nature communications

Article

https://doi.org/10.1038/s41467-024-51382-x

scConfluence: single-cell diagonal integration with regularized Inverse Optimal Transport on weakly connected features

Received: 11 March 2024

Accepted: 6 August 2024

Published online: 05 September 2024

Check for updates

Jules Samaran[1], Gabriel Peyré[2] & Laura Cantini[1]

The abundance of unpaired multimodal single-cell data has motivated a growing body of research into the development of diagonal integration methods. However, the state-of-the-art suffers from the loss of biological information due to feature conversion and struggles with modality-specific populations. To overcome these crucial limitations, we here introduce scConfluence, a method for single-cell diagonal integration. scConfluence combines uncoupled autoencoders on the complete set of features with regularized Inverse Optimal Transport on weakly connected features. We extensively benchmark scConfluence in several single-cell integration scenarios proving that it outperforms the state-of-the-art. We then demonstrate the biological relevance of scConfluence in three applications. We predict spatial patterns for *Scgn*, *Synpr* and *Olah* in scRNA-smFISH integration. We improve the classification of B cells and Monocytes in highly heterogeneous scRNA-scATAC-CyTOF integration. Finally, we reveal the joint contribution of *Fezf2* and apical dendrite morphology in Intra Telencephalic neurons, based on morphological images and scRNA.

In the last decade, single-cell transcriptomics (scRNA) has revolutionized our understanding of the diversity of cells constituting living tissues[1–3]. Since then, a new milestone has been reached with the introduction of high-throughput sequencing technologies allowing us to measure additional molecular modalities, such as chromatin accessibility (scATAC)[4,5] and methylation (snmC)[6], at the resolution of the single cell. More recently, technologies allowing the joint measurement of different single-cell modalities from the same cell (i.e. paired data) have been proposed[7–15]. Examples of these cutting-edge sequencing technologies are CITE-seq, simultaneously measuring RNA and surface protein abundance by leveraging oligonucleotide-conjugated antibodies[8], and 10X Genomics Multiome platform, quantifying RNA and chromatin accessibility by microdroplet-based isolation of single nuclei.

Different single-cell modalities describe complementary facets of the cell; their joint analysis is thus expected to provide tremendous power to uncover cellular identities[16]. For achieving this aim, paired single-cell multimodal data represent an ideal resource[17,18] and numerous methods have been designed for their integration[19–22]. Nevertheless, paired data are still rare and limited in the number of modalities that they contain (maximum three)[23]. Single-cell multimodal data profiled from different cells of the same biological condition, i.e. unpaired data, thus represent a precious resource for accessing different molecular facets of a cell and better understanding its identity.

The integration of unpaired single-cell multimodal data, i.e. diagonal integration, is more challenging than paired integration[24]. Indeed, comparing cells from different modalities is not straightforward, as they are described by different features (e.g. genes, peaks, proteins). The aim of diagonal integration is to define a low-dimensional latent space shared by all modalities. In this shared latent space, cells should be arranged according to their biological similarity, independently from their modality of origin. Providing such a biologically meaningful modality alignment of cells, different from

[1]Institut Pasteur, Université Paris Cité, CNRS UMR 3738, Machine Learning for Integrative Genomics Group, Paris, France. [2]CNRS and DMA de l'Ecole Normale Supérieure, CNRS, Ecole Normale Supérieure, Université PSL, Paris, France. e-mail: laura.cantini@pasteur.fr

the many potential artificial alignments that overlap cells from different cell types, is extremely challenging.

To guide cell alignment between modalities in the shared latent space, diagonal integration leverages prior biological information[24]. Indeed, connections between the features of different modalities are generally known in biology. For instance, chromatin peaks can be mapped to genes based on their proximity to gene promoter regions, thus enabling the computation of gene activity measurements[25,26]. Similarly, protein-coding genes and their corresponding proteins can be used as connections between scRNA-seq and proteomic data. Most of the state-of-the-art methods use this prior biological knowledge to convert all modalities to the same features and then handle the alignment similarly to batch effect correction[27–29]. However, this conversion can result in an important loss of biological information as features across modalities are weakly connected. Indeed, across-modality feature connections are often rare and noisy. For example, protein-coding genes are a subset of all the expressed genes, and not all possible chromatin peaks are close to the promoter of a gene. This problem becomes even more challenging once the features measured in one modality are few due to technological limitations (e.g. targeted CyTOF providing only a few proteins quantified across cells). State-of-the-art methods not requiring modality conversions also exist[30,31]. However, they still depend on the assumption that most features can be reliably connected across modalities. In addition, many state-of-the-art methods[27,28,30,31] ignore the possibility that a population of cells (cell type/state) can be present only in one modality, which is frequently the case for unpaired data.

Here, we propose scConfluence, a diagonal integration method combining uncoupled autoencoders, which reduce the dimensionality of the original data to a shared latent space and account for potential batch effects, together with regularized Inverse Optimal Transport (rIOT)[32], which aligns cells across modalities in the shared latent space by leveraging weakly connected features. By employing rIOT to ensure modality alignment, scConfluence can independently process the complete set of original features through autoencoders while utilizing only the connected features for aligning cell embeddings. Therefore, our approach does not suffer from the loss of biological information generally resulting from modality conversion prior to dimension reduction. In addition, thanks to the unbalanced relaxation of Optimal Transport[33], scConfluence can also deal with cell types absent in a modality thus overcoming all the major limitations of the state-of-the-art.

We extensively benchmark scConfluence with respect to the state-of-the-art in several scRNA-surface protein and scRNA-scATAC integration problems. This in-depth comparison proves that scConfluence's embeddings outperform the state-of-the-art across a wide variety of datasets. We further demonstrate scConfluence's robustness, accuracy, and general applicability in addressing three diverse and crucial biological questions. First, we integrate scRNA-seq and smFISH profiled from mouse somatosensory cortex and predict *Scgn*, *Synpr*, and *Olah* to have spatial patterns of expression amenable for further biological investigation. Second, scConfluence's integration of scRNA-seq, scATAC-seq, and CyTOF improves the classification of B cells and Monocytes in highly heterogeneous human PBMC datasets. Finally, scConfluence integrates neuronal morphological images with scRNA-seq from the mouse primary motor cortex revealing the joint contribution of the Transcription Factor *Fezf2* and apical dendrite morphology to information processing in Intra Telencephalic neurons.

scConfluence is highly modular, allowing its generalization to the new integration scenarios that will arise as a consequence of the continuous single-cell technological developments (e.g. single-cell metabolomics). scConfluence is implemented as an extensively documented open-source Python package seamlessly integrated within the scverse ecosystem[34] and is available at https://github.com/cantinilab/scconfluence.

Results

scConfluence a new method for diagonal single-cell multimodal integration

We developed scConfluence, a method for diagonal integration combining uncoupled autoencoders with regularized Inverse Optimal Transport (rIOT) on weakly connected features.

As shown in Fig. 1a, the inputs of scConfluence are single-cell data from M modalities represented by the matrices $\mathbf{X}^{(p)} \in \mathbb{R}^{n^{(p)} \times d^{(p)}}$ with $p \in [1..M]$, where rows correspond to cells and columns to features (e.g. genes, chromatin peaks, proteins). The cells of $\mathbf{X}^{(p)}$ can come from multiple experimental batches. As discussed in the Introduction, although each modality is grounded in a different feature space, across-modality connections between some features can be defined based on prior biological knowledge. Therefore, we expect that for all pairs of modalities (p,p'), we have access to $\mathbf{Y}^{(p,p')} \in \mathbb{R}^{n^{(p)} \times d^{(p,p')}}$ and $\mathbf{Y}^{(p',p)} \in \mathbb{R}^{n^{(p')} \times d^{(p,p')}}$, conversions of $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(p')}$ to common features, respectively. For example, if p corresponds to scRNA and p' is scATAC, $\mathbf{Y}^{(p,p')}$ and $\mathbf{Y}^{(p',p)}$ correspond to the RNA count matrix and the gene activity matrix derived from peak accessibility counts, respectively.

scConfluence makes use of both the original data $\mathbf{X}^{(p)}$ and the converted data $\mathbf{Y}^{(p,p')}$ to learn low-dimensional cell embeddings $\mathbf{Z}^{(p)} \in \mathbb{R}^{n^{(p)} \times d_z}$ in a shared latent space of dimension d_z. These embeddings can then be used for visualization and clustering, useful for discovering subpopulations of cells, and for imputation of features across modalities (Fig. 1b).

For each modality p, scConfluence trains an autoencoder $\mathrm{AE}^{(p)}$ on $\mathbf{X}^{(p)}$ using modality-specific architectures[35] and reconstruction losses $\mathcal{L}_{AE^{(p)}}$ in order to retain all the complementary information brought by each modality. $\mathrm{AE}^{(p)}$ also performs batch correction by learning cell embeddings independent from their experimental batches of origin (see "Methods" section). While frameworks based on autoencoders have been already designed in the context of diagonal integration[29,31,36], the innovation of scConfluence is the combined use of Optimal Transport and regularized Inverse Optimal Transport (rIOT) for aligning cells in the shared latent space. Optimal transport (OT) is a mathematical toolkit for comparing high-dimensional point clouds[37] that is gaining traction for addressing various problems in single-cell genomics: single-cell multi-omics cell matching[38,39], paired multi-omics integration[20,39], trajectory inference[39–42] and predicting single-cell perturbation responses[43]. Solving the OT problem produces a correspondence map, i.e. transport plan, between point clouds based on their relative positions (see "Methods" section). rIOT aims at addressing the inverse problem by inferring the relative positions of points based on a given transport plan[32]. scConfluence makes innovative use of both OT and rIOT by first solving an OT problem leveraging weakly connected features ($\mathbf{Y}^{(p,p')}$ and $\mathbf{Y}^{(p',p)}$) to find a transport plan $\mathbf{P}^{(p,p')}$ across modalities and then using rIOT on $\mathbf{P}^{(p,p')}$ to adjust the cell embeddings inferred by $\mathrm{AE}^{(p)}$ and $\mathrm{AE}^{(p')}$.

In more detail, we first use $\mathbf{Y}^{(p,p')}$ and $\mathbf{Y}^{(p',p)}$ to compute a distance matrix between cells from different modalities which we then leverage to find an unbalanced Optimal Transport plan $\mathbf{P}^{(p,p')} \in \mathbb{R}_+^{n^{(p)} \times n^{(p')}}$ (see "Methods" section for a definition of Unbalanced Optimal Transport). $\mathbf{P}^{(p,p')}$ provides a partial correspondence map between cells of modalities p and p' which we aim to leverage to determine the relative positions of cell embeddings in the shared latent space. This specific goal corresponds to the rIOT problem that we described above. In scConfluence, this is achieved by minimizing the loss $\mathcal{L}_{IOT}^{(p,p')}$ which penalizes distances between rows of $\mathbf{Z}^{(p)}$ and $\mathbf{Z}^{(p')}$ which are coupled by $\mathbf{P}^{(p,p')}$. See "Methods" section for a more formal explanation of the connection between our approach and rIOT. While $\mathcal{L}_{IOT}^{(p,p')}$ leverages biological prior knowledge to attract corresponding cells across modalities, it is not always sufficient to completely overlap them in the shared latent space. To address this, we add to the loss, as a regularization term, the unbalanced Sinkhorn divergence[44] between the cell embeddings of each pair of modalities ($\mathcal{L}_{reg}^{(p,p')}$). $\mathcal{L}_{reg}^{(p,p')}$, based on OT, is

https://doi.org/10.1038/s41467-024-51382-x

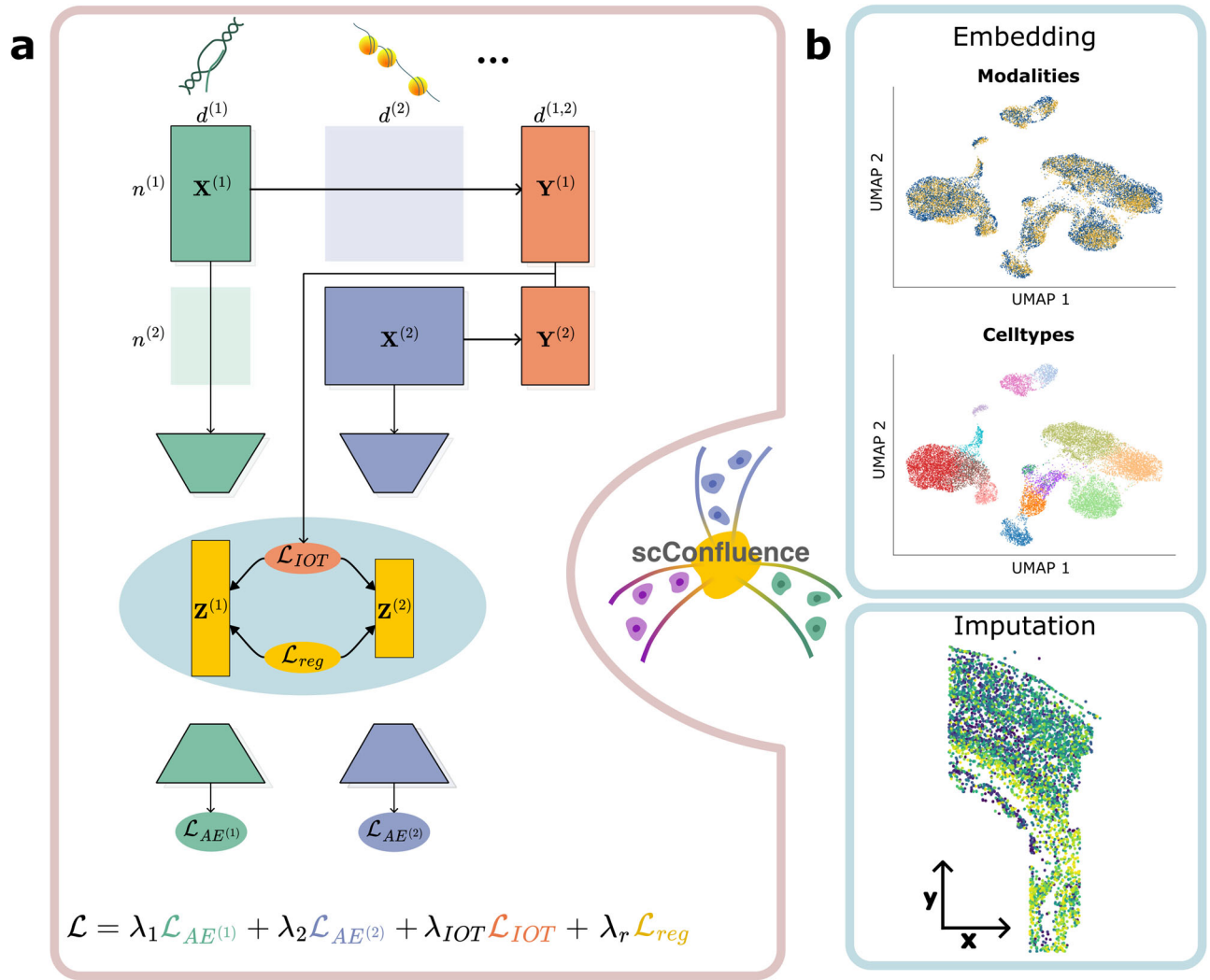

Fig. 1 | The scConfluence framework for diagonal integration. a Schematic representation of the framework simplified to only two modalities ($M = 2$). While the original data matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are inputted to their respective auto-encoders, converted feature matrices $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ (shorter notations for $\mathbf{Y}^{(1,2)}$ and $\mathbf{Y}^{(2,1)}$) are used to compute an Optimal Transport plan across the two modalities. The IOT loss $\mathcal{L}_{IOT}$ computed thanks to the transport plan and the regularization loss $\mathcal{L}_{reg}$ constituting together the rIOT constraint, are used to enforce the alignment of modalities in the shared latent space. **b** Examples of two outputs of scConfluence: cell embeddings can be visualized using 2D projections and clustered to discover new cell subpopulations, they can also be used to impute features across modalities.

frequently used in machine learning to minimize the distance between high-dimensional point clouds (see "Methods" section). The gradients of both $\mathcal{L}_{IOT}^{(p,p')}$ and $\mathcal{L}_{reg}^{(p,p')}$ are back-propagated through the modality encoders in order to improve the across-modality alignment of cell embeddings. In addition, in both $\mathcal{L}_{IOT}^{(p,p')}$ (using an unbalanced transport plan) and $\mathcal{L}_{reg}^{(p,p')}$ (i.e. unbalanced Sinkhorn divergence), Unbalanced Optimal Transport achieves a tradeoff between aligning all cells (as in regular OT) and avoiding artificial alignments for cells that have no suitable match in the other modality. Therefore, scConfluence is able to deal with cell populations present only in one modality.

The final loss optimized over the parameters of the $\mathrm{AE}^{(p)}$ with stochastic gradient descent is thus:

$$\mathcal{L} = \sum_{p=1}^{M} \lambda_p \mathcal{L}_{\mathrm{AE}(p)} + \sum_{1 \le p < p' \le M} \lambda_{IOT} \mathcal{L}_{IOT}^{(p,p')} + \lambda_r \mathcal{L}_{reg}^{(p,p')} \quad (1)$$

scConfluence separately uses all original features for dimensionality reduction in order to retain all the complementary information brought by each modality and leverages common information in the form of connected features to align cells with rIOT. Therefore, our innovative combined use of OT and rIOT allows scConfluence to avoid the loss of biological information generally resulting from modality conversion in state-of-the-art methods. As a consequence, scConfluence is much more robust to integration problems where very few features are connected across modalities (e.g. scRNA-surface protein data integration). In addition, the quality of scConfluence's modality alignment depends on the transport plan $\mathbf{P}^{(p,p')}$ which relies only on the relative distances derived from the converted data $\mathbf{Y}^{(p,p')}$ and $\mathbf{Y}^{(p',p)}$. As a consequence, scConfluence can better deal with situations where strong batch effects between modalities are present in the converted data space. Furthermore, while state-of-the-art methods strictly enforce the complete mixing of cells across modalities, scConfluence, through the use of unbalanced OT, can cope with large discrepancies between the cell populations present in each modality. scConfluence is thus able to integrate single-cell modalities even when they do not contain the same cell types.

We extensively benchmarked scConfluence against five state-of-the-art methods: Seurat (v3.0), Liger, MultiMAP, Uniport, and scGLUE[27–31]. Seurat, Liger, and MultiMAP are widely used single-cell unpaired multi-omics integration methods in the computational

biology community. Uniport is the main alternative to our method also using OT. Finally, scGLUE is the most recent and best-performing method in the NeurIPS challenge on Open Problems in Single-Cell Analysis[45].

scConfluence outperforms the state-of-the-art on the integration of unbalanced cell populations

One of the main challenges of diagonal single-cell multi-omics integration is the need to deal with unbalanced cell populations. This requires aligning shared cell populations, independently of their size, and preserving modality-specific ones. We thus benchmarked scConfluence with the state-of-the-art based on its ability to integrate single-cell modalities sharing only a fraction of cell populations. As using simulated data based on distributional assumptions would favor methods making the same assumptions, we here designed a benchmark using scCATseq data profiled from HeLa, HCT, and K562 cancer cell lines[15]. The choice of these data comes from the need to work with well-separated clusters, for which cell lines are an ideal example. In addition, having an equivalent proportion of cells per cluster in the two modalities allows us to design scenarios with different levels of unbalanceness in the cell populations. Of note, while scCATseq provides a joint profiling of scRNA and scATAC from exactly the same cell, the cell pairing information has not been used here as input of the various methods. To then test to which extent unbalanced cell populations affect the results of diagonal integration we modified the scCATseq data to represent three realistic situations: (i) removing half of K562 scRNA cells; (ii) removing all K562 scRNA cells and (iii) removing completely K562 scRNA cells and HCT scATAC cells. See Fig. 2a for a schematic representation.

A successful integration method should: (i) produce biologically meaningful integrated cell embeddings, i.e. organizing cells according to cell types and states, and (ii) align cells profiled from different modalities (e.g. scRNA, scATAC) that are paired or at least from the same cell type/state We used the purity score[20] to evaluate (i). For (ii), we used three scores: Fraction Of Samples Closer Than the True Match (FOSCTTM) (modified from refs. 31,38,46 as explained in "Methods" section), to evaluate the closeness of paired cells, connectivity[47], to assess whether cells from the same cell type are close to each other independently of their modality of origin, and transfer accuracy[48], to measure the proximity between corresponding cell types across modalities in the shared latent space (see "Methods" section for all scores mentioned). Concerning MultiMAP, its output used for downstream analyses is a neighborhood cell graph only encoding closest interactions. This link thresholding in the neighborhood cell graph results in artificially low performances with FOSCTTM. For this reason, FOSCTTM was not reported for MultiMAP.

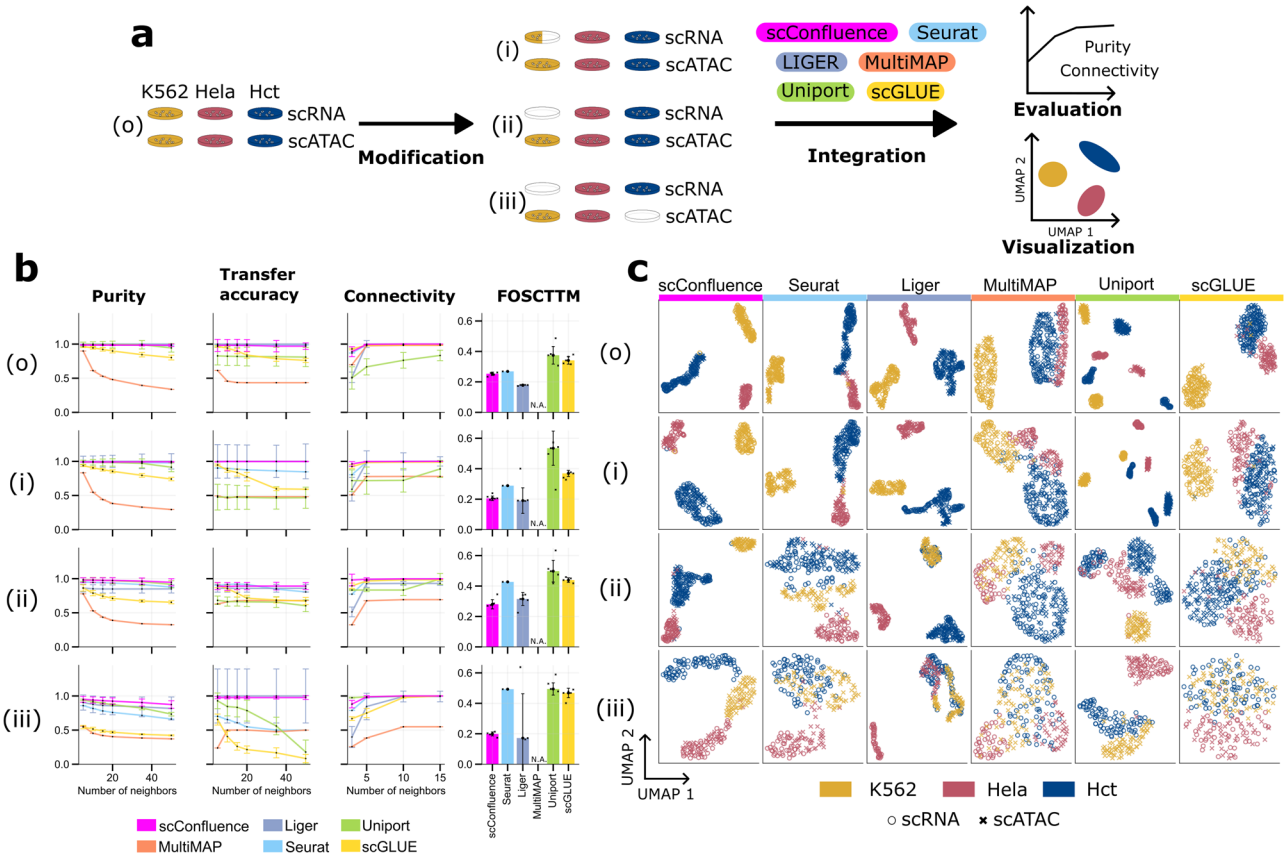

Fig. 2 | Benchmarking cell embeddings in unbalanced cell lines. a Schematic representation of the benchmarking process. Four scenarios are here considered: removing half of K562 scRNA cells, removing all K562 scRNA cells, and removing completely K562 scRNA cells and HCT scATAC cells; **b** Purity, Transfer accuracy, Connectivity, and Fraction Of Samples Closer Than the True Match (FOSCTTM) scores are here reported for the six benchmarked methods (scConfluence, Seurat, Liger, MultiMAP, Uniport, and scGLUE) on the four controlled settings derived from the cell lines data as described in (**a**). Since purity, transfer accuracy, and connectivity scores are based on nearest neighbors graphs, the plots report their behavior for various sizes of neighborhood (x-axis). Error bars in the plots specify the standard deviation across $n = 5$ random initialization seeds for each method and they are centered on the median result. Inside bar plots, small dark stars represent individual seed results. Source data are provided as a Source Data file; **c** The six columns of this panel provide UMAP visualizations for the six benchmarked methods (scConfluence, Seurat, Liger, MultiMAP, Uniport, and scGLUE) on the same four controlled settings derived from the cell lines data. Different colors in these UMAP plots correspond to the three different cell lines present in the data while the shape of the point markers corresponds to the modality of origin of each cell (scRNA, scATAC).

As expected, all methods showed decreasing performances when the scenarios became less balanced. scConfluence outperformed the state-of-the-art in all scenarios, proving more robustness to variabilities in cell populations' proportions (see Fig. 2b, c). For the remaining methods, MultiMAP and scGLUE struggled the most to group cells based on their cell line of origin, while MultiMAP and Uniport were less performant in mixing modalities from shared populations. This can be observed also in the UMAP plots (Fig. 2c). Regarding LIGER, the results here displayed concern its performances once setting the number of latent dimensions to three. This choice particularly advantages the method that fails to integrate the two modalities for other values of the latent dimensions (see Supplementary Fig. 1). In addition, even when using three latent dimensions, LIGER displays high variability in every score across different runs in most scenarios.

scConfluence outperforms the state-of-the-art in scRNA-surface protein and scRNA-scATAC integration

To then benchmark scConfluence vs the state-of-the-art on larger and more realistic diagonal integration scenarios, we considered two 10X Genomics Multiome (scRNA+scATAC) datasets: (i) *PBMC 10X*, a human PBMC dataset with 9378 cells per modality (ii) *OP Multiome*, a human bone marrow dataset, with 69,249 cells per modality profiled from different sites and donors constituting a total of 13 batches[49]; plus two CITE-seq (scRNA+surface protein) datasets: (i) *BMCITE*, a human bone marrow dataset with 30,672 cells per modality where 23 surface protein levels were measured[27] (ii) *OP Cite*, a human bone marrow dataset with 90,261 cells per modality profiled from different sites and donors constituting a total of 12 batches and with 134 surface proteins[49]. These are gold-standard datasets in multi-omics integration, already used to benchmark state-of-the-art methods[27,31,45,49]. We chose paired multi-omics data to test diagonal integration in order to have ground-truth matching between cells, useful for evaluating the performances of the various methods. Of note, the data have been treated as unpaired by the various methods and the cell pairing information has only been used for performance evaluation. In addition, the data are provided with high-quality cell labels useful for performance evaluation. For details on the data see Supplementary Table 1 and for their pre-processing see "Methods" section.

The benchmarking of performances is based on the same scores described in the previous section. In addition, to assess the ability of the methods to capture additional substructure inside cell types we added the "cell type FOSCTTM" score (i.e. FOSCTTM considering only cells from the same cell type), which penalizes the random alignment of cells from the same cell type across modalities (see "Methods" section). This score was not relevant in the context of cell lines as no additional heterogeneity is expected.

Regarding scRNA-scATAC integration (Fig. 3b), scConfluence is the best-performing method, in three out of four evaluation scores (Purity, Transfer accuracy, and connectivity). Concerning FOSCTTM, scGLUE has the best performances, immediately followed by scConfluence and Uniport. All methods perform better on *PBMC 10X* than *OP Multiome*. This is not surprising as *OP Multiome* contains more cell populations and strong batch effects, corresponding to several donors and sequencing sites. Of note, on this dataset, scConfluence performs best for batch correction. Indeed, the graph connectivity score

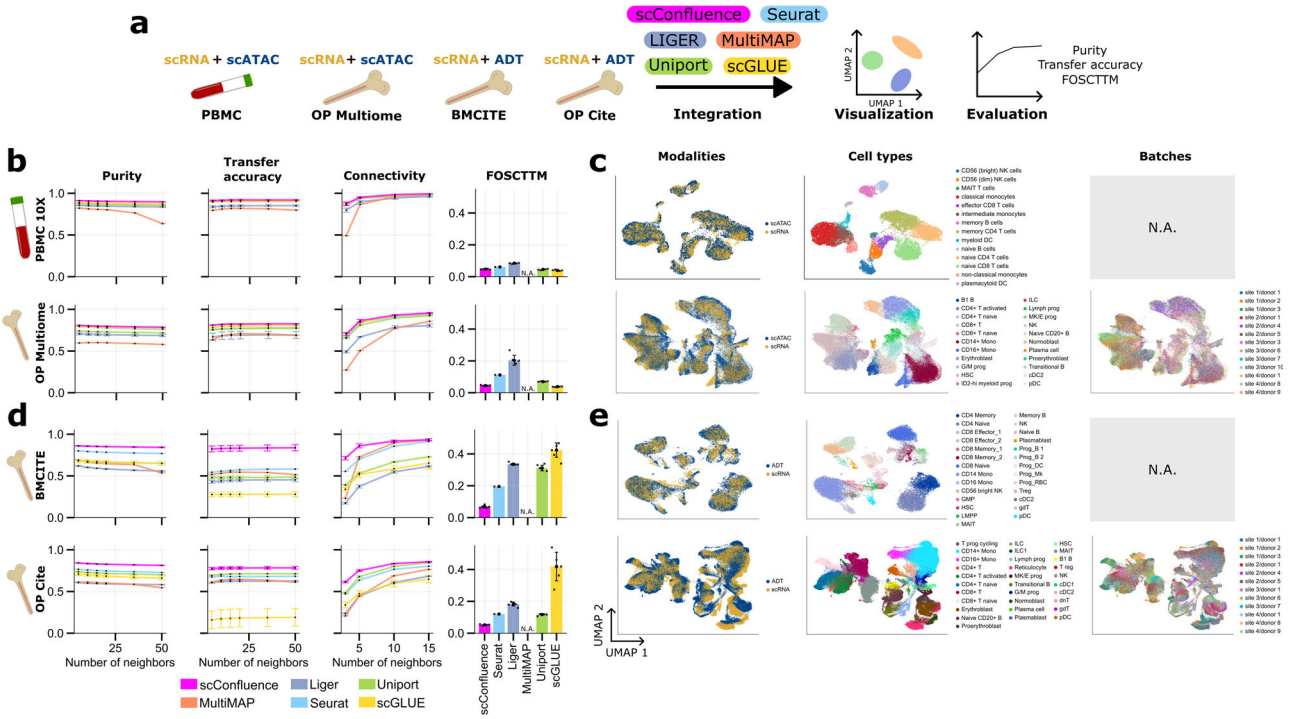

Fig. 3 | Cell embedding benchmark in gold-standard scRNA-surface protein and scRNA-scATAC datasets. a Schematic representation of the benchmarking process; **b** Purity, Transfer accuracy, Connectivity, and FOSCTTM scores for the six benchmarked methods (scConfluence, Seurat, Liger, MultiMAP, Uniport, and scGLUE) in two scRNA-scATAC datasets profiled from PBMC and bone marrow. Error bars in the plots specify the standard deviation across $n = 5$ random initialization seeds for each method and they are centered on the median result. Inside bar plots, small dark stars represent individual seed results. Source data are provided as a Source Data file; **c** UMAP visualizations of scConfluence's cell embeddings in the same datasets as (**b**). Cells are colored based on their modality of origin, their cell type annotation, or their batch of origin (when multiple batches are present in the data), respectively; **d** Same scores and methods as (**b**), but computed on the two scRNA-surface protein datasets of the benchmark profiled from bone marrow. Error bars in the plots specify the standard deviation across $n = 5$ random initialization seeds for each method and they are centered on the median result. Inside bar plots, small dark stars represent individual seed results. Source data are provided as a Source Data file.; **e** UMAP visualizations of scConfluence's cell embeddings on the two scRNA-surface protein datasets with cells colored according to the same rules as (**c**).

captures both the mixing of cells from different modalities and different batches. Overall, for scRNA-scATAC integration, scConfluence is the method achieving the best compromise between producing a biologically meaningful integrated cell embedding and aligning cells profiled from scRNA and scATAC. In Fig. 3c, UMAP visualizations illustrate the quality of the integration results obtained by scConfluence, with respect to the mixing of the modalities, the correction of batch effects, and the alignment of annotated cell types. For all other methods see Supplementary Figs. 2 and 3.

In scRNA and surface protein integration (Fig. 3d), scConfluence largely outperformed the state-of-the-art based on all four metrics on both datasets. On *BMCITE*, the relative improvement of scConfluence with respect to the second best is 9% in purity, 45% in transfer accuracy (corresponding to over 30% of the cells better classified by our method), and 66% in FOSCTTM. The performance gap is smaller on *OP Cite*, but still sizable with a relative improvement of 10% in purity, 10% in transfer accuracy (corresponding to over 5% of the cells better classified by our method), and 50% in FOSCTTM. The observed gap can be explained by the need of state-of-the-art methods for a large number of connections between the features of different modalities. This is not the case when integrating scRNA and surface protein data. For instance, in *BMCITE*, only 23 features are connected between the two modalities. As a consequence, most state-of-the-art methods have to subset the scRNA features to 23 protein-coding genes, thus discarding most of the information contained in the data. Moreover, scGLUE also struggles to align modalities since its prior feature graph contains thousands of nodes but only 23 edges. In addition, on *OP CITE*, scConfluence performs best for batch correction based on graph connectivity.

The quality of our integration is highlighted by the UMAP visualizations in Fig. 3e. While on *BMCITE* the modalities are completely mixed, on *OP Cite* a non-perfect mixing can be observed for a few cell types/states (e.g. reticulocytes, erythroblasts, and lymphoid progenitors). However, the integration of *OP Cite* data is a particularly challenging task, where a good tradeoff needs to be found between overlapping cells from different data modalities, correcting batch effects in each modality, and defining a biologically meaningful integrated cell embedding (i.e. organizing cells according to cell types and states). Based on the evaluation in Fig. 3d, scConfluence is the method achieving the best tradeoff. All other state-of-the-art methods suffer more in at least one of these objectives (Supplementary Figs. 4 and 5). For instance, LIGER completely overlaps the two modalities but provides integrated cell embeddings less biologically coherent than scConfluence.

Finally, scConfluence, as most of the methods, proves the ability to capture substructure inside cell types (see Supplementary Fig. 6) with cell type FOSCTTM scores significantly below the baseline of 0.5. This proves that embedding methods can highlight cellular heterogeneity at a finer resolution than cell types.

scConfluence robustly integrates scRNA and smFISH from the mouse cortex, predicting genes with relevant spatial patterns

The phenotypic behavior of a cell, i.e. the cell state, results from the joint activity of the molecular regulation inside the cell and the influence of neighboring cells. Working with gene expression across space (e.g. in tissue context) is thus crucial to better characterize cell states. However, the possibility to jointly measure at single-cell and high-throughput resolution both spatial position and gene expression is still rare[50]. At the same time, other existing data have important limitations. On one hand, spatial high-plex imaging data (e.g. smFISH[51–53], starMAP[54]) are limited by the possibility of only measuring a few genes (~100–1000 genes)[55]. On the other hand, scRNA sequencing allows to sequence the full transcriptome but breaks tissues apart thus losing the spatial information[1]. Integrating these two types of data is thus the best opportunity we have to shed light on the role of spatial context in cell state definition.

With this aim, we applied scConfluence to integrate two gold-standard datasets profiled from the mouse somatosensory cortex: (i) smFISH data of 33 selected marker genes measured in 4530 cells[56]; (ii) Smartseq2 data of ~20k genes (including the 33 of the previous dataset) measured across 3005 cells[57]. As shown in Fig. 4a, two outputs of scConfluence have been considered: (i) cell embeddings, whose quality is evaluated based on the same criteria used above (except for FOSCTTM since the data is unpaired) and (ii) imputations of the expression levels of unmeasured genes in the smFISH experiment. scConfluence's results are here compared with the same state-of-the-art methods as before, with the only addition of GimVI[58] which was specifically designed for scRNA and spatial high-plex imaging data.

Regarding the quality of cell embeddings, scConfluence outperforms all state-of-the-art methods according to cell type purity, transfer accuracy and graph connectivity (Fig. 4b, c, Supplementary Fig. 7). Thus, scConfluence proved again the ability to leverage a small number of common features to perform diagonal integration. Regarding the smFISH imputations, scConfluence enables us to predict features across modalities by connecting the smFISH encoder with the scRNA decoder. Indeed, the scRNA decoder can take as input a cell embedding from any modality and output its estimated scRNA profile. To evaluate the quality of the imputations, as done in[58], we created multiple scenarios holding out ~10% of the smFISH genes (see "Methods" section). The proximity between the imputed and the ground-truth smFISH measurements was then calculated based on average and median Spearman correlations (aSCC and mSCC), as in ref. 29. The Spearman correlation is a natural choice for this task[29,58] since it is less sensitive to outliers and focuses on the monotonic relationship (not necessarily linear) between pairs of observations. This is particularly relevant since we are interested in rewarding imputations that reflect the ground-truth's pattern of expression rather than its absolute values. As shown in Fig. 4d, gene imputation is very challenging, as aSCC and mSCC values are relatively low even for the most performant methods (median score around 0.1–0.2). Overall, according to both mSCC and aSCC, scConfluence is among the best-performing methods. In Fig. 4e, the quality of the imputations of scConfluence can be assessed also visually for the genes *Sox10*, *Kcnip2*, *Plp1* (all other genes are available in Supplementary Fig. 8). The results suggest that scConfluence captures the major patterns of spatial variation in the ground-truth. In particular, *Sox10*, and *Plp1* exhibit higher expression in oligodendrocytes, see ref. 56 for brain region annotation. *Kcnip2* displays higher expression in pyramidal neurons (both in the hippocampus and upper layers) and in inhibitory neurons from the caudoputamen. scConfluence can infer expression in those regions but fails to capture the layered structure in pyramidal neurons. Nevertheless, in comparison, GimVI which was designed specifically for imputation fails to impute major patterns of spatial variation as shown in Supplementary Fig. 9.

In addition, for the genes measured in scRNA but not in the smFISH data, scConfluence predicted some interesting spatial patterns (Fig. 4F). In particular, for *Pnoc*, *Hapln2*, and *Cux2*, known markers of inhibitory neurons, oligodendrocytes, and upper neuronal layers respectively, scConfluence imputed smFISH profiles coherent with existing studies[57,59]. Finally, scConfluence also suggests additional genes having interesting spatial patterns: *Scgn*, highly expressed in the excitatory neurons from layers 4 and 6, *Synpr*, highly expressed in the region corresponding to the caudoputamen, and *Olah*, highly expressed in hippocampal and layer 6 neurons. These last results prove the ability of scConfluence to provide new relevant biological hypotheses to be followed up experimentally.

https://doi.org/10.1038/s41467-024-51382-x

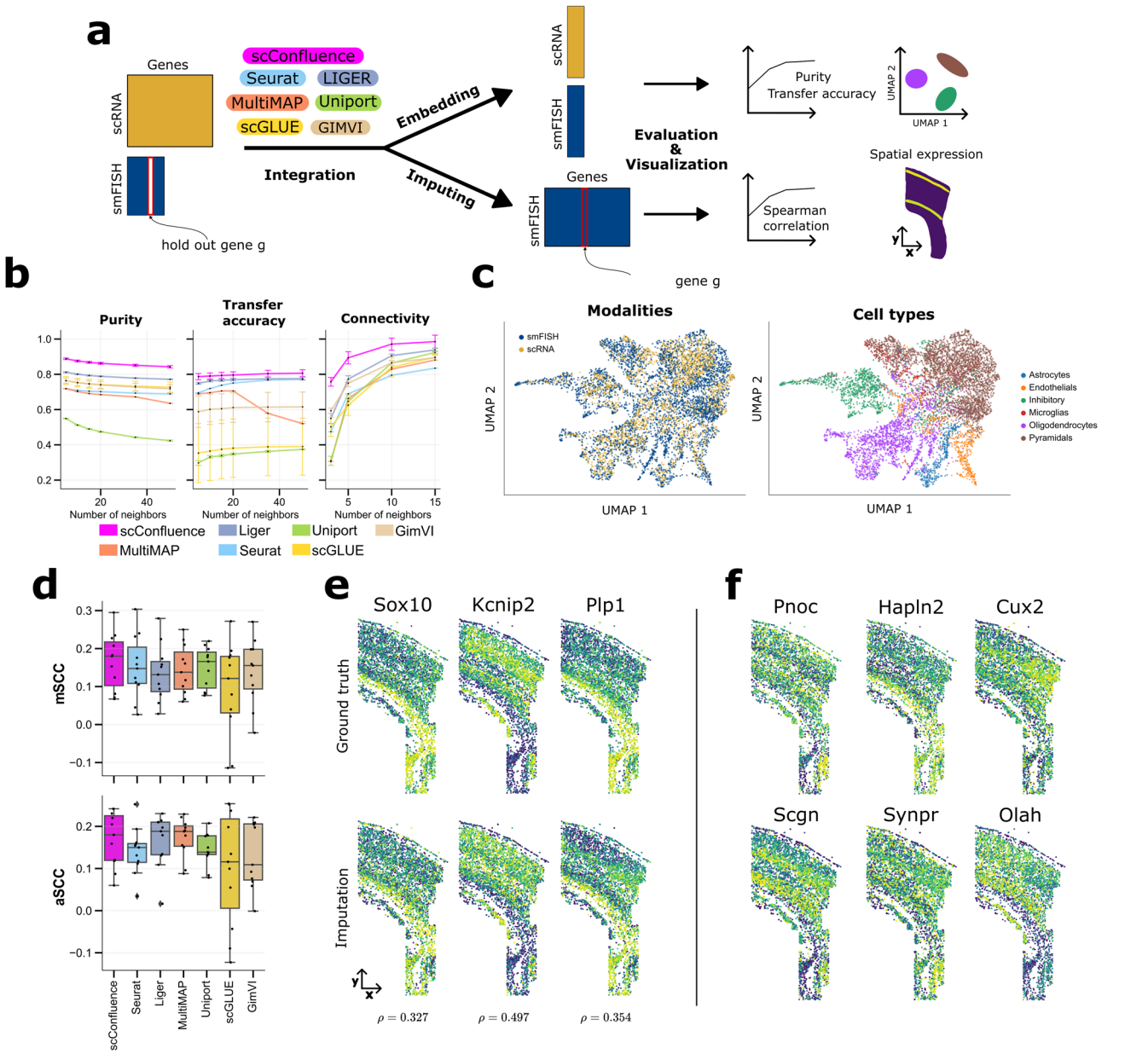

Fig. 4 | Cell embeddings and gene imputations resulting from scRNA and smFISH integration in mouse somatosensory cortex. a Schematic representation of the integration and imputation process; **b** Purity, Transfer accuracy, and Connectivity scores of the seven benchmarked methods (scConfluence, Seurat, Liger, MultiMAP, Uniport, and scGLUE, GimVI). Error bars in the plots specify the standard deviation across $n = 5$ random initialization seeds for each method and they are centered on the median result. Inside bar plots, small dark stars represent individual seed results. Source data are provided as a Source Data file; **c** UMAP visualizations of scConfluence's cell embeddings colored by the modalities of origin and their cell type annotations; **d** Boxplots of average and median Spearman correlation coefficients (aSCC and mSCC) between real and imputed smFISH genes across $n = 11$ imputation scenarios (no statistical method was used to predetermine sample size). In the boxplots, the center line, box limits, and whiskers denote the median, upper and lower quartiles, and 1.5× interquartile range, respectively. Black dots over the boxplots correspond to individual data points. Source data are provided as a Source Data file; **e** Spatial pattern of expression of scConfluence's imputations (bottom) on three held-out smFISH genes and their ground-truth pattern of expression (top). Spearman correlations between the ground-truth and imputed counts are written at the bottom. **f** scConfluence's imputed spatial pattern of expression of six scRNA genes not measured in the smFISH experiment.

scConfluence integrates highly heterogeneous scRNA, scATAC, and cyTOF leveraging their complementarity to improve cell type identification in PBMCs

A crucial challenge in biology is to take advantage of the complementarity between different data modalities to achieve a better understanding of cellular heterogeneity. While this is easier to achieve when the data are profiled from the same set of cells (e.g. 10X Multiome, CITE-seq), it becomes more challenging on unpaired data. Here, we bring this challenge to its extreme by performing diagonal integration of three PBMC single-cell omics data profiled from different cells, different donors, and by different laboratories. The aim is to test to which extent scConfluence takes advantage of the complementarity between different data modalities despite the significant across-dataset variations.

We thus applied scConfluence to the diagonal integration of three human PBMC datasets extracted in highly heterogeneous settings: (i) Seq-Well-based scRNA-seq dataset of 16,627 cells[60]; (ii) 10X Genomics scATAC-seq (Chromium platform) dataset of 21,261 cells[61] and (iii)

single-cell resolution mass cytometry (Helios CyTOF system) dataset where 48 proteins were measured in 43,232 cells[62]. This configuration is particularly challenging for diagonal integration as in most real applications the different modalities would have been extracted from a single group of donors in comparable conditions, a situation characterized by much lower biological and technical variations.

For each of the three datasets, cell type annotations were provided in their original publication. Strong discrepancies could be observed in the depth of annotation of most of the cell types. For example, B cells in scATAC are divided into naive, memory, and plasma; in CyTOF instead, they are divided into naive, memory, and double negative, and in scRNA they are merged in a single B cell population. In addition, some cell types were modality-specific, for example, MAIT T cells for CyTOF, and plasma cells for scATAC data. Such discrepancies might be due to the absence of such cell types in some modalities, to their misclassification, or to differences in annotation depth in the original studies.

scConfluence successfully integrated all three modalities in a common latent space where cells were organized according to cell types and states independently from their modality of origin (see Fig. 5b–e). Indeed, as it can be already observed from the UMAP of the three omics integration (Fig. 5b–d), cells from different modalities, and corresponding to the same cell type annotation overlap in the latent space. In addition, once clustering cells in the integrated latent space (Fig. 5f), the obtained clusters are consistent with the annotations of each modality (see Fig. 5g–i). However, our integrative analysis also provides additional information (Fig. 5f–i). The cells annotated as B cells in scRNA are split into three clusters from the three omics integration (Fig. 5g, clusters: 0, 1, 2). In scATAC (Fig. 5h), these three clusters correspond to cells annotated as memory, naive, and plasma B

