A Comparison of Dimension Reduction Methods for Cytometry by Time-of-Flight Data

Kaiwen Wang[1], Yuqiu Yang[1,2], Fangjiang Wu[2], Bing Song, Ph.D.[2], Xinlei Wang, Ph.D.[1,*], Tao Wang, Ph.D.[2,3,*]

1 Department of Statistical Science, Southern Methodist University, Dallas, TX, USA, 75275.
2 Quantitative Biomedical Research Center, Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.
3 Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.
* Corresponding authors

Short title: Dimension Reduction Methods for CyTOF Data

Key words: CyTOF, scRNA-seq, proteomics, dimension reduction, benchmark

ABSTRACT

Recently developed single-cell profiling technologies, such as single cell sequencing (scRNA-seq) and mass cytometry (CyTOF), provide critical biological insights from a large number of cells and a large number of features. While experimental and informatic techniques around scRNA-seq are much more advanced, research around CyTOF data analysis has severely lagged behind. However, CyTOF profiles the proteomics makeup of the single cells and are more closely related to disease phenotypes than scRNA-seq. CyTOF data are also dramatically different from scRNA-seq data in many aspects. This calls for the evaluation and development of statistical methods for CyTOF data. Dimension reduction (DR) is one of the most critical first steps of analyses of scRNA-seq and CyTOF data. Here, we benchmark 20 of these methods on 110 real and 425 synthetic CyTOF datasets, including 10 Imaging CyTOF datasets, for accuracy, scalability and usability. In particular, we checked the concordance of DR for CyTOF data against scRNA-seq data that were generated from the same samples. We found that SAUCIE and scvis surprisingly perform well across all categories, whereas UMAP particularly excels in downstream analyses and MDS preserves local and global structure well. Our results highlight the complementarity of existing tools, and that the choice of method should depend on the underlying data structure of the single cells. Based on these results, we develop a set of guidelines to help users select the best DR method for their dataset. Our freely available data and evaluation pipeline will aid in the development of improved DR algorithms specifically tailored to analyze data generated from the increasingly popular CyTOF technique.

BACKGROUND

Recently developed single-cell profiling technologies hold the promise to provide critical insights into the cellular heterogeneity in the tissues of various biological conditions, developmental trajectories of the single cells, and how cells communicate with each other. Researchers in the field of genomics and proteomics have separately developed single cell profiling technologies, mainly single cell RNA-sequencing (scRNA-seq) and mass cytometry (CyTOF)(1–4), and their variants. While experimental and informatic techniques around scRNA-seq are much more advanced, research around CyTOF data analysis has severely lagged behind. However, CyTOF profiles the proteomics makeup of the single cells and are more closely related to clinical phenotypes than scRNA-seq. CyTOF can also additionally capture post-translational modifications(5), which is beyond the reach of scRNA-seq. CyTOF is also becoming increasingly popular, at an exponential rate similar to scRNA-seq (**Sup. Fig. 1**). This significant inadequacy calls for the development of sophisticated, robust and innovative informatics methodologies for the analysis of CyTOF data.

There are key differences between CyTOF and scRNA-seq, which makes it questionable to directly translate scRNA-seq analytical methods for use on CyTOF data. For example,

scRNA-seq usually profiles several thousand genes' RNA expression for several thousand cells, while CyTOF profiles up to 120 proteins' expression for 10 to 100 times more cells. The large number of cells enable CyTOF to capture rare populations that may be missed by scRNA-seq data. CyTOF data also uniquely suffers from time-dependent signal decay during the generation of the data by the CyTOF machine. Furthermore, CyTOF data are continuous in nature while scRNA-seq data are inherently integer count. CyTOF is also free from the drop-out issue that is commonly seen in scRNA-seq data(6–8). On the other hand, more and more studies are now trying to generate matched CyTOF and scRNA-seq data for the single cells from the same research subjects(9–11). These two data types are of different nature, and thus are highly complementary with each other. Therefore, analytical procedures need to ensure concordant conclusions to be drawn from integratively analyzing CyTOF and scRNA-seq data. Unfortunately, these issues haven't been addressed thoroughly by researchers, causing difficulties for valid interpretation for the CyTOF data.

The first step of single cell data analysis is data exploration, which is usually achieved through Dimension Reduction (DR), and followed by visualization, cell type assignment, differential expression analyses, *etc*. Proper DR is fundamental to all downstream analyses, and improper DR could bias all conclusions drawn from the CyTOF and scRNA-seq data. Traditional linear DR algorithms such as Principal Component Analyses have existed for decades. With the rise of scRNA-seq, numerous algorithms such as tSNE, uMap(12), SAUCIE(13), ZIFA(14), PHATE(12), scvis(15), and Diffusion map(16), have been developed to capture the complicated non-linear relationships in the high-dimensional scRNA-seq data. However, it's unclear which of these DR methods is the best for CyTOF. The best DR method for scRNA-seq does not necessarily extrapolate to CyTOF.

To fill in this void, we systematically reviewed and compared the performances of 20 DR methods for CyTOF, based on 110 real CyTOF datasets and 425 simulated datasets that cover a comprehensive spectrum of tissue types, composition of cell types and cell type abundances. Of particular interest, we also included 10 Imaging CyTOF/mass cytometry (IMC) datasets(17), which is an expansion of mass cytometry that enables the capturing of the spatial information of the cells (**Sup. Fig. 2**). We also evaluated which of the DR algorithms, when applied on CyTOF data, will be able to generate conclusions more concordant with the scRNA-seq data generated from the same samples. Overall, we found significant complementarity of existing DR tools, and that the choice of method should depend on the underlying data structure of the single cells. We created an interactive set of guidelines (available at dbai.biohpc.swmed.edu/cytof-playground), which gives context-specific recommendations for DR method usage. Our evaluation also highlights some challenges for current methods, and our evaluation strategy can be useful to spearhead the development of new tools that accurately perform DR for CyTOF data, especially in consideration of matched scRNA-seq data.

RESULTS

Overall benchmarking review

In this study, we collected a total of 20 DR methods published by various researchers (**Sup. Table 1**, and **Fig. 1a**), including those developed specifically for single cell-type of data and general purpose DR methods. We benchmarked these DR methods on a total of 110 real CyTOF datasets from various sources (**Fig. 1b**, and **Sup. Table 2**), including both peripheral blood and solid tissues, and also 425 simulated CyTOF datasets of diverse characteristics, using our Cytomulate algorithm (**Fig. 1c**).

For scoring the accuracy of the DR methods, we chose a total of 16 metrics in 6 main categories, characterizing different aspects of the performances of the DR methods (**Sup. Table 3**). The most important criteria were visually demonstrated in **Fig. 1d**. These 6 categories are (1) global structure preservation, (2) local structure preservation, (3) cluster reconstruction (assessing whether similar clusters can be reconstructed each time in the DR space), (4) cluster concordance (concordance of the clusters reconstructed in the DR space and the original space), (5) scRNA concordance (concordance with scRNA-seq data generated from the same samples when available), and (6) cell type-clustering concordance (concordance of the clusters reconstructed in the DR space and the cell typing results). In our study, the DR methods were assessed and ranked mainly by these metrics focusing on the accuracy of DR. But we also assessed and reported their scalability with respect to the number of cells and protein markers, stability of the predictions after subsampling the datasets; and the usability of the tool in terms of software and documentation (**Fig. 1a**, and **Sup. Table 3**).

The data compendium consisted of both real datasets (**Sup. Table 2**), which provide the highest biological relevance, and also synthetic datasets, which offer the most comprehensive and controllