3-IT cells remains separated. Examination of the original subtype annotations reveals that these non-overlapping cells predominantly belong to differental methylated gene defined subtypes that are largely absent in MTG, indicating genuine biological divergence rather than residual batch effects (Fig. R36c,d). In contrast, L2/3-IT subtypes shared between MTG and Pro are

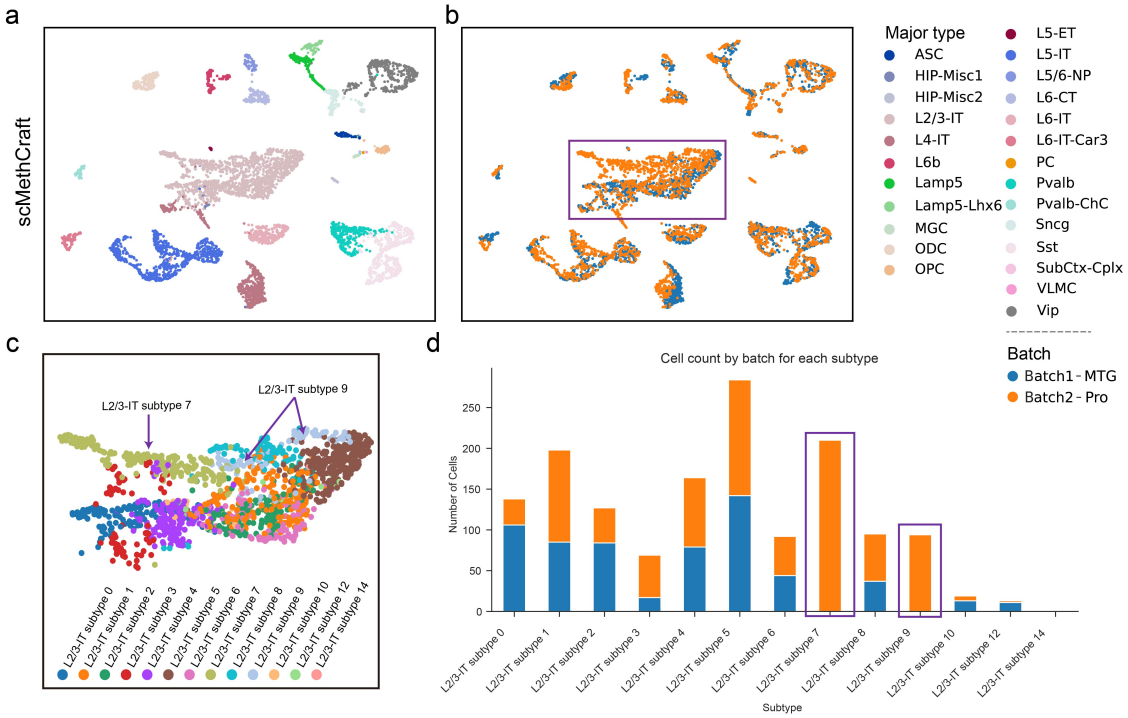


Fig. R36 scMethCraft preserves L2/3-IT subtypes during batch integration. a-b, UMAP visualization of the integrated embeddings provided by scMethCraft on the MTG-Pro dataset colored by **(a)** major type and **(b)** batch. The purple box indicates L2/3-IT cells. **c,** The L2/3-IT cells of scMethCraft's integrated embedding, colored by subtype. **d,** Subtype-specific batch composition of L2/3-IT cells.

well integrated and remain proximal in the embedding space, preserving subtype-specific coherence. Together, these results demonstrate that scMethCraft integrates biologically similar populations across batches while maintaining separation for region-specific subtypes, thereby avoiding overcorrection and preserving meaningful biological heterogeneity.

Taken together, these results indicate that scMethCraft performs robust and biologically faithful data integration in more realistic scenarios involving partially overlapping cell type compositions. By effectively aligning shared populations while preserving unshared or biologically distinct cell types, scMethCraft achieves a balance between batch effect removal and the retention of biologically meaningful heterogeneity. We have discussed the issue of overcorrection in the revised main text and Supplementary Figs. S8-S9.

COMMENT 5

5. In the assessment of cell annotation performance, the authors have performed clustering across a variety of resolutions until they obtain the same number of clusters as are annotated cell types. This likely inflates performance of the cell annotation method, as the authors ensure the clustering is

somewhat similar to the cell annotations by optimizing the resolution. Instead the authors should show performance with clustering at a fixed resolution, chosen independently of the cell type labels.

RESPONSE TO COMMENT 5

We thank the reviewer for this insightful comment regarding the evaluation protocol. To address the reviewer's concern, we additionally performed clustering on all 11 datasets using the Louvain algorithm implemented in Scanpy[6] with a default resolution of 1, chosen independently of the cell type labels. As shown in Fig. R37, scMethCraft still achieves the best performance across five different evaluation metrics, demonstrating that its advantage is consistent and robust under different clustering parameter settings.

Nevertheless, after careful consideration, we retain the evaluation strategy in which the number of clusters is matched to the number of annotated cell types. The rationale for this choice is as follows. First, in the evaluation of dimensionality reduction or imputation methods for single-cell sequencing data, clustering cells into a number of groups equal to the known cell type labels is a widely adopted and well-accepted practice, and has been extensively used in prior studies[8, 16, 31-33]. This strategy aims to assess whether the learned low-dimensional representations are capable of separating biologically meaningful cell populations, rather than optimizing unsupervised clustering performance. Second, the latent spaces produced by different dimensionality reduction methods can vary substantially in structure and scale. Under a fixed clustering resolution, the resulting number of clusters may differ dramatically across methods, which can lead to an unfair comparison that is confounded by clustering behavior rather than reflecting the intrinsic quality of the learned representations. Matching the number

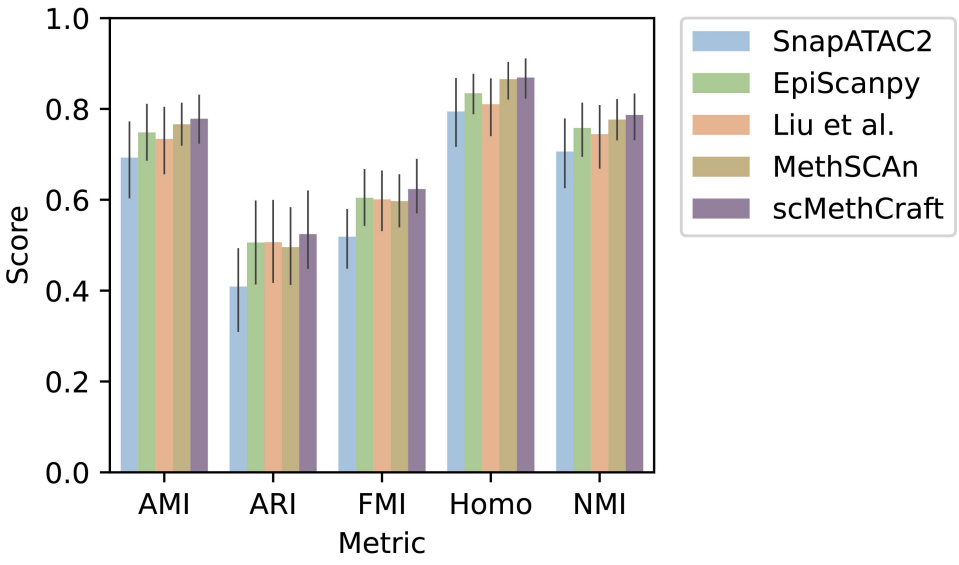


Fig. R37 Cell clustering performance of embeddings generated by different methods evaluated at the default clustering resolution using five metrics across 11 datasets.

of clusters to the annotated cell types helps mitigate this issue and enables a more controlled and equitable evaluation across methods.

COMMENT 6

Furthermore, it was unclear where the ground truth cell type labels used for evaluating cell annotation performance originated from.

RESPONSE TO COMMENT 6

We apologize for not clarifying this point in the manuscript. The ground truth cell type labels of brain datasets used for evaluation were obtained directly from the original studies associated with each dataset. In the original publications, cell types were annotated based on low-methylated gene markers. For the newly added PBMC dataset, we leveraged the paired scRNA-seq data to perform cell-type annotation based on well-established marker genes. We have now clarified the source of these labels in the Methods section of the revised manuscript to avoid any confusion for readers.

COMMENT 7

Finally, more in-depth benchmarking comparisons with other methods are needed. This includes Epiclomal, DeepCpG, MambaCpG, Melissa.

RESPONSE TO COMMENT 7

We thank the reviewer for this suggestion and for highlighting several important methods in the single-cell DNA methylation analysis. We agree that comprehensive benchmarking is essential, and below we clarify the scope of our comparisons and the rationale for including or excluding specific methods.

Epiclomal[26] is a probabilistic clustering method based on hierarchical mixture models that jointly performs clustering and imputation on sparse single-cell DNA methylation data. Melissa[20] is a Bayesian hierarchical model that leverages local methylation patterns to cluster cells and uncover epigenetic variability. DeepCpG[25] and MambaCpG[27] are deep learning-based methods primarily designed to predict CpG-level methylation states from local DNA sequence context, often incorporating information from neighboring CpG sites. These methods have played an important role in advancing single-cell methylation analysis and have substantially enriched the scDNAm data analysis toolbox.

However, all the above methods operate at single-CpG site resolution, either for site-level prediction or cell clustering. While this strategy is effective for datasets with relatively small numbers of cells, it becomes increasingly impractical as the dataset size grows. With the rapid increase in throughput of scDNAm technologies, a more modern and widely adopted analysis paradigm is to aggregate methylation signals across genomic regions to obtain average methylation levels within fixed bins, thereby constructing a cell-by-region scDNAm matrix. scMethCraft is specifically designed for this form of region-based scDNAm data and focuses on downstream tasks such as imputation, dimensionality reduction, and integration. As a result, these methods differ substantially from scMethCraft in terms of data input, application scenario, and modeling objective, making a direct comparison inappropriate.

In addition, the scalability of these CpG-level methods limits their applicability to large-scale datasets. In our study, the datasets typically contain thousands of cells. We attempted to apply Epiclomal, Melissa, DeepCpG, and MambaCpG to these datasets, but none of them could be completed within practical memory (up to 1 TB) or time (168 hours) constraints. Similar scalability limitations have been explicitly reported in the peer-review discussion of the MethSCAn study published in Nature Methods, which provides a detailed analysis of why several earlier scDNAm methods are no longer suitable as baselines for large-scale studies (https://static-content.springer.com/esm/art%3A10.1038%2Fs41592-024-02347-x/MediaObjects/41592_2024_2347_MOESM2_ESM.pdf).

To improve the completeness and fairness of our benchmarking, and following suggestions from other reviewers, we have now added comparisons with MethSCAn[3] and the bin-based pipeline following Liu et al.[2], both of which are matrix-based approaches and can be applied to datasets containing thousands of cells. The expanded benchmarking results show that scMethCraft consistently achieves superior performance across multiple datasets and in both dimensionality reduction and imputation tasks, including comparisons with the newly added baseline methods (Fig. R38). We have revised the manuscript accordingly to incorporate these additional comparisons and to better highlight these strengths (Figs. 2, 4, Supplementary Figs. S1-S5, Supplementary Notes S2, S6).

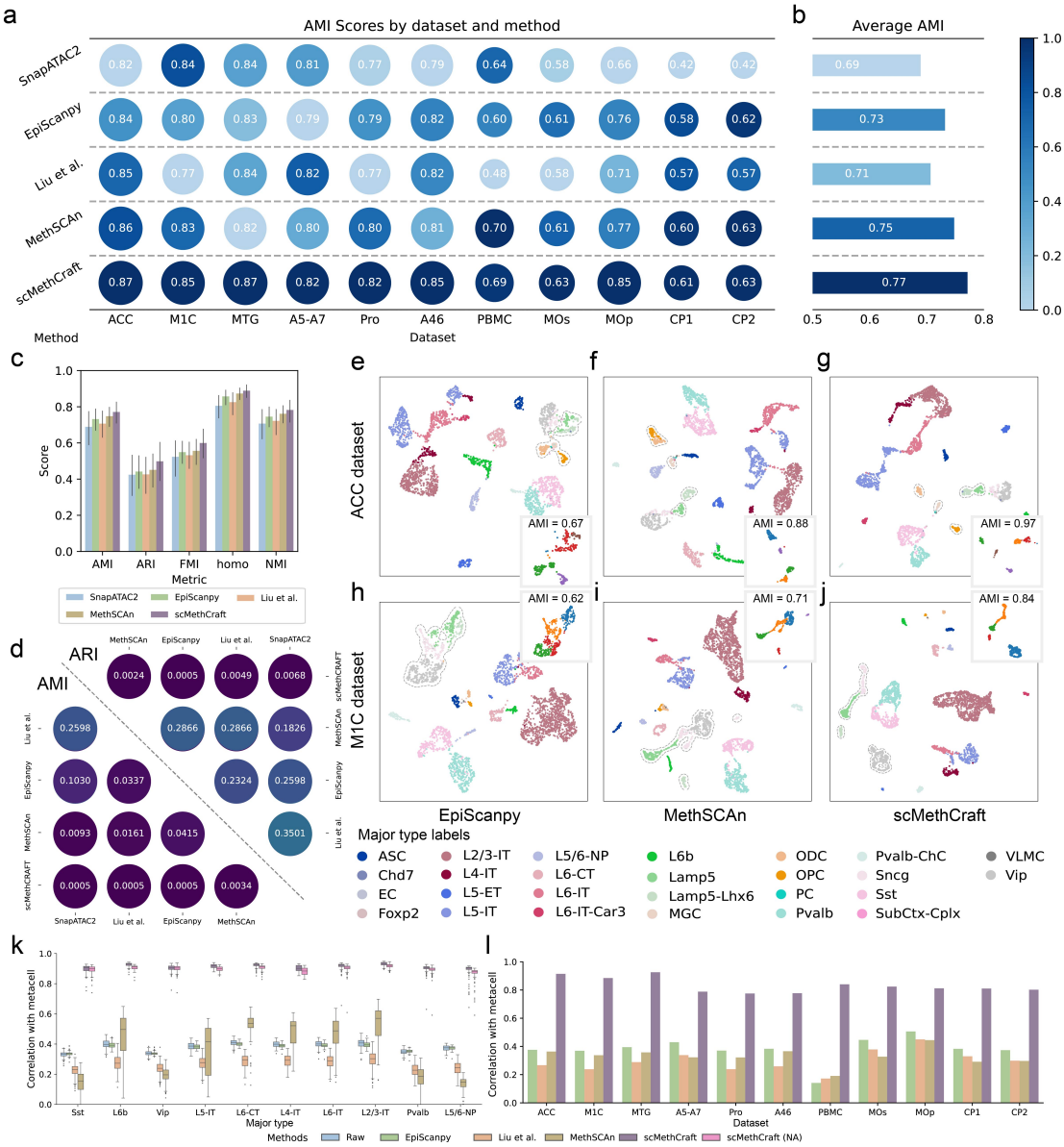


Fig. R38 scMethCraft consistently achieves superior performance in both dimensionality reduction and imputation tasks. a, Cell clustering performance assessed by AMI scores with respect to major cell types across different dimensionality reduction methods. Colors represent min-max normalized AMI scores within each dataset, where lighter blue indicates the lowest-performing method and darker blue indicates the highest-performing method. **b,** Average AMI scores with respect to major cell types, based on embeddings provided by different methods across 11 independent datasets. **c,** Cell clustering performance of embeddings generated by different methods was evaluated using five metrics across 11 datasets. **d,** One-sided paired Wilcoxon signed-rank tests were used to compare performance between methods. Each cell shows the P-value for the hypothesis that the method in the column outperforms the method in the row in terms of AMI or ARI across the 11 datasets. **e-j,** UMAP visualization of cell embeddings provided by different methods on the ACC and M1C datasets. Main panels are colored by major cell types, while inset panels are colored by cell clusters. The AMI values shown were computed only using the cell types highlighted by the boxes in the main panels. **k,** Spearman correlation between the average methylation level of each single cell and its corresponding meta-cell, computed for the raw matrix and for enhanced matrices generated by different methods across cell types in the ACC dataset. scMethCraft (NA) denotes correlations computed using only regions originally annotated as NA in the raw matrix. **l,** Spearman correlation between enhanced datasets and their corresponding meta-cells across 11 datasets processed by different methods.

REVIEWER #4 (REMARKS TO THE AUTHOR)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

RESPONSE

We thank the reviewer for the co-review and acknowledge this statement.

References

1. Tian, W. et al. Single-cell DNA methylation and 3D genome architecture in the human brain. *Science* **382**, eadf5357 (2023).
2. Liu, H. et al. DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120–128 (2021).
3. Kremer, L.P. et al. Analyzing single-cell bisulfite sequencing data with MethSCAn. *Nat. Methods* **21**, 1616–1623 (2024).
4. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
5. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
6. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
7. Xiong, L. et al. SCALE method for single-cell ATAC-seq analysis via latent feature extraction. *Nat. Commun.* **10**, 4576 (2019).
8. Tang, S. et al. scCASE: accurate and interpretable enhancement for single-cell chromatin accessibility sequencing data. *Nat. Commun.* **15**, 1629 (2024).
9. Li, W.V. & Li, J.J. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* **9**, 997 (2018).
10. Wang, L. et al. Loss of NARS1 impairs progenitor proliferation in cortical brain organoids and leads to microcephaly. *Nat. Commun.* **11**, 4038 (2020).
11. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015).
12. Wiseman, J.A., Dragunow, M. & I.-H. Park, T. Cell Type–Specific Nuclei Markers: The Need for Human Brain Research to Go Nuclear. *Neuroscientist* **29**, 41–61 (2023).
13. Saito, Y. et al. Differential NOVA2-mediated splicing in excitatory and inhibitory neurons regulates cortical development and cerebellar function. *Neuron* **101**, 707–720. e705 (2019).
14. Huang, Y. et al. Expression of Transcription Factor Satb2 in Adult Mouse Brain. *Anat. Rec.* **296**, 452–461 (2013).
15. Danese, A. et al. EpiScanpy: integrated single-cell epigenomic analysis. *Nat. Commun.* **12**, 5228 (2021).
16. Zhang, K., Zemke, N.R., Armand, E.J. & Ren, B. A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat. Methods* **21**, 217–227 (2024).
17. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
18. Teo, A.Y.Y., Squair, J.W., Courtine, G. & Skinnider, M.A. Best practices for differential accessibility analysis in single-cell epigenomics. *Nat. Commun.* **15**, 8805 (2024).
19. Gu, C., Liu, S., Wu, Q., Zhang, L. & Guo, F. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. *Cell Res.* **29**, 110–123 (2019).

20. Kapourani, C.-A. & Sanguinetti, G. Melissa: Bayesian clustering and imputation of single-cell methylomes. *Genome Biol.* **20**, 61 (2019).
21. Yuan, H. & Kelley, D.R. scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional neural networks. *Nat. Methods* **19**, 1088–1096 (2022).
22. Kapourani, C.-A., Argelaguet, R., Sanguinetti, G. & Vallejos, C.A. scMET: Bayesian modeling of DNA methylation heterogeneity at single-cell resolution. *Genome Biol.* **22**, 114 (2021).
23. Kipf, T. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
24. Tang, J. et al. CaMelia: imputation in single-cell methylomes based on local similarities between cells. *Bioinformatics* **37**, 1814–1820 (2021).
25. Angermueller, C., Lee, H.J., Reik, W. & Stegle, O. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol.* **18**, 1–13 (2017).
26. PE de Souza, C. et al. Epiclomal: probabilistic clustering of sparse single-cell DNA methylation data. *PLoS Comput. Biol.* **16**, e1008270 (2020).
27. Zhao, Q. et al. MambaCpG: an accurate model for single-cell DNA methylation status imputation using mamba. *Brief. Bioinform.* **26** (2025).
28. Lister, R. et al. Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341**, 1237905 (2013).
29. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
30. Mo, A. et al. Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369–1384 (2015).
31. Li, Z. et al. Chromatin-accessibility estimation from single-cell ATAC-seq data with scOpen. *Nat. Commun.* **12**, 6386 (2021).
32. Chen, S. et al. RA3 is a reference-guided approach for epigenetic characterization of single cells. *Nat. Commun.* **12**, 2177 (2021).
33. Chen, X. et al. EpiAgent: foundation model for single-cell epigenomics. *Nat. Methods* **22**, 2316–2327 (2025).

To the reviewers,

We sincerely thank all four reviewers for their thoughtful comments and positive feedback on our manuscript, “Dissecting epigenetic heterogeneity in single-cell DNA methylomes with a unified framework”.

In particular, regarding Reviewer #3’s concerns about the differential analysis methods employed, we have extensively reviewed the relevant literature and conducted further discussion and analysis. Based on these efforts, we have revised the Results, Methods, and Discussion sections of the manuscript to clarify the rationale, limitations, and potential improvements of the differential analysis performed in scMethCraft. All changes are highlighted in red in the revised manuscript.

Below, we provide point-by-point responses to the reviewers’ comments, with the reviewers’ points in black and our responses in blue.

REVIEWER #1 (REMARKS TO THE AUTHOR)

The authors have successfully addressed my comments, and I have no further concerns.

RESPONSE

We are thankful to the reviewer for approval of the revised manuscript. The constructive comments provided by the reviewer have helped us significantly enhance our manuscript.

REVIEWER #2 (REMARKS TO THE AUTHOR)

The revised manuscript includes a number of additional inclusions that satisfy my initial comments.

RESPONSE

We sincerely appreciate the reviewer's recognition of the revised manuscript, and we once again express our gratitude for the valuable comments provided by the reviewer, which helped us improve our manuscript.

REVIEWER #3 (REMARKS TO THE AUTHOR)

In their revised manuscript, the authors have addressed most of my initial comments. The additional benchmarking and discussion of other tools is a good addition, as is the additional integration evaluations. However, I still have concerns regarding the differential testing methods used in the manuscript, and more broadly the implied recommendation for users to follow this approach with their own data using scMethCraft.

RESPONSE

We would like to express our gratitude to the reviewer for their recognition of our work and for the valuable constructive comments. We have thoroughly reviewed the relevant literature and carefully considered and discussed the points raised.

COMMENT 1

The authors have now used Bonferroni correction, and state that it helps mitigate the inflation of significance arising from non-independence among observations. This is not accurate: Bonferroni controls the family-wise error rate and assumes independent observations, it does not mitigate inflated significance arising from non-independent observations.

Adopting Bonferroni correction addresses the issue of multiple testing, but fundamentally the differential test is statistically flawed due to the violated assumption of independent observations. This is an important point, because the results of these tests will not be reliable, and enabling this in the scMethCraft framework is concerning as it may lead to false discoveries in the field.

RESPONSE TO COMMENT 1

We thank the reviewer for the valuable comment regarding the method used to identify differentially methylated regions. In the previous round of review, the reviewer noted that "The methods also do not state that any multiple testing correction was performed on the p-values obtained." Following the reviewer's suggestion, we applied a Bonferroni correction to adjust for multiple testing in the revised manuscript. We note, however, that while Bonferroni correction controls the family-wise error rate under independent observations, it does not address inflated significance resulting from non-independence among cells. We sincerely apologize for not explicitly discussing this point in our previous response. After receiving the reviewer's additional comments, we carefully revisited the relevant literature and discussed this issue in greater depth, and we provide a more comprehensive clarification below.

First, we acknowledge that in the scMethCraft-enhanced data, a certain degree of correlation may indeed be introduced among cells, which may in a strict sense violate the independence assumption underlying the Wilcoxon rank-sum test. In fact, the question of whether the independence assumption holds is itself a common issue in single-cell sequencing data analysis. Multiple cells originating from the same biological sample often share developmental origins and biological backgrounds; therefore, even in the absence of any data augmentation or smoothing procedures, correlations may still exist among cells, as also noted by Reviewer #1. Although such correlations mean that the assumptions of this statistical test are not perfectly satisfied, previous studies have shown that the Wilcoxon rank-sum test remains one of the most widely used approaches for differential analysis in single-cell epigenomic data analysis[1]. Many widely used single-cell sequencing data analysis toolkits, including Seurat[2] (default test for FindMarkers), Scanpy[3] (sc.tl.rank_genes_groups), and EpiScanpy[4] (epi.tl.rank_features), employ the Wilcoxon rank-sum test to identify differential features between cell populations. A large number of previous single-cell studies have also applied Wilcoxon-based tests to detect differentially methylated or accessible regions in single-cell epigenomic sequencing data[5-15].

Moreover, although data imputation or enhancement may increase correlations among cells, applying the Wilcoxon rank-sum test to imputed or enhanced data for differential analysis remains a common practice in studies of sequencing data imputation methods. For example, SAVER performs the Wilcoxon rank-sum test on imputed scRNA-seq datasets to identify differentially expressed genes[16]. scGNN enhances scRNA-seq data using a graph neural network and then applies the Wilcoxon rank-sum test to the enhanced data to detect differentially expressed genes[17]. netNMF-sc ranks differentially expressed genes using the Wilcoxon rank-sum test on imputed data[18]. DeepIMB imputes microbiome data with a deep neural network and performs differential analysis using the Wilcoxon rank-sum test on the imputed data[19]. In addition, review studies have used the Wilcoxon rank-sum test to conduct comprehensive evaluations of different scRNA-seq data imputation methods[20].

Second, for scDNAm data, which are typically represented as regional methylation ratios rather than count-based measurements as in scRNA-seq or scATAC-seq data, performing rigorous statistical inference for differential analysis remains a challenging and open problem. During the revision process, we surveyed a number of widely used differential analysis methods in single-cell sequencing studies and attempted to adapt them in our manuscript. However, count-based methods that are extensively applied to scRNA-seq data, such as DESeq2[21] and edgeR[22], cannot be directly applied to scDNAm data, as the latter substantially deviates from the discrete negative binomial distribution assumption. Developing a dedicated

differential testing framework for single-cell DNA methylation data is itself a separate and technically challenging research problem, which falls beyond the scope of the present study.

Third, we would like to clarify the issue of false positives and false discoveries. Due to the inherent characteristics of single-cell sequencing technologies, such as high missingness, uneven coverage, and limited numbers of biological samples, it is difficult for any differential analysis method to completely avoid false positives in practice. Therefore, in real-world studies, the validation of differentially expressed genes or differentially methylated regions typically relies on multiple lines of evidence, including visualization, prior biological knowledge, multi-omics integration, and downstream experimental validation. In this study, the primary goal of differential region analysis is to identify candidate regions exhibiting clear cell-type-specific methylation patterns, serving as an exploratory analysis to demonstrate the ability of scMethCraft to recover underlying biological signals. Our results show that, after model-based enhancement, biological patterns in the data become more clearly distinguishable, which facilitates the identification of a greater number of regions with potential biological relevance in differential analysis. At the same time, we observe that the differentially methylated regions identified by this pipeline effectively capture methylation differences across cell types and are consistent with their known biological functions.

In summary, although the data enhanced by scMethCraft do not fully satisfy the independence assumption of the Wilcoxon rank-sum test, (1) many existing studies have adopted the Wilcoxon rank-sum test for differential analysis without explicitly accounting for potential correlations; (2) applying the Wilcoxon rank-sum test to augmented or imputed data is also a common and effective practice; (3) although some methods developed for scRNA-seq data attempt to address the lack of independence among cells, their direct application to scDNAm data is hindered by the distinct characteristics of methylation data; (4) developing a statistically rigorous hypothesis testing framework tailored for scDNAm data is beyond the scope of this study; and (5) in this work, the use of the Wilcoxon rank-sum test is primarily intended to identify candidate regions with clear cell-type-specific methylation patterns, as part of an exploratory analysis to demonstrate the ability of scMethCraft to recover underlying biological signals, rather than to provide definitive statistical significance. We have further supported this objective through multiple downstream analyses, which collectively demonstrate the potential of scMethCraft under this setting. Therefore, we hope the reviewer can appreciate our use of the Wilcoxon rank-sum test as a pragmatic and reasonable choice given the current state of the field.

We also acknowledge that the description in the original manuscript was not sufficiently rigorous. Specifically, we did not adequately discuss this statistical assumption or its potential implications, which may have led readers to misinterpret the analysis pipeline as a statistically

rigorous approach for differential region inference. Without proper clarification, readers might regard differential analysis based on augmented data as a standard statistical inference procedure and apply it directly to their own data, potentially increasing the risk of false positives. We fully understand the reviewer's concern in this regard. Accordingly, in the revised manuscript, we have further clarified the main text, explicitly stating that the differential region analysis in this study is primarily intended to prioritize candidate regions with potential cell-type-specific patterns, rather than to serve as a strict statistical inference. In other words, the goal of this analysis is to rank candidate regions associated with cell types, rather than to provide definitive statistical significance. In addition, we have incorporated corresponding clarifications, emphasizing that statistical analyses performed on scMethCraft-enhanced data should be interpreted with caution. We also recommend examining the results from multiple perspectives and performing cross-validation with the original data to obtain more robust and reliable conclusions.

We sincerely thank the reviewer again for the constructive comments. It has helped us more clearly articulate the scope of applicability of our method and its underlying statistical assumptions, thereby further improving the rigor and interpretability of the manuscript.

REVIEWER #4 (REMARKS TO THE AUTHOR)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

RESPONSE

We thank the reviewer for the co-review and acknowledge this statement.

References

1. Teo, A.Y.Y., Squair, J.W., Courtine, G. & Skinnider, M.A. Best practices for differential accessibility analysis in single-cell epigenomics. *Nat. Commun.* **15**, 8805 (2024).
2. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
3. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
4. Danese, A. et al. EpiScanpy: integrated single-cell epigenomic analysis. *Nat. Commun.* **12**, 5228 (2021).
5. Janssens, J. et al. Decoding gene regulation in the fly brain. *Nature* **601**, 630–636 (2022).
6. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142–146 (2015).
7. Sarropoulos, I. et al. Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science* **373**, eabg4696 (2021).
8. Chen, A.F. et al. NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 547–553 (2022).
9. Jin, W. et al. Semi-automated IT-scATAC-seq profiles cell-specific chromatin accessibility in differentiation and peripheral blood populations. *Nat. Commun.* **16**, 2635 (2025).
10. Corces, M.R. et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).
11. Gu, C., Liu, S., Wu, Q., Zhang, L. & Guo, F. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. *Cell Res.* **29**, 110–123 (2019).
12. You, M. et al. Single-cell epigenomic landscape of peripheral immune cells reveals establishment of trained immunity in individuals convalescing from COVID-19. *Nat. Cell Biol.* **23**, 620–630 (2021).
13. Luo, C. et al. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genomics* **2** (2022).
14. Prompsy, P. et al. Interactive analysis of single-cell epigenomic landscapes with ChromSCape. *Nat. Commun.* **11**, 5702 (2020).
15. Floc'hlay, S. et al. Shared enhancer gene regulatory networks between wound and oncogenic programs. *Elife* **12**, e81173 (2023).
16. Huang, M. et al. SAVER: gene expression recovery for single-cell RNA sequencing. *Nat. Methods* **15**, 539–542 (2018).
17. Wang, J. et al. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat. Commun.* **12**, 1882 (2021).
18. Elyanow, R., Dumitrascu, B., Engelhardt, B.E. & Raphael, B.J. netNMF-sc: leveraging gene–gene interactions for imputation and dimensionality reduction in single-cell expression analysis. *Genome Res.* **30**, 195–204 (2020).
19. Song, H., Hosain, M.M. & Park, T. DeepIMB: Imputation of non-biological zero counts in microbiome data. *Genes Genom.* **48**, 117–129 (2026).
20. Hou, W., Ji, Z., Ji, H. & Hicks, S.C. A systematic evaluation of single-cell RNA-sequencing imputation methods. *Genome Biol.* **21**, 218 (2020).

21. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
22. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).