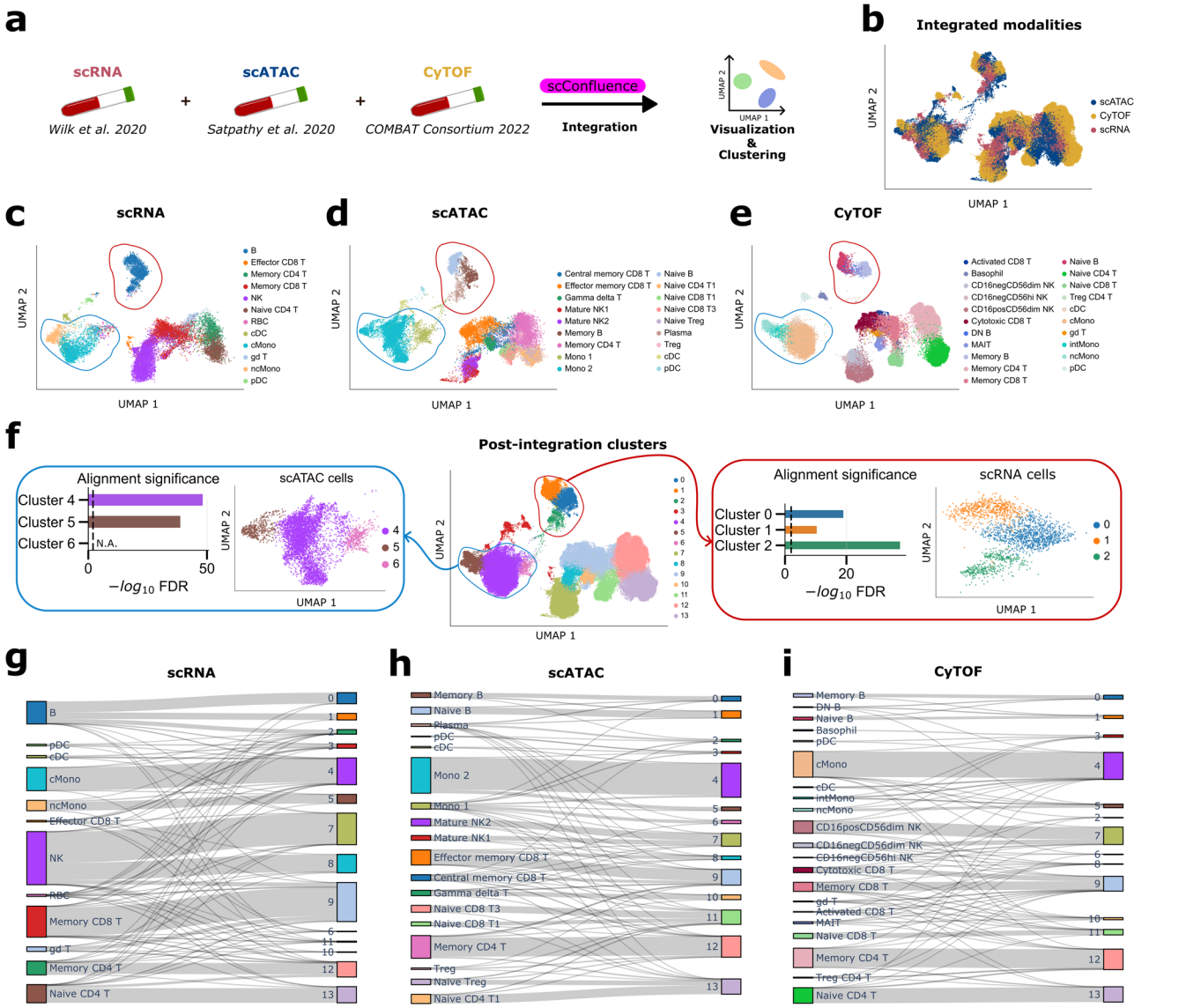

Fig. 5 | Tri-omics integration and subclustering of PBMC data. a Schematic representation of the integration; **b** UMAP visualization of all the integrated cell embeddings colored by their modality of origin; **c**–**e** UMAP visualization of scConfluence's integrated cell embeddings plotted one modality at a time and colored by their cell type annotation of origin. The red circles highlight B cells which are already sub-annotated in scATAC and CyTOF. The blue circles highlight monocytes that are already sub-annotated in scRNA and CyTOF; **f** UMAP visualization of all the integrated cell embeddings colored based on inferred cluster annotations. Additional plots are provided for ATAC monocytes and RNA B cells which have been subclustered. The significance of the overlap between the marker genes obtained from scRNA and scATAC for each subcluster (Fisher's exact test) is plotted. The dashed vertical line corresponds to FDR = 0.01. No alignment significance score is reported for cluster 6 as it only contains cells from the scATAC experiment. Source data are provided as a Source Data file; **g**–**i** Sankey diagrams displaying the comparison between cell annotations in their original publication and in our integrative analysis. Source data are provided as a Source Data file.

cells. Similar conclusions can be derived from the CyTOF annotation (Fig. 5i). We can thus assume that the cells classified in scRNA as cluster 0–2 also correspond respectively to memory, naive, and plasma B cells. However, clustering the scRNA cells on their own would have not allowed us to identify plasma cells (see Supplementary Fig. 10). scConfluence's integration thus had a crucial role in re-annotating the scRNA B cell cluster into appropriate subpopulations. We then further verified whether this subclustering of B cells in scRNA corresponds to a real biological signal or to the random splitting of scRNA B cells driven by the artificial mixing of cells across modalities. With this aim, we identified the differentially expressed genes in clusters 0–2 for both scRNA and scATAC-derived gene activity, separately. CyTOF was excluded from this analysis because of the low number of features (only 48 proteins). We then tested the significance of their intersection (see "Methods" section, Fig. 5f, Supplementary Table 2), finding an overlap of 30 genes (corresponding to a −log10FDR of 19) for cluster 0, 12 genes for cluster 1 (corresponding to a −log10FDR of 10) and 232 genes for cluster 2 (corresponding to a −log10FDR of 37). All of them being well beyond the standard FDR threshold of 0.01 proves that clusters 0–2 share the same differentially expressed genes in scRNA and scATAC. In addition, the common differentially expressed genes contain known markers of memory, naive, and plasma B cells: *AIM2* and *RALGPS2*[22] for memory B cells; *BTG1*, *TCL1A*, and *YBX3*[22] for naive B cells and *MCL1*[63] for plasma B cells. Taken together these results thus confirm that the splitting of scRNA cells annotated as B cells into three subclusters (0–2), is not the result of an artificial modality alignment, but corresponds to real biological signals not identified in the previous unimodal scRNA analysis[60].

B cells are not the only example of cell populations benefitting from single-cell multi-omic integration. Monocytes are also annotated differently across single-cell omics data. Indeed, the scRNA study clusters them into classical and non-classical; CyTOF divides them into classical, non-classical, and intermediate; scATAC splits them into Mono 1 and Mono 2. scConfluence's integration of these three omics data divides monocytes into three clusters (4, 5, and 6), 4 and 5 having a good correspondence with classical and non-classical monocytes, respectively (see Fig. 5g, i). As shown in Fig. 5i, intermediate monocytes tend to cluster in the shared latent space together with non-classical monocytes (cluster 5), probably due to the fact that the clustering algorithm is splitting cell populations into discrete groups while this is a continuum of cells. In addition, the Mono 2 population of scATAC is split into clusters 4 and 5, thus containing both classical and non-classical monocytes. On the opposite, cluster 6 only corresponds to Mono 1 from scATAC, possibly representing a different state of monocytes not fitting within the classical/non-classical subdivision. To confirm such conclusions, we ran the same statistical test as earlier (Fig. 5f, Supplementary Table 3) and found an intersection of differentially expressed genes between scRNA and scATAC of 226 genes for cluster 4 (corresponding to a −log10FDR of 48) and 80 genes for cluster 5 (corresponding to a −log10FDR of 39). In addition, the shared differentially expressed genes contained *CD14*, a known marker of classical monocytes, for cluster 4 and *CD16*, a known marker of non-classical monocytes, for cluster 5. Concerning cluster 6, composed only of scATAC cells, the overexpression of *CD2* and *CCR7* (log2 fold change of 5.61 and 5.40, respectively) could be observed, possibly suggesting that cluster 6 is a group of monocytes transitioning into Dendritic Cells[64,65] (see Supplementary Table 4).

Finally, our integration revealed the mislabeling of a subset of cells annotated as Natural Killer (NK) cells in the original publication. Indeed, these cells formed a distinct subcluster (cluster 8) in which both markers of CD8 T cells (*CD3E)* and NK cells (*NCAM1*) were expressed, as shown in Supplementary Fig. 11. This enabled us to identify them as NKT cells, a heterogeneous group of T cells that share properties of both T cells and NK cells[66].

scConfluence integrates scRNA and neuronal morphologies highlighting morphological heterogeneity in neuronal cell types of mouse motor cortex

The experiments above were focused on molecular data (e.g. transcriptomics, epigenomics, and proteomics), but single-cell analysis can also benefit from other data modalities, such as imaging. A classical situation where imaging data play a key role is the study of neurons. Indeed, morphology imaging data provide a different classification of neocortical neurons with respect to scRNA data. An example of classification based on manual annotation of morphologies divides mouse neocortical interneurons into 15 groups[67] representing different subgroups of Martinotti, neurogliaform, basket, single-bouquet, bitufted, bipolar, double-bouquet, chandelier cell, shrub, horizontally elongated, pyramidal and deep-projecting. On the other hand, in scRNA mouse motor cortex neurons have been classified into 90 populations[68], corresponding to different subpopulations of Lamp5, Sncg, Vip, Sst, Pvalb pyramidal tract, near-projecting, Cortico Thalamic (CT), Extra Telencephalic (ET) and Intra Telencephalic neurons (IT). The integration of these two data modalities has thus a crucial role in unraveling neural heterogeneity and its associated biological functions[69]. This is an extremely challenging task that could not be tackled by the other state-of-the-art methods, as no natural connection exists between the pixels of an image and the features of scRNA data (i.e. genes).

We considered a dataset of 1214 adult mouse primary motor cortex cells profiled with Patch-seq, providing scRNA-seq, neuronal morphologies, and electrophysiology measurements. The dataset is classified, based on scRNA, into Lamp5, Sncg, Vip, Sst, Pvalb, CT, ET, and IT neurons extracted from layers 1, 2/3, 5, and 6[70]. Out of the 1214 cells, only 625 cells were profiled for both scRNA and morphologies, while for the remaining 589 cells only scRNA was available. This is not surprising as Patch-seq is difficult to master, thus implying the production of data containing some modalities and missing others, typical scenario of interest for diagonal integration. As shown in Supplementary Fig. 12a, cells from scRNA perfectly organize according to the cell labels obtained in ref. 70. On the contrary, Supplementary Fig. 12b shows that the scRNA labels do not fully capture the heterogeneity present in the morphology data, thus further suggesting that this modality contains complementary information. We thus investigated the role of such complementarity, by integrating with scConfluence the 625 available morphologies together with the 589 scRNA profiles (Fig. 6a). Since scRNA-seq profiles are available for both groups of cells (Fig. 6a), we can use the measured genes as the connected features to build the **Y** matrices. These measurements are ideal to compute a reliable transport plan across the modalities as they come from the same sequencing technology and dataset.

The cells in scConfluence's shared latent space were broadly organized according to the previously defined scRNA populations (Fig. 6b). At the same time, morphological heterogeneity could be detected in some of these populations. For example, as shown in Fig. 6c, excitatory neurons (CT, ET, IT) are organized into three morphological categories: "tufted", "untufted" and "other" based on the visual inspection of their apical dendrites[71]. Most of the CT neurons are untufted and other, ET neurons are mainly tufted, finally, IT neurons result in a continuum progression from tufted to untufted. This progression seems associated with their layer of origin. For example, tufted IT neurons tend to be from layers 2/3 and 5, while untufted IT neurons are mostly from layer 6. Such morphological heterogeneity is extremely relevant as the geometry of tuft dendrites has an impact on the integrative properties of excitatory neurons[72–74]. In addition, we observe a higher expression of the Transcription Factor *Fezf2* in tufted IT neurons from layer 5 (see Fig. 6d). This result is concordant with the hypothesis that *Fezf2* expression is required for the maintenance of tuftness in IT neurons[75,76]. However, we also observe tufted cells not expressing *Fezf2* as well as untufted cells expressing *Fezf2*, thus raising the possibility that other factors might be involved in such a process.

https://doi.org/10.1038/s41467-024-51382-x

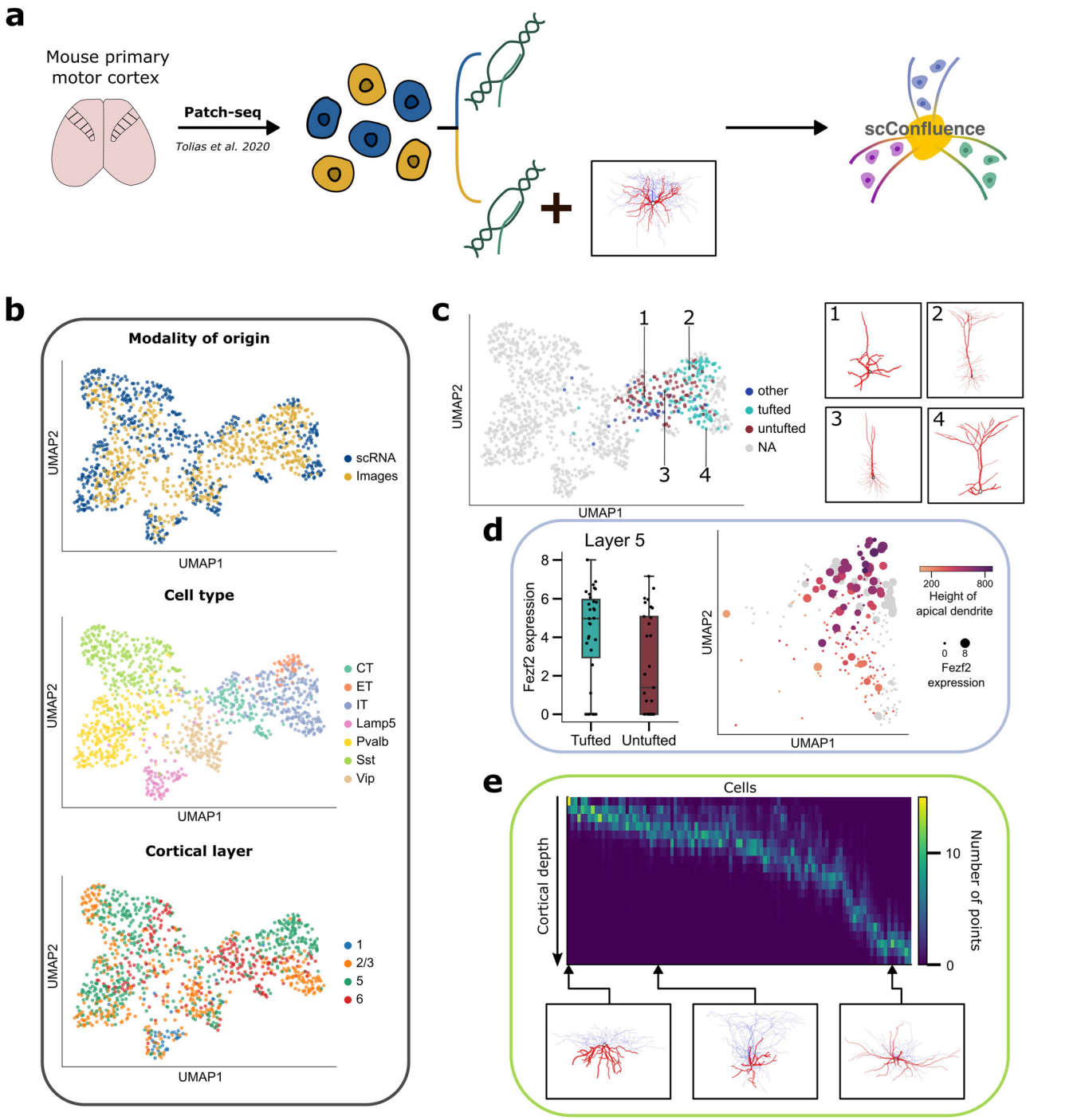

Fig. 6 | Integration of scRNA-seq and neuronal morphologies in the mouse primary motor cortex. a Schematic representation of the integration; **b** UMAP visualizations of the integrated cell embeddings colored by their modality of origin, their cell type annotations, and their cortical layers of origin; **c** UMAP visualization of the integrated cell embeddings colored by their morphological labels which are only available for excitatory neurons. The terms 'tufted' and 'untufted' correspond to visual inspection of the neurons' apical dendrites; some examples of neuronal morphologies are displayed next to the UMAP plot; **d** Pattern of expression of *Fezf2* in IT neurons. The boxplots on the left shows the distribution of expression of *Fezf2* in untufted ($n = 29$) and tufted ($n = 31$) IT neurons from layer 5. The center line, box limits, and whiskers denote the median, upper and lower quartiles, and 1.5× interquartile range, respectively. Black dots over the boxplots correspond to individual data points. Source data are provided as a Source Data file. The UMAP plot of IT neurons shows the correlated pattern of variation of *Fezf2* expression (corresponding to the size of the points) and the height of apical dendrites (corresponding to the color gradient); **e** Heatmap representing the depth profiles of Sst neurons' axons perpendicular to the pia. Cells have been sorted based on the depth of their soma.

Focusing then on all IT neurons, both the expression of *Fezf2* and the length of apical dendrites display a continuous gradient along the same one-dimensional manifold (Fig. 6d). In agreement with this, both *Fezf2* activity and length of apical dendrites have been independently found to be highly correlated with calcium signaling[77,78], which is connected to dendritic excitability through calcium electrogenesis[79,80]. Our observation has particular biological relevance as it could represent not only a simple association but a causal effect of *Fezf2* on the

morphology of IT neurons resulting in a regulation of dendritic excitability. This hypothesis is supported by the fact that *Fezf2* has been already shown to play a key role in the determination of the function, dendritic morphology, and molecular differentiation of CT neurons[81].

Furthermore, Somatostatin-expressing neurons (Sst), which are known to be morphologically diverse[82], seem to be organized according to their layer of origin, with layer 2/3, layer 5, and layer 6 moving from left to right in the last UMAP plot of Fig. 6b. This laminal organization is associated with a morphological pattern of variation, as shown by the axonal depth profiles in Fig. 6e. In layer 2/3 we observe a higher presence of Martinotti cells extending their axons up to layer 1. Indeed, Martinotti cells are known to make contact in layer 1 with the distal tuft dendrites of pyramidal cells[83]. On the other hand, deeper layers contain more non-Martinotti cells which seem to often target neurons inside their own layer.

Discussion

The impressive abundance of unpaired multimodal single-cell data has motivated a growing body of research into the development of integration methods. However, the state-of-the-art suffers from two major drawbacks: (i) the loss of biological information due to across-modalities feature conversion and (ii) the presence of populations only profiled in one data modality.

We introduced scConfluence, a method for single-cell diagonal integration combining uncoupled autoencoders with regularized Inverse Optimal Transport (rIOT) on weakly connected features. scConfluence produces informative cell embeddings in a shared latent space by leveraging the complementarity of multiple modalities profiled from different groups of cells. This aim is achieved by using autoencoders on the full data matrices, allowing simultaneous dimensionality reduction and batch correction of different unpaired data modalities, together with rIOT on connected features to align cells in the shared latent space. This approach allows scConfluence to leverage prior knowledge without discarding the modality-specific features which also provide relevant biological information.

Unlike the state-of-the-art, scConfluence does not rely on the assumption that most features are strongly connected across modalities. Indeed, as soon as such connections allow us to compute meaningful relative distances between cell populations the integration will be successful. This can be achieved even when there are few connected features, as in smFISH-scRNA integration, or when such connections are not perfect, as for proteins and scRNA integration[84]. In addition, the use of unbalanced Optimal Transport allows us to account for the presence of cell populations not shared across modalities.

We extensively benchmarked scConfluence in several scRNA-surface protein and scRNA-scATAC integration problems proving that it outperforms the state-of-the-art. We then explored scConfluence's ability to tackle complex and crucial biological questions. First, we integrated with scConfluence scRNA and smFISH profiled from mouse somatosensory cortex and we imputed spatial patterns of expression for *Scgn*, *Synpr*, and *Olah* relevant for future biological investigations. Second, scConfluence's integration of scRNA-seq, scATAC-seq, and CyTOF in highly heterogeneous human PBMC datasets refined the classification of B cells and Monocytes. Finally, through the integration of neuronal morphological images with scRNA-seq from the mouse primary motor cortex, scConfluence shed light on the combined impact of *Fezf2* expression and apical dendrite morphology on information processing in Intra Telencephalic neurons.

A challenging aspect for scConfluence and all the state-of-the-art is the need to properly deal with rare cell populations. Indeed, rare populations are harder to detect as they are under-represented in parameter estimation. This is even more challenging for methods relying on mini-batch gradient descent (such as scConfluence, scGLUE, and Uniport). Indeed, rare populations are much less likely to be simultaneously sampled from each modality in the mini-batches. At the same time, mini-batch optimization is necessary to scale to millions of cells. In addition, scConfluence, as much as all other state-of-the-art diagonal integration methods, relies on connections between features of different modalities. Such connections are not always available, as for example when integrating electrophysiology measurements with gene expression profiled from different neurons.

One of the main advantages of scConfluence is its modularity, allowing the users to choose their preferred unimodal dimensionality reduction method. For the modalities analyzed in this paper (scRNA-seq, scATAC-seq, CyTOF, smFISH, Patch-seq) ad-hoc autoencoders are proposed. However, for new modalities the users can choose whether to use a classical fully connected autoencoder with the L_2 loss or a more tailored solution available in the literature. Such a tailored solution could be a novel autoencoder architecture or even any parametric dimension reduction model that can be optimized with stochastic gradient descent. Future developments could further improve the performances of scConfluence by plugging-in more advanced dimensionality reduction models recently developed or soon-to-be developed.

Regarding future perspectives, while this work is focused on unpaired multimodal data, paired multimodal data also starts to increasingly accumulate. We can thus expect a relevant need for methods able to jointly integrate these two types of multimodal data. In this setting, paired data would represent a very reliable prior knowledge to guide the alignment of unpaired cells. In addition, they could possibly bring new biological information, not already encoded in the single data modalities. Future developments of scConfluence should be aimed at tackling this intriguing emerging challenge.

Methods

Notations

For two vectors $\mathbf{u} \in \mathbb{R}^{n_u}$ and $\mathbf{v} \in \mathbb{R}^{n_v}$, we use the notations:

$(\mathbf{u} \otimes \mathbf{v})_{ij} = u_i v_j$ and $(\mathbf{u} \oplus \mathbf{v})_{ij} = u_i + v_j$. For two matrices $\mathbf{U} \in \mathbb{R}^{n \times d}$ and $\mathbf{V} \in \mathbb{R}^{n \times d}$ of identical dimensions, we'll use the scalar product notation $\langle , \rangle$ to denote the Frobenius inner product $\langle \mathbf{U}, \mathbf{V} \rangle = \sum_{i,j} U_{ij} V_{ij}$.

Optimal transport

Optimal Transport (OT), as defined by Monge[85] and Kantorovich[86], aims at comparing two probability distributions by computing the plan transporting one distribution to the other with a minimal cost.

While the OT theory has been developed in the general case of positive measures, our application only involves point clouds which are uniform discrete measures $\sum_{i=1}^{n} \frac{1}{n} \delta_{\mathbf{a}_i}$ where the set of $\mathbf{a}_i$ is the support of the point clouds. Therefore, to avoid adding unnecessary complexity in the notations we will denote the probability measures just as the set of positions $\mathbf{a}$.

The classical OT distance, also known as the Wasserstein distance, between two point clouds $\mathbf{a} \in \mathbb{R}^{n_1 \times d}$ and $\mathbf{b} \in \mathbb{R}^{n_2 \times d}$ is defined as:

$$\mathrm{OT}(\mathbf{a}, \mathbf{b}, c) = \min_{\mathbf{P} \in \Pi(n_1, n_2)} \langle \mathbf{P}, \mathbf{c}(\mathbf{a}, \mathbf{b}) \rangle \tag{2}$$

Where $\Pi(n_1, n_2) = \{\mathbf{P} \in \mathbb{R}_+^{n_1 \times n_2} \text{ s.t. } \mathbf{P} = \frac{1}{n_1}\mathbf{1}, \mathbf{P}^{\mathrm{T}} = \frac{1}{n_2}\mathbf{1}\}$ and c is a ground cost function used to compute the pairwise dissimilarity matrix $\mathbf{c}(\mathbf{a}, \mathbf{b}) = c\left(\mathbf{a}_i, \mathbf{b}_j\right)_{1 \le i \le n_1, 1 \le j \le n_2} \in \mathbb{R}_+^{n_1 \times n_2}$ that encodes the cost of transporting mass from one point (e.g. cell) to another. In this uniform discrete case, the coupling $\mathbf{P} \in \Pi(n_1, n_2)$ is a matrix that represents how the mass in the point cloud $\mathbf{a}$ is moved from one point to another in order to transform $\mathbf{a}$ into $\mathbf{b}$.

As real data often contains outliers to which OT is highly sensitive, a more robust extension of OT called unbalanced OT[33] has been

developed.

$$\mathrm{OT}^{\tau}(\mathbf{a},\mathbf{b},c) = \min_{\mathbf{P}\in\Pi(n_1,n_2)} \langle \mathbf{P}, \mathbf{c}(\mathbf{a},\mathbf{b})\rangle + \tau \mathrm{D}\left(\mathbf{P}\mathbf{1}\middle|\frac{1}{n_1}\mathbf{1}\right) + \tau \mathrm{D}\left(\mathbf{P}^{\mathrm{T}}\mathbf{1}\middle|\frac{1}{n_2}\mathbf{1}\right) \quad (3)$$

where τ is a positive parameter controlling the looseness of the relaxation.

In this formulation, the hard constraint on the marginals of the optimal plan is replaced with a soft penalization D which measures the discrepancy between the marginals of the transport plan $\mathbf{P}$ and the uniform distributions on $\mathbf{a}$ and $\mathbf{b}$. While setting $\tau = +\infty$ recovers the balanced OT problem (Eq. 2), using $\tau < +\infty$ allows the transport plan to discard outliers and deal with unbalanced populations. Indeed, in (Eq. 3), unbalanced OT achieves a tradeoff between the constraint to conserve the mass by transporting all of $\mathbf{a}$ onto $\mathbf{b}$ and the aim to minimize the cost of transport. When an outlier is too costly to transport, it is therefore discarded from the plan. A classical choice for D is the Kullback–Leibler divergence. It is defined for two discrete probability distributions represented as vectors of probabilities $\mathbf{p}$ and $\mathbf{q}$ as $\mathrm{KL}(\mathbf{p}|\mathbf{q}) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$. The Total Variation (TV) distance defined as $\mathrm{TV}(\mathbf{p},\mathbf{q}) = |p_i - q_i|$ is also frequently used. The main difference between those two options is that when using TV, each point is either fully transported or discarded while using KL leads to transporting for each point a fraction of the mass which smoothly decreases as the cost of transport increases. We use both in different parts of our methods (see "Optimal Transport solvers").

Adding an entropic regularization to the objective function of (Eq. 2) results in a new optimization problem noted as $\mathrm{OT}_{\varepsilon}(\mathbf{a},\mathbf{b},c)$, where ε is a positive parameter quantifying the strength of the regularization.

$$\mathrm{OT}_{\varepsilon}(\mathbf{a},\mathbf{b},c) = \min_{\mathbf{P}\in\Pi(n_1,n_2)} \langle \mathbf{P}, \mathbf{c}(\mathbf{a},\mathbf{b})\rangle + \varepsilon \mathrm{KL}(\mathbf{P}|\mathbf{a}\otimes\mathbf{b}) \quad (4)$$

While setting $\varepsilon = 0$ recovers the unregularized OT problem (Eq. 2), using $\varepsilon > 0$ makes the problem ε-strongly convex. It can be solved computationally much faster than its unregularized counterpart with the GPU-enabled Sinkhorn algorithm[87].

This entropic regularization can be used in the same fashion in (Eq. 3) to obtain the following problem:

$$\begin{aligned}\mathrm{OT}_{\varepsilon}^{\tau}(\mathbf{a},\mathbf{b},c) = \min_{\mathbf{P}\in\Pi(n_1,n_2)} &\langle \mathbf{P}, \mathbf{c}(\mathbf{a},\mathbf{b})\rangle + \varepsilon \mathrm{KL}(\mathbf{P}|\mathbf{a}\otimes\mathbf{b}) + \tau \mathrm{D}\left(\mathbf{P}\mathbf{1}|\frac{1}{n_1}\mathbf{1}\right) \\ &+ \tau \mathrm{D}\left(\mathbf{P}^{\mathrm{T}}\mathbf{1}|\frac{1}{n_2}\mathbf{1}\right)\end{aligned} \quad (5)$$

While $\mathrm{OT}_{\varepsilon}^{\tau}$ provides a scalable (thanks to the Sinkhorn algorithm) and robust (thanks to the unbalanced relaxation) way to estimate the distance between point clouds, it shouldn't be used as is for machine learning applications. Indeed, it suffers from a bias when $\varepsilon > 0$ and is not a proper metric for measures. In particular, $\mathrm{OT}_{\varepsilon}^{\tau}(\mathbf{a},\mathbf{a},c) > 0$. To solve this issue, a debiased version of (Eq. 5) has been introduced as the unbalanced Sinkhorn divergence[44]:

$$\mathrm{S}_{\varepsilon}^{\tau}(\mathbf{a},\mathbf{b},c) = \mathrm{OT}_{\varepsilon}^{\tau}(\mathbf{a},\mathbf{b},c) - \frac{1}{2}\mathrm{OT}_{\varepsilon}^{\tau}(\mathbf{a},\mathbf{a},c) - \frac{1}{2}\mathrm{OT}_{\varepsilon}^{\tau}(\mathbf{b},\mathbf{b},c) \quad (6)$$

The Sinkhorn divergence $\mathrm{S}_{\varepsilon}^{\tau}$ on the other hand is very well suited to define geometric loss functions for fitting parametric models in machine learning applications. Not only is it robust and scalable but it also verifies crucial theoretical properties such as being positive, definite, convex, and metrizing the convergence in law.

To designate optimal transport problems, we'll use the unified notations $\mathrm{OT}_{\varepsilon}^{\tau}$ and $\mathrm{S}_{\varepsilon}^{\tau}$ for all cases with $\tau = +\infty$ referring to the balanced case and $\varepsilon = 0$ referring to the unregularized case.

scConfluence

scConfluence takes as inputs data from M modalities with $M \geq 2$ where each modality's data comes in the form of a matrix $\mathbf{X}^{(p)} \in \mathbb{R}^{n^{(p)} \times d^{(p)}}$ where the $n^{(p)}$ rows correspond to cells and the $d^{(p)}$ columns are the features that are measured in the p th modality (e.g. genes, chromatin peaks, proteins). For each modality the vector $\mathbf{s}^{(p)}$ whose entries are the batch indexes of the cells in $\mathbf{X}^{(p)}$ is also available. Additionally, for all pairs of modalities (p,p'), we have access to $\mathbf{Y}^{(p,p')} \in \mathbb{R}^{n^{(p)} \times d^{(p,p')}}$ and $\mathbf{Y}^{(p',p)} \in \mathbb{R}^{n^{(p')} \times d^{(p,p')}}$ which correspond to $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(p')}$ translated to a common feature space. The method to obtain $\mathbf{Y}^{(p,p')}$ for each modality is detailed later in "Building the common features matrix" section.

ScConfluence leverages all these inputs simultaneously but in different components to learn low-dimensional cell embeddings $\mathbf{Z}^{(p)} \in \mathbb{R}^{n_p \times d_z}$ in a shared latent space of dimension d_z. For each modality p, we use one autoencoder $\mathrm{AE}^{(p)}$ on $\mathbf{X}^{(p)}$ with modality-specific architectures and reconstruction losses $\mathcal{L}_{\mathrm{AE}_p}$, see the "Training details" section.

While variational autoencoders have become widely popular in single-cell representation learning, we decided not to use them. Indeed, variational autoencoders are trained by optimizing the ELBO which contains two terms, one for the reconstruction of the data and one which is the Kullback–Leibler divergence between the variational posterior and the prior distribution. This second term has been found to aim at a goal conflicting with the reconstruction and to lead to worse inference abilities[88]. With this in mind, we used classical autoencoders with an additional regularization. In our architecture, the encoder still outputs parameters of a Gaussian with diagonal covariance as a variational model would, but instead of forcing this distribution to be close to an uninformative Gaussian prior, we simply add a constant (0.0001) to the outputted standard deviation of the posterior distribution so that our model does not converge to a deterministic encoder during training. This stochasticity in the encoder acts as a regularization against overfitting as it forces the decoder to learn a mapping that is robust to small deviations around latent embeddings.

To handle batch effects within modalities, the batch information $\mathbf{s}^{(p)}$ is used as a covariate of the decoder as done in existing autoencoder-based methods for omics data[35]. Conditioning the decoding of the latent code on its batch index allows our AEs to decouple the biological signal from the sample-level nuisance factors captured in different batches.

Meanwhile, the $\mathbf{Y}^{(p,p')}$ matrices are leveraged to align cells across modalities using Optimal Transport. For each pair of modalities (p,p'), we use the Pearson similarity (see Implementation details) to compute the cost matrix $\mathbf{c}_{corr}\left(\mathbf{Y}^{(p,p')},\mathbf{Y}^{(p',p)}\right)$. Indeed, while the squared L_2 distance is classically used in OT, the Pearson similarity has been shown to better reflect differences between genomic measurements[89]. Using this cost matrix, we derive the unbalanced Optimal Transport Plan $\mathbf{P}^{(p,p')}$ which reaches the optimum in $\mathrm{OT}_{\varepsilon}^{\tau}\left(\mathbf{Y}^{(p,p')},\mathbf{Y}^{(p',p)},\mathbf{c}_{\mathrm{corr}}\right)$. $\mathbf{P}^{(p,p')}$ thus provides a partial plan to match corresponding cells from different modalities in the latent space. Using the unbalanced relaxation of OT to compute $\mathbf{P}^{(p,p')}$ enables scConfluence to efficiently deal with cell populations present only in one modality. Indeed, cell populations that are not shared across modalities will have a higher transport cost and are more likely to be part of the mass discarded by the unbalanced OT plan. Once $\mathbf{P}^{(p,p')}$ is obtained, it provides a correspondence map between modalities which determines which embeddings should be brought closer in the latent space. Since diagonal integration's goal is to embed closely cells that are biologically similar, we enforce a loss term whose specific goal is this:

$$\mathcal{L}_{IOT}^{(p,p')} = \left\langle \mathbf{P}^{(p,p')}, \mathbf{c}_{L_2}\left(\mathbf{Z}^{(p)},\mathbf{Z}^{(p')}\right)\right\rangle \quad (7)$$

where c_{L_2} is the squared L_2 distance such that $\mathbf{c}_{L_2}\left(\mathbf{Z}^{(p)},\mathbf{Z}^{(p')}\right) = \left(||\mathbf{Z}_i^{(p)} - \mathbf{Z}_j^{(p')}||_2^2\right)_{1\leq i\leq n^{(p)}, 1\leq j\leq n^{(p')}}$.

Minimizing $\mathcal{L}_{IOT}^{(p,p')}$ leads to reducing the distance only between the cell embeddings which are matched by $\mathbf{P}^{(p,p')}$. We add to this loss a regularization term which reduces the global distance between the set of embeddings in $\mathbf{Z}^{(p)}$ and those in $\mathbf{Z}^{(p')}$. This allows us to make sure that we do not only juxtapose corresponding cell populations from different modalities, but that they overlap in the shared latent space. To enforce this regularization, we use the unbalanced Sinkhorn divergence (Eq. 6) as both its computational and theoretical properties make it an ideal regularization function for our goal.

All those different objectives contribute together to the following final loss which we optimize over the parameters of the neural networks $\mathrm{AE}^{(p)}$ with stochastic gradient descent:

$$\mathcal{L} = \sum_{p=1}^{M} \lambda_p \mathcal{L}_{\mathrm{AE(p)}} + \sum_{1\le p<p'\le M} \lambda_{IOT} \left\langle \mathbf{P}^{(p,p')}, \mathbf{c}_{L_2}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right)\right\rangle + \lambda_r S_\varepsilon^\tau\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}, c_{L_2}\right) \tag{8}$$

Where the λ_p, λ_{IOT} and λ_r are positive weights controlling the contribution of each different loss terms.

Connection to regularized Inverse Optimal Transport

Our final loss (Eq. 8) can be decomposed in two main objectives, on one side the reconstruction losses whose goal is to extract the maximum amount of information out of each modality, on the other side the alignment loss $\mathcal{L}_{\mathrm{align}}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right)$, whose goal is to align cells across modalities in the shared latent space.

$$\mathcal{L}_{\mathrm{align}}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right) = \lambda_{IOT}\left\langle \mathbf{P}^{(p,p')}, \mathbf{c}_{L_2}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right)\right\rangle + \lambda_r S_\varepsilon^\tau\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}, c_{L_2}\right) \tag{9}$$

There is an intimate connection between $\mathcal{L}_{\mathrm{align}}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right)$ and the theory of Inverse Optimal Transport (IOT).

Regularized Inverse Optimal Transport (rIOT)[32] refers to the problem of learning a pairwise dissimilarity matrix $\mathbf{C}$ from a given transport plan $\mathbf{P} \in \Pi(n_1, n_2)$, with a certain regularization on $\mathbf{C}$. In our case, it can be formalized as the following convex optimization problem:

$$\mathrm{rIOT}_\varepsilon(\mathbf{P}) = \min_{\mathbf{a},\mathbf{b}} \mathrm{KL}\left(\mathbf{P} | \mathbf{Q}_\varepsilon(\mathbf{a},\mathbf{b})\right) + R(\mathbf{a},\mathbf{b}) \tag{10}$$

where $\mathbf{Q}_\varepsilon(\mathbf{a},\mathbf{b})$ is the balanced optimal transport plan achieving the optimum in $\mathrm{OT}_\varepsilon^{+\infty}\left(\mathbf{a},\mathbf{b},c_{L_2}\right)$ and R is a user-defined regularization. In our case, we want this regularization to force points coupled by $\mathbf{P}$ to completely overlap.

We prove that in the particular case of balanced plans, which corresponds to setting $\tilde{\tau} = +\infty$ and $\tau = +\infty$ in our method, and with the regularizing function $R(\mathbf{a},\mathbf{b}) = \frac{1}{\tilde{\varepsilon}}\mathrm{OT}_\varepsilon^{+\infty}\left(\mathbf{a},\mathbf{b},c_{L_2}\right) + \frac{\lambda_r}{\varepsilon\lambda_{IOT}}\mathrm{S}_\varepsilon^{+\infty}\left(\mathbf{a},\mathbf{b},c_{L_2}\right)$, minimizing $\mathcal{L}_{\mathrm{align}}$ with respect to $\mathbf{Z}^{(p)}$ and $\mathbf{Z}^{(p')}$ is equivalent to solving $\mathrm{rIOT}_\varepsilon\left(\mathbf{P}^{(p,p')}\right)$. More formally, we prove that:

$$\mathrm{argmin}_{\mathbf{Z}^{(p)},\mathbf{Z}^{(p')}} \mathcal{L}_{\mathrm{align}}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right) = \mathrm{argmin}_{\mathbf{Z}^{(p)},\mathbf{Z}^{(p')}} \mathrm{KL}\left(\mathbf{P} | \mathbf{Q}_\varepsilon\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right)\right) + R\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right) \tag{11}$$

The proof of Eq. (11) uses the following lemma (See Supplementary Note 1).

Lemma: Let $\mathbf{a}$ and $\mathbf{b}$ be two point clouds of size n_1 and n_2 respectively. Given $\mathbf{P} \in \Pi(n_1, n_2)$ and denoting as $\mathbf{Q}_\varepsilon(\mathbf{a},\mathbf{b})$ the balanced entropic optimal transport plan achieving the optimum in $OT_\varepsilon^{+\infty}\left(\mathbf{a},\mathbf{b},c_{L_2}\right)$, the following equality holds:

$$\mathrm{KL}\left(\mathbf{P} | \mathbf{Q}_\varepsilon(\mathbf{a},\mathbf{b})\right) = \left\langle \mathbf{P}, \log(\mathbf{P})\right\rangle + \frac{1}{\varepsilon}\left\langle \mathbf{P}, \mathbf{c}_{L_2}(\mathbf{a},\mathbf{b})\right\rangle - \frac{1}{\varepsilon}\mathrm{OT}_\varepsilon^{+\infty}\left(\mathbf{a},\mathbf{b},c_{L_2}\right) - 1 \tag{12}$$

Using the lemma (Eq. 12) and the definition of R we prove (Eq. 11) by rewriting $rIOT_\varepsilon\left(\mathbf{P}^{(p,p')}\right)$ as:

$$\begin{aligned}
\mathrm{rIOT}_\varepsilon\left(\mathbf{P}^{(p,p')}\right) &= \min_{\mathbf{Z}^{(p)},\mathbf{Z}^{(p')}} \left\langle \mathbf{P}^{(p,p')}, \log \mathbf{P}^{(p,p')}\right\rangle + \frac{1}{\varepsilon}\left\langle \mathbf{P}, \mathbf{c}_{L_2}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right)\right\rangle \\
&\quad - \frac{1}{\varepsilon}\mathrm{OT}_\varepsilon^{+\infty}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}, c_{L_2}\right) - 1 \\
&\quad + \left(\frac{1}{\varepsilon}\mathrm{OT}_\varepsilon^{+\infty}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}, c_{L_2}\right) + \frac{\lambda_r}{\varepsilon\lambda_{IOT}}\mathrm{S}_\varepsilon^{+\infty}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}, c_{L_2}\right)\right) \\
&= \min_{\mathbf{Z}^{(p)},\mathbf{Z}^{(p')}} \left\langle \mathbf{P}^{(p,p')}, \log \mathbf{P}^{(p,p')}\right\rangle + \frac{1}{\varepsilon}\left\langle \mathbf{P}^{(p,p')}, \mathbf{c}_{L_2}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right)\right\rangle \\
&\quad + \frac{\lambda_r}{\varepsilon\lambda_{IOT}}\mathrm{S}_\varepsilon^{+\infty}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}, c_{L_2}\right) - 1 \\
&= \min_{\mathbf{Z}^{(p)},\mathbf{Z}^{(p')}} \left\langle \mathbf{P}^{(p,p')}, \log \mathbf{P}^{(p,p')}\right\rangle + \frac{1}{\varepsilon\lambda_{IOT}}\mathcal{L}_{\mathrm{align}}\left(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')}\right) - 1
\end{aligned} \tag{13}$$

By noticing in (Eq. 13) that neither $\langle \mathbf{P}^{(p,p')}, \log \mathbf{P}^{(p,p')}\rangle$ nor the scaling factor $\frac{1}{\varepsilon\lambda_{IOT}}$ depends on $(\mathbf{Z}^{(p)}, \mathbf{Z}^{(p')})$, we obtain (Eq. 11).

Training details

Neural network architectures. The encoders and decoders are three-layer neural networks with ReLU activation functions inspired by the architecture of the scVI VAE. We used a latent dimension of 16 for all datasets but adapted the number of neurons in hidden layers to the dimensionality of the datasets (see Supplementary Table 5). On scATAC and scRNA datasets which contained thousands of features, we did a first dimension reduction with PCA and used the 100 principal components as inputs of the encoder while the decoder outputted a reconstruction in the original feature spaces which were compared with the data prior to the PCA projection. For proteomic and smFISH modalities which contained much fewer features, we reduced the number of layers of both encoders and decoders to two. We used the same decoder architecture as scVI with the Zero Inflated Negative Binomial (ZINB) likelihood for the reconstruction loss on scRNA data. For other modalities, however, we replaced the scVI decoder with a simple fully connected multi-layer perceptron and used the squared L_2 distance as the reconstruction loss.

Optimal transport solvers. We used the Python package *POT* to compute the plans $\mathbf{P}^{(p,p')}$ with the function ot.partial.partial_wasserstein. This implementation of unbalanced optimal transport uses the Total variation distance for the penalization of marginals. It is parameterized by the Lagrangian multiplier m associated with $\tilde{\tau}$ to control the unbalancedness of the plan. m is a parameter between 0 and 1 which quantifies how much mass is transported by the optimal plan. The use of TV to penalize the unbalanced relaxation allows $\mathbf{P}^{(p,p')}$ to completely ignore cell populations that are identified to have no equivalent in other modalities. We set by default $m = 0.5$ which produces robust performances when prior information about the level of unbalanceness of the data is not available, see Supplementary Fig. 13. If additional information is available, m can be set accordingly to obtain better results, a higher value being better for situations where the data is more balanced across modalities. We use no entropic regularization in the computation of $\mathbf{P}^{(p,p')}$ ($\tilde{\varepsilon} = 0$) as POT's CPU implementation was already fast enough on our mini-batches for us to afford to avoid using an approximation.

For the unbalanced Sinkhorn divergence we used the Python package *Geomloss*[90] which has very efficient GPU implementations with a linear memory footprint. Indeed, while it cannot take as input a custom cost matrix as POT does, when the cost function is the squared L_2 distance (as is the case for our regularization term) *Geomloss* uses KeOps[91] to implement efficient operations with a small memory footprint and automatic differentiation. Geomloss uses the KL to penalize the unbalanced relaxation. We used the following hyperparameters: "p" = 2, "blur" = 0.01 (which corresponds to $\varepsilon = 0.0001$), "scaling" = 0.8, "reach" = 0.3 (which corresponds to $\tau = 0.09$).

Training hyperparameters. All models were optimized using the PyTorch lightning library. We used the ADAMW optimizer[92] with a learning rate of 0.003. The batch size was set to 256 times the number of modalities. 20% of the dataset was held-out for validation and an early stopping was triggered when the validation loss didn't improve for 40 epochs. As commonly done in the state-of-the-art[29,31,35], we then use all samples (both train and validation) after training to compute cell embeddings and evaluation metrics. In our task, the goal is to encode the whole given dataset on which the model was trained. Unseen samples and generalizability are not relevant for this problem since the models we train are not meant to be then used on different datasets at inference time. There is no information leakage either since the ground-truth information (i.e. cell type labels and pairing information) used to evaluate the methods are not used during training. λ_p was set to 1.0 for all modalities except for ATAC where it was set to 5.0 due to the larger amount of content measured in the ATAC modality and this was the case for all datasets without further need for tuning. For λ_{IOT} which controls the impact of the IOT term inside the full loss, the default value was set to 0.01. Nonetheless, this term can be tuned depending on the reliability and quality of the connected features used to compute the IOT loss term. Indeed, in situations where stronger connections were available across modalities, e.g. when comparing gene expression measurements to gene expression measurements in the mouse cortex datasets, we increased the value of this parameter to 0.05 but we would not recommend using values outside of the [0.01, 0.1] range. The λ_r hyperparameter controls the importance of the Sinkhorn regularization term whose goal is to force corresponding populations across modalities to overlap in the latent space. Theoretically, the best value for λ_r is the lowest which allows a complete overlapping of cells across modalities. We found 0.1 to be a good default value for this hyperparameter but decreased it to 0.03 when integrating 3 modalities at a time since we are summing three regularization terms in this case. We wouldn't recommend using a value outside of the [0., 0.5] range. See our documentation for more details on those hyperparameters at https://scconfluence.readthedocs.io/en/latest/.

Computational runtime. See Supplementary Table for a comparison of the computation times of the different methods on the PBMC 10X dataset (where approximately 10k cells were profiled for both modalities). Overall, scConfluence's running time is comparable with the state-of-the-art. It is indeed slower than the three CPU methods (Seurat, liger, and MultiMAP). Nonetheless, these methods operate on the full dataset in one go and therefore can't be applied to very large datasets. On the other hand, scConfluence is faster than the other two neural network-based methods, all three being much more scalable due to their mini-batch approach.

Outputs of scConfluence

Once the scConfluence model is trained, it can be used to obtain mainly two kinds of output. Firstly, the trained encoders from each modality are used to encode all cells in the integrated latent space. Those cell embeddings can then be used for any downstream analysis such as clustering. Secondly, the model can be used to impute features across modalities by composing the encoder of one modality with the decoder of another modality. While this can be very useful, users should keep in mind that this technique doesn't provide exact "translations" and because of that, this approach is not cycle-consistent, as shown in Supplementary Fig. 14.

Data preprocessing

Details of the different datasets used are available in the Supplementary Table 1. Additionally, the proportions of cell types present in each modality for unpaired datasets are available in Supplementary Figs. 15–18.

scRNA preprocessing. We performed quality control filtering of cells on the proportion of mitochondrial gene expression, the number of expressed genes, and the total number of counts (using Muon's filter_obs). Quality control filtering of genes was performed on the number of cells expressing the gene (using Muon's filter_var). We then kept a copy of the raw counts data before applying the log-normalization which consists of normalizing counts for each cell so that they sum to 10,000 (using Scanpy's normalize_total) and then log transforming them (using Scanpy's log1p). To subselect genes we took the union between the set of 3000 most variable genes in the normalized counts (using Scanpy's highly_variable_genes with flavor = 'seurat') and the set of 3000 most variable genes in raw counts (using Scanpy's highly_variable_genes with flavor = 'seurat_v3'). Finally, the log-normalized counts were used to compute the first 100 principal components which served as the input of the decoder while we kept a copy of the raw counts to evaluate the output of the decoder using the ZINB likelihood (except for the Patch-seq dataset where we used a fully connected decoder with the squared L_2 loss on the log-normalized counts).

scATAC preprocessing. We performed quality control filtering of cells on the number of open peaks and the total number of counts (using Muon's filter_obs). Quality control filtering of peaks was performed on the number of cells where the peak is open (using Muon's filter_var). We didn't apply any further subselection of the peaks after the quality control. Cells were normalized using the TF-IDF normalization (using Muon's tfidf). Finally, the first 100 principal components of the normalized data were used as input to the encoder while the unreduced TF-IDF normalized data was used to evaluate the output of the decoder with a squared L_2 loss.

Protein preprocessing (in Cite-seq and CyTOF). Since the number of measured proteins is small and this data is less noisy than scRNA or scATAC, no quality control or feature selection was performed. We normalized the data using Muon's implementation of the Center Log Ratio technique. This processed data was used for both the encoder and the decoder (with a squared L_2 loss).

smFISH preprocessing. We performed quality control filtering of cells on the proportion of mitochondrial gene expression, the number of expressed genes, and the total number of counts (using Muon's filter_obs). Quality control filtering of genes was performed on the number of cells expressing the gene (using Muon's filter_var). For the smFISH gene counts we used the same normalization technique as in the original study: we normalized by both the total number of molecules of all genes in each cell and the sum of each gene over all cells. This processed data was used for both the encoder and the decoder (with a squared L_2 loss).

Patch-seq morphologies preprocessing. We retrieved the neuronal morphologies as 3D point clouds stored in.SWC files and did not have to do any quality control since only high-quality morphologies could be reconstructed. We then used the *NeuroM* package[93] to load the

morphologies and project them onto the xy-plane (which is actually the xz plane since y and z were switched in the raw files) while coloring each point according to its neuronal compartment type (dendrites in red, axons in blue and soma in black). We then input those images in Google's Inception v3 pre-trained deep neural network to extract features by retrieving the output of the last layer (with 2,048 dimensions). We then concatenated all these feature vectors in a matrix. This processed data was used for both the encoder and the decoder (with a squared L_2 loss).

Building the common features matrix. The first step to construct the cross-modality cost matrices consisted in obtaining the $\mathbf{Y}^{(p,p')}$ and $\mathbf{Y}^{(p',p)}$ matrices.

With scRNA and scATAC data, this consisted in obtaining the gene activity matrix and subsetting the two matrices to the set of common genes. We obtained the gene activities using different techniques depending on the metadata available for each dataset. For the cell lines data we used Maestro[26], for the Multiome PBMC data we used Signac[25], the gene activities for the Open problems Multiome dataset had been already computed by the authors with Signac, and for the tri-omics PBMC dataset we ran the R script provided by the authors on the GitHub repository of their study https://github.com/GreenleafLab/10x-scATAC-2019/blob/master/code/04_Run_Cicero_v2.R using Cicero[94].

With scRNA and Protein data, this consisted in manually inspecting the genecards website to find for each protein its associated coding gene and then subsetting the RNA and Protein data to the pairs available in both modality's features.

With scATAC and Protein, we did the same as with RNA and Protein after obtaining the gene activities from ATAC.

With RNA and smFISH, since all genes measured in the smFISH experiment were also measured in the scRNA dataset we simply subset the scRNA genes to keep only the common genes.

With RNA and Patch-seq morphologies, since for both groups of cells we had access to the scRNA measurements we could directly use those as common features.

Building the biological cost matrix. Having obtained the converted data matrices $\mathbf{Y}^{(p,p')}$ and $\mathbf{Y}^{(p',p)}$, we then applied to each modality's data (ATAC gene activities were treated as RNA) the same preprocessing as described earlier. We then scaled both $\mathbf{Y}$ matrices (except for the Patch-seq since we were comparing scRNA data from the same dataset) and computed the cost matrix by using the correlation distance between each pair of cells from the two modalities using scipy's *cdist*.

Baselines

Seurat. We compare scConfluence to Seurat v3 as the v3 refers to the version aimed at tackling diagonal integration. In practice, we used the R package *Seurat v4.3.0* which finds anchor pairs between cells from different modalities by searching for Mutual Nearest Neighbors after having reduced the dimension of the data with Canonical Correlation Analysis (CCA). Before running the CCA, all modalities are converted to the same features so we followed the same protocol as described above in "Building the common features matrix" section, as it coincides with the indications described in the tutorials available in the Seurat documentation. We ran the Seurat method with default parameters, except for the Protein and smFISH datasets where we set the latent dimension to 15 since the default number of latent dimensions was close to or even higher than the number of features measured. For gene imputation in the scRNA-smFISH experiment, we used the TransferData function as indicated in the documentation.

LIGER. We compare scConfluence to Liger using the R package *rliger v1.0.0*. Liger relies on integrative non-negative matrix factorization (NMF) to perform diagonal integration and also requires as a first step to convert all modalities to common features. We did this step in the same way as for Seurat. For all datasets except the cell lines, we ran Liger with default parameters. On the cell lines simulated experiment, using the default setting of 30 latent dimensions resulted in the embeddings from different modalities being completely separated. Since the latent dimensions can be interpreted as clusters in NMF we used this to set the number of latent dimensions to 3 which greatly improved Liger's results. We could not tune other baselines similarly for this experiment as the dimension of their latent space can't be interpreted similarly and this did provide a competitive advantage to liger since we used the knowledge that there were 3 main clusters in the dataset (which usually can't be known when integrating new datasets). For the Protein and smFISH datasets we set the latent dimension to 15 since the default number of latent dimensions was close to or even higher than the number of features measured. For gene imputation in the scRNA-smFISH experiment, we used a knn regression with the scRNA embeddings serving as reference to predict the expression levels of held-out genes for smFISH embeddings.

MultiMAP. We compare scConfluence to MultiMAP using the Python package *MultiMAP v0.0.1*. MultiMAP is a generalization of the popular UMAP method[95] to the unpaired multimodal setting. MultiMAP combines intra-modality distances with prior knowledge-based cross-modality distances to recover geodesic distances between all cells on a single latent manifold which can then be projected on $\mathbb{R}^2$ for visualization. Intra-modality distances are computed based on low-dimensional projections of the data after preprocessing. We followed the documentation for this step although dimension reduction was not necessary for smFISH and proteomic datasets where the number of features was already lower than a hundred. To compute distances across modalities we converted pairs of modalities to a common feature space we did as described above in "Building the common features matrix" section, as it coincides with the indications described in the tutorials available in the MultiMAP documentation. MultiMAP was run with default parameters on all datasets. For gene imputation in the scRNA-smFISH experiment, we used knn regression with the scRNA embeddings serving as a reference to predict the expression levels of held-out genes for smFISH embeddings.

Uniport. We compare scConfluence to Uniport using the Python package *Uniport v1.2.2*. Uniport uses one encoder which takes as input cells from all modalities converted to common features while using modality-specific decoders to reconstruct each modality's features. It also leverages unbalanced Optimal Transport in the latent space to force different modalities to mix in the latent space. For feature conversion, we proceed as described above in "Building the common features matrix" section, as it coincides with the indications described in the tutorials available in the Uniport documentation. We ran Uniport with default parameters on all datasets. For gene imputation in the scRNA-smFISH experiment, we used the scRNA decoder to map the embeddings of smFISH cells to the scRNA domain as described in the Uniport documentation.

scGLUE. We compare scConfluence to scGLUE using the Python package *scglue v0.3.2*. scGLUE simultaneously trains one variational autoencoder per modality and one graph variational autoencoder which learns feature embeddings based on a prior knowledge-based guidance graph containing connections between features from different modalities. We followed scGLUE's documentation to construct the guidance graph for scRNA and scATAC integration. For scRNA and Protein integration where no documentation was available, we created a graph where each coding gene was linked to its associated protein. For scRNA and smFISH integration, we created a graph with links between each smFISH-measured gene and the same gene in the scRNA data. We ran scGLUE with default parameters on all datasets, except for

the Protein and smFISH datasets where we set the latent dimension to 15 since the default number of latent dimensions was close or even higher than the number of features measured. For gene imputation in the scRNA-smFISH experiment, we used the scRNA decoder to map the embeddings of smFISH cells to the scRNA domain as in other autoencoder-based methods.

GimVI. We compare scConfluence to GimVI using the Python package *scvi-tools v0.16.4*. GimVI is only applicable to scRNA and smFISH integration and simultaneously trains one autoencoder per modality while enforcing mixing between modalities in the latent space with a discriminative neural network trained in an adversarial way. We ran GimVI with default parameters and performed gene imputation as described in its documentation: we used the scRNA decoder to map the embeddings of smFISH cells to the scRNA domain.

Evaluation metrics

We used several scoring functions to assess the quality of the embeddings provided by each method throughout the benchmarking. All methods were run with five different random seeds and we reported the median score, except for Seurat which contains no randomness and could therefore be run with one seed only. Apart from FOSCTTM, all metrics are based on the k-nearest neighbor graph of embeddings. To give a complete overview of the performance of the methods, we computed those metrics with k taking all values in {5,10,15,20,35,50}. Those metrics are therefore displayed as curves whose x-axis corresponds to the values of k.

For the MultiMAP method whose output is not an embedding but a graph whose edge weights represent similarities between integrated cells, we can use this graph to compute nearest neighbors. Additionally, for the *OP Multiome* and *OP Cite* datasets which contained more than 60,000 cells per modality, we evaluated the methods after the training using a subset of 20,000 cell embeddings as the metrics were too expensive to compute on the full results of each method. We suppose in the following that only two modalities are being integrated, as is the case for all benchmarked datasets.

Notations. For each cell i from the pth modality, we denote as $N_k^{(p)}(i)$ the k nearest neighbors of the cell's embedding $\mathbf{z}_i^{(p)}$ in the integrated latent space. c_i denotes the cell type label of cell i.

Purity. The purity score measures the average proportion of an integrated cell's k-nearest neighbors that share the sample's cell type annotation[20]. It thus varies between 0 and 1 with a higher score indicating a stronger performance.

The score can be written as $\frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1}\frac{|\{t\in N_k^{(1)}(i)|c_t=c_i\}|}{k}+\sum_{j=1}^{n_2}\frac{|\{t\in N_k^{(2)}(j)|c_t=c_j\}|}{k}\right)$.

Transfer accuracy. The transfer accuracy is the accuracy of a k-nearest neighbor classifier using one modality as a reference and the other modality as a query. Since both modalities can be the reference and the query, we compute the results of both classifications and report the average of the two scores.

Graph connectivity. The graph connectivity metric assesses how well cells with the same cell type label are connected in the kNN graph representation of the embeddings[47]. This score can be used to detect whether there exist discrepancies in the integrated latent space between cells from different modalities or experimental batches. For each different cell type c, we denote as $G_k(c)$ the subset of the integrated kNN graph containing only cells with label c. We compute for each cell type c the score s_c equal to the size of the largest connected component in $G_k(c)$ divided by the number of cells with label c. The final graph connectivity score is the average of the cell type scores s_c.

FOSCTTM. The Fraction Of Samples Closer Than the True Match (FOSCTTM) metric has been used before to evaluate diagonal integration methods on paired multimodal datasets where both modalities are measured in the same cells[31,38,46]. Since this metric is only designed for paired datasets we can suppose that there are exactly n cells for each modality and that they are ordered such that the ith cell in the first modality is the true match of the ith cell in the second modality. FOSCTTM aims at comparing for every cell from modality p the distance to its true match and the distance to all other cells in the opposite modality which we denote as p' (since $p \in \{1,2\}$, $p' = 3 - p$).

It is classically defined as:

$$\text{FOSCTTM} = \frac{1}{2n}\sum_{i=1}^{n}\frac{r_i^{(1)}}{n}+\frac{r_i^{(2)}}{n} \tag{14}$$

where $r_i^{(p)} = \left|\left\{j \in [1..n]\, s.t.\, d\left(\mathbf{z}_i^{(p)}, \mathbf{z}_j^{(p')}\right) < d\left(\mathbf{z}_i^{(p)}, \mathbf{z}_i^{(p')}\right)\right\}\right|$

However, in this paper, we compute it in a slightly different way. In the previous formula (Eq. 14), we replace $r_i^{(p)}$ with $\tilde{r}_i^{(p)} = \left|\left\{j \in [1..n] \setminus \{i\}\, s.t.\, d\left(\mathbf{z}_i^{(p)}, \mathbf{z}_j^{(p)}\right) < d\left(\mathbf{z}_i^{(p)}, \mathbf{z}_i^{(p')}\right)\right\}\right|$. This means that rather than only assessing how close the true match of a cell's embedding is compared to cells from the opposite modality, we assess how close the true match of the cell's embedding is compared to all cells from both modalities.

With this formula, we can simultaneously evaluate whether the mixing of the two modalities in the shared latent space is complete and verify that corresponding populations are accurately matched across modalities. Other metrics originally designed to assess batch effect correction are often used to evaluate the mixing of modalities such as the batch entropy of mixing[96] but these don't penalize artificial alignments. The complete overlapping of cells from different modalities only matters if those cells are biologically equivalent and this is assessed by our modified formulation of the FOSCTTM.

Cell type FOSCTTM. Additionally, we reported in Supplementary Fig. 6 one other version of the FOSCTTM metric called "cell type FOSCTTM" which only considers cells from the same cell type.

This additional metric allows us to measure whether paired embeddings are closer to each other than other cells from the same cell type, thus enabling us to find out whether the integration learns sub-cell type structures. More formally, the cell type FOSCTTM is defined as:

$$\text{celltypeFOSCTTM} = \frac{1}{2n}\sum_{i=1}^{n}\frac{\hat{r}_i^{(1)}+\hat{r}_i^{(2)}}{2\sum_{k=1}^{n}\mathbf{1}_{(c_i=c_k)}} \tag{15}$$

where $\hat{r}_i^{(p)} = r_i^{(p)} + \left|\left\{j \in [1..n] \setminus \{i\}\, s.t.\, d\left(\mathbf{z}_i^{(p)}, \mathbf{z}_j^{(p')}\right) < d\left(\mathbf{z}_i^{(p)}, \mathbf{z}_j^{(p')}\right) \text{ and } c_i = c_j\right\}\right|$.

Gene imputation

Both the scRNA-seq and smFISH datasets were downloaded using the scvi-tools helper functions *load_cortex()* and *load_smfish()* which select only the overlapping cell types to ensure the consistency of the imputation task. We then divided the 33 genes measured in the smFISH experiment into eleven disjoint groups of 3 genes. Each of these groups corresponded to a different scenario where the three genes were removed from the smFISH data and held-out and then imputed by each method.

Considering the prior knowledge used to connect features across modalities was very reliable (as we could map genes in the smFISH experiment with themselves in the scRNA-seq without any errors) we increased the weight of the IOT loss λ_{IOT} from 0.01 to 0.05 in this experiment.

For both methods which were autoencoder based, Uniport and scGLUE, we used the same technique as us to perform the imputation. For all other methods, we used knn regression using the scRNA-seq

embeddings as a reference to predict the expression levels of held-out genes in smFISH embeddings.

We used the Spearman correlation to quantify the similarity between the imputed values of a gene across all cells with the ground-truth held-out values. It is defined as the Pearson correlation between the rank values of those two vectors. As in the benchmarking section, we ran each method on each scenario with five initialization seeds (except Seurat which contains no stochasticity). For each gene in each scenario, we kept the median Spearman correlation across the five seeds. We then reported one score per imputation scenario and plotted the eleven scores as a violin plot. We can aggregate the Spearman correlation of the three genes forming each scenario using the average or the median therefore we report both the average and median Spearman correlations (aSCC and mSCC).

For the visualization of the imputations, we made use of the recorded 2D positions of the smFISH cells to plot the cells as they are located in the tissue. To better visualize spatial patterns of the imputations, we used the histogram equalization technique on the imputed values.

Tri-omics integration

We removed very rare cell types (containing less than 0.5% of the whole dataset) from all three datasets. this resulted in the removal of ATAC "Immature NK", "Basophil", "Naive CD8 T2" and "Naive CD4 T2"cells as well as CyTOF "Plasmablasts", "cDC1", "CLA + HLADR + NK cells", "Activated gd T cells", "Unclassified", "HLADR + CD38− CD4 T cells", "Cytotoxic CD4 T cells", "DN T cells" and "Activated CD4 T cells".

We clustered the cell embeddings from all modalities using Scanpy's Louvain with a resolution of 0.5.

We then focused on the B cells and monocyte clusters which we reclustered with a resolution of 0.2.

To assess whether the subclusters we found were correctly aligned across modalities we used the same methods as described in scGLUE[31] except that we only used scRNA and scATAC-derived gene activities. Indeed, including the CyTOF data in this analysis would have resulted in removing too many features to be able to design a statistical test with sufficient power. For each of the two populations we subclustered (B cells and monocytes), we tested for significant overlap in cell type marker genes. For both gene expression and gene activities, the cell type markers were identified using Scanpy's one-versus-rest Wilcoxon rank-sum test with the following criteria: FDR < 0.05 and log fold change > 0. The significance of marker overlap was determined by Fisher's exact test.

Patch-seq

We removed the cells that were labeled as "unclassified" or which belonged to rare cell types (there were less than 15 cells labeled either as "Scng" or "NP").

For the scRNA modality, we didn't use the scVI decoder with a ZINB loss but rather just a fully connected decoder with an L_2 loss on the log-normalized counts as it fitted better the data. Similarly to the smFISH/scRNA experiment, we increased the weight of the IOT loss λ_{IOT} to 0.05. Indeed our prior knowledge about connections between features across modalities consisted in connecting each gene with itself as scRNA measurements were available for both modalities. However, in contrast with the smFISH experiment where only a few dozen of genes had been measured in both modalities, here all genes could be connected, making the prior information much stronger than in previous cases. This resulted in the Sinkhorn regularization not being necessary to obtain a good mixing of the two modalities, hence we set λ_r to 0. Moreover, the sets of cells from the two modalities being actually two independent subsets from the exact same dataset, we could expect very little heterogeneity between the cell populations present in each modality and increased the transported mass parameter m from 0.5 to 0.75 in this experiment.

Statistics and reproducibility

The detailed statistical tests were indicated in figures or associated legends where applicable. No statistical method was used to predetermine sample size. No data were excluded from the analyses. Complete randomization was performed for allocating groups. Our study does not involve group allocation that requires blinding.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All relevant data supporting the key findings of this study are available within the article and its Supplementary Information files. **Cell lines**. We retrieve a scCATseq (RNA + ATAC) dataset with 205 cells from three cancer cell lines (HCT116, HeLa-S3, K562). Data is available in the Supplementary Materials of the original publication[15]. **PBMC 10X**. We retrieve a 10X Genomics Multiome (RNA + ATAC) dataset available at https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-no-cell-sorting-10-k-1-standard-2-0-0. **OP Multiome and OP Cite**. We retrieve a Multiome (RNA + ATAC) and a Cite-seq bone marrow dataset from the Open Problems challenge[49]. The GEO accession number is GSE194122 and the data is available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194122. **BMCITE**. We retrieve a CITE-seq (RNA + ADT) bone marrow dataset from Stuart et al.[27], the GEO accession number is GSE128639 and the data is available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128639. **Smartseq cortex**. We retrieve a scRNA-seq mouse somatosensory cortex dataset from Zeisel et al.[57] using scvi-tools's helper function *scvi.data.cortex*. The data is available at https://storage.googleapis.com/linnarsson-lab-www-blobs/blobs/cortex/expression_mRNA_17-Aug-2014.txt. **smFISH**. We retrieve an osmFISH mouse somatosensory cortex dataset from Codeluppi et al.[56] using scvi-tools's helper function scvi.data.cortex. The data is available at http://linnarssonlab.org/osmFISH/osmFISH_SScortex_mouse_all_cells.loom. **3omics RNA**. We retrieved a scRNA-seq dataset of PBMCs from a COVID study[60] and selected cells from all healthy patients. The GEO accession number is GSE150728 and the data is available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150728. **3omics ATAC**. We retrieve a scATAC-seq dataset of PBMCs and Bone marrow cells from an hematopoietic study in which we select the four batches of PBMCs ("PBMC_Rep1", "PBMC_Rep2", "PBMC_Rep3", "PBMC_Rep4"). The GEO accession number is GSE129785 and the data is available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129785. **3omics CyTOF**. We retrieve a CyTOF dataset of PBMCs from a COVID study in which we select an experimental batch of healthy cells (Batch B). The data is available at https://doi.org/10.5281/zenodo.5139560 under the name "CBD-KEY-CYTOF-WB.tar.gz". **Patch neurons**. We retrieve a Patch-seq dataset of mouse primary motor cortex cells[70]. The scRNA counts are available with GEO accession number GSE163764 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163764 and neuronal morphological reconstructions are available at https://download.brainimagelibrary.org/3a/88/3a88a7687ab66069/. Source data are provided with this paper.

Code availability

Package. The Python package for scConfluence is hosted at https://github.com/cantinilab/scconfluence[97]. It can be installed easily by running pip install scconfluence. *Reproducibility*. Code to reproduce the experiments and figures is available at https://github.com/cantinilab/scc_reproducibility.

References

1. Method of the Year 2013. *Nat. Methods* **11**, 1–1 (2014).
2. Potter, S. S. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.* **14**, 479–492 (2018).
3. Papalexi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).
4. Cusanovich, D. A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
5. Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method for single cell chromatin accessibility profiling. *Nat. Commun.* **9**, 5345 (2018).
6. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
7. Lee, J., Hyeon, D. Y. & Hwang, D. Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.* **52**, 1428–1442 (2020).
8. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
9. Clark, S. J. et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).
10. Angermueller, C. et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).
11. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
12. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).
13. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).
14. Swanson, E. et al. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *eLife* **10**, e63632 (2021).
15. Liu, L. et al. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nat. Commun.* **10**, 470 (2019).
16. Miao, Z., Humphreys, B. D., McMahon, A. P. & Kim, J. Multi-omics integration in the age of million single-cell data. *Nat. Rev. Nephrol.* **17**, 710–724 (2021).
17. Method of the year 2019: single-cell multimodal omics. *Nat. Methods* **17**, 1–1 (2020).
18. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many. *Nat. Methods* **17**, 11–14 (2020).
19. Argelaguet, R. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).
20. Huizing, G.-J., Deutschmann, I. M., Peyré, G. & Cantini, L. Paired single-cell multi-omics data integration with Mowgli. *Nat. Commun.* **14**, 7711 (2023).
21. Gayoso, A. et al. Joint probabilistic modeling of single-cell multiomic data with totalVI. *Nat. Methods* **18**, 272–282 (2021).
22. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
23. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* **24**, 695–713 (2023).
24. Xu, Y. & McCord, R. P. Diagonal integration of multimodal single-cell data: potential pitfalls and paths forward. *Nat. Commun.* **13**, 3505 (2022).
25. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
26. Wang, C. et al. Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol.* **21**, 198 (2020).
27. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
28. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
29. Cao, K., Gong, Q., Hong, Y. & Wan, L. A unified computational framework for single-cell data integration with optimal transport. *Nat. Commun.* **13**, 7419 (2022).
30. Jain, M. S. et al. MultiMAP: dimensionality reduction and integration of multimodal data. *Genome Biol.* **22**, 346 (2021).
31. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022).
32. Ma, S., Sun, H., Ye, X., Zha, H. & Zhou, H. Learning cost functions for optimal transport. Preprint at https://doi.org/10.48550/arXiv.2002.09650 (2021).
33. Séjourné, T., Peyré, G. & Vialard, F.-X. Unbalanced optimal transport, from theory to numerics. *Handb. Numer. Anal.* **24**, 407–471 (2023)
34. Virshup, I. et al. The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat. Biotechnol.* **41**, 604–606 (2023).
35. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
36. Yang, K. D. et al. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nat. Commun.* **12**, 31 (2021).
37. Gabriel Peyré and Marco Cuturi, "Computational Optimal Transport: With Applications to Data Science", *Found. Trends Mach. Learn.* **11**, 355–607 (2019).
38. Demetci, P., Santorella, R., Sandstede, B., Noble, W. S. & Singh, R. SCOT: single-cell multi-omics alignment with optimal transport. *J. Comput. Biol.* **29**, 3–18 (2022).
39. Klein, D. et al. Mapping cells through time and space with moscot. Preprint at https://doi.org/10.1101/2023.05.11.540374 (2023).
40. Huguet, G. et al. Manifold interpolating optimal-transport flows for trajectory inference. *Adv. Neural Inf. Process. Syst.* **35**, 29705–29718 (2022).
41. Schiebinger, G. et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943.e22 (2019).
42. Forrow, A. & Schiebinger, G. LineageOT is a unified framework for lineage tracing and trajectory inference. *Nat. Commun.* **12**, 4940 (2021).
43. Bunne, C. et al. Learning single-cell perturbation responses using neural optimal transport. *Nat. Methods* **20**, 1759–1768 (2023).
44. Séjourné, T., Feydy, J., Vialard, F.-X., Trouvé, A. & Peyré, G. Sinkhorn divergences for unbalanced optimal transport. Preprint at https://doi.org/10.48550/arXiv.1910.12958 (2023).
45. Lance, C. et al. Multimodal single cell data integration challenge: Results and lessons learned. In *Proc. NeurIPS 2021 Competitions and Demonstrations Track* 162–176 (PMLR, 2022).
46. Itai, Y., Rappoport, N. & Shamir, R. Integration of gene expression and DNA methylation data across different experiments. *Nucleic Acids Res.* **51**, 7762–7776 (2023).
47. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).

48. Cao, K., Bai, X., Hong, Y. & Wan, L. Unsupervised topological alignment for single-cell multi-omics integration. *Bioinformatics* **36**, i48–i56 (2020).
49. Luecken, M. et al. A sandbox for prediction and integration of DNA, RNA, and proteins in single cells. In *Proc. Neural Information Processing Systems Track on Datasets and Benchmarks*, Vol. 1 (2021).
50. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
51. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
52. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743–748 (2012).
53. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
54. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
55. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* **381**, eabq4964 (2023).
56. Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
57. Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
58. Lopez, R. *et al*. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. Preprint at https://doi.org/10.48550/arXiv.1905.02269 (2019).
59. Nieto, M. et al. Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex. *J. Comp. Neurol.* **479**, 168–180 (2004).
60. Wilk, A. J. et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat. Med.* **26**, 1070–1076 (2020).
61. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
62. Ahern, D. J. et al. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* **185**, 916–938.e58 (2022).
63. Peperzak, V. et al. Mcl-1 is essential for the survival of plasma cells. *Nat. Immunol.* **14**, 290–297 (2013).
64. Di Pucchio, T. et al. CD2+/CD14+ monocytes rapidly differentiate into CD83+ dendritic cells. *Eur. J. Immunol.* **33**, 358–367 (2003).
65. Tang-Huau, T.-L. et al. Human in vivo-generated monocyte-derived dendritic cells and macrophages cross-present antigens through a vacuolar pathway. *Nat. Commun.* **9**, 2570 (2018).
66. Väyrynen, J. P. et al. Spatial organization and prognostic significance of NK and NKT-like cells via multimarker analysis of the colorectal cancer microenvironment. *Cancer Immunol. Res.* **10**, 215–227 (2022).
67. Jiang, X. et al. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**, aac9462 (2015).
68. Yao, Z. et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021).
69. Peng, H. et al. Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174–181 (2021).
70. Scala, F. et al. Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021).
71. Laturnus, S. C. & Berens, P. MorphVAE: generating neural morphologies from 3D-Walks using a variational autoencoder with spherical latent space. In *Proc. 38th International Conference on Machine Learning* 6021–6031 (PMLR, 2021).
72. Eberhardt, F., Herz, A. V. M. & Häusler, S. Tuft dendrites of pyramidal neurons operate as feedback-modulated functional subunits. *PLoS Comput. Biol.* **15**, e1006757 (2019).
73. Ramaswamy, S. & Markram, H. Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron. *Front. Cell. Neurosci.* **9**, 233 (2015).
74. Rhodes, P. A. & Llinás, R. R. Apical tuft input efficacy in layer 5 pyramidal cells from rat visual cortex. *J. Physiol.* **536**, 167–187 (2001).
75. Tantirigama, M. L. S., Oswald, M. J., Duynstee, C., Hughes, S. M. & Empson, R. M. Expression of the developmental transcription factor Fezf2 identifies a distinct subpopulation of layer 5 intratelencephalic-projection neurons in mature mouse motor cortex. *J. Neurosci.* **34**, 4303–4308 (2014).
76. Tantirigama, M. L. S. et al. Fezf2 expression in layer 5 projection neurons of mature mouse motor cortex. *J. Comp. Neurol.* **524**, 829–845 (2016).
77. Clare, A. J., Day, R. C., Empson, R. M. & Hughes, S. M. Transcriptome profiling of layer 5 intratelencephalic projection neurons from the mature mouse motor cortex. *Front. Mol. Neurosci.* **11**, 410 (2018).
78. Galloni, A. R., Laffere, A. & Rancz, E. Apical length governs computational diversity of layer 5 pyramidal neurons. *eLife* **9**, e55761 (2020).
79. Cain, S. M. & Snutch, T. P. Contributions of T-type calcium channel isoforms to neuronal firing. *Channels* **4**, 475–482 (2010).
80. Higley, M. J. & Sabatini, B. L. Calcium signaling in dendritic spines. *Cold Spring Harb. Perspect. Biol.* **4**, a005686 (2012).
81. Tsyporin, J. et al. Transcriptional repression by FEZF2 restricts alternative identities of cortical projection neurons. *Cell Rep.* **35**, 109269 (2021).
82. Nigro, M. J., Hashikawa-Yamasaki, Y. & Rudy, B. Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex. *J. Neurosci.* **38**, 1622–1633 (2018).
83. Wang, Y. et al. Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat. *J. Physiol.* **561**, 65–90 (2004).
84. Slavov, N. Single-cell proteomics: quantifying post-transcriptional regulation during development with mass-spectrometry. *Development* **150**, dev201492 (2023).
85. Monge, G. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. R. Sci.* 666–704 (1781).
86. Kantorovich, L. On the transfer of masses (in Russian). *Dokl. Akad. Nauk* **37**, 227–229 (1942).
87. Cuturi, M. Sinkhorn distances: lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, Vol. 26 (Curran vAssociates, Inc., 2013).
88. Zhao, S., Song, J. & Ermon, S. InfoVAE: balancing learning and inference in variational autoencoders. *Proc. AAAI Conf. Artif. Intell.* **33**, 5885–5892 (2019).
89. Kim, T. et al. Impact of similarity metrics on single-cell RNA-seq data clustering. *Brief. Bioinform.* **20**, 2316–2326 (2019).
90. Feydy, J. et al. Interpolating between optimal transport and MMD using sinkhorn divergences. In *Proc. Twenty-Second International Conference on Artificial Intelligence and Statistics* 2681–2690 (PMLR, 2019).
91. Charlier, B., Feydy, J., Glaunès, J. A., Collin, F.-D. & Durif, G. Kernel operations on the GPU, with Autodiff, without memory overflows. *J. Mach. Learn. Res.* **22**, 1–6 (2021).
92. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. Preprint at https://doi.org/10.48550/arXiv.1711.05101 (2019).
93. Arnaudon, A. et al. NeuroM. Zenodo https://doi.org/10.5281/zenodo.10630119 (2024).
94. Pliner, H. A. et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871.e8 (2018).

95. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. *J. Open Source Softw.* **3**, 861 (2018).
96. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
97. Samaran, J., Peyré, G. & Cantini, L. scConfluence: single-cell diagonal integration with regularized inverse optimal transport on weakly connected features, https://github.com/cantinilab/scconfluence, https://doi.org/10.5281/zenodo.12736947 (2024).

Acknowledgements

The project leading to this manuscript has received funding from the European Union (ERC StG, MULTIview-CELL, 101115618) and the French government under the management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). In addition, this work has been funded by the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005) and the INSERM project ITMO Cancer MIC APL-EpiNet. The work of G. Peyré was supported by the French government under the management of Agence Nationale de la Recherche as part of the 'Investissements d'avenir' program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). This work was performed using HPC resources from GENCI–IDRIS (Grant 2021-*AD011013214*). We acknowledge the help of the HPC Core Facility of the Institut Pasteur and Mélanie Ridel and Déborah Philipps for the administrative support.

Author contributions

J.S., G.P., and L.C. designed and planned the study. J.S. and L.C. wrote the paper. G.P. revised the manuscript. J.S. developed the tool and performed all the analyses.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51382-x.

Correspondence and requests for materials should be addressed to Laura Cantini.

Peer review information *Nature Communications* thanks Jiadong Mao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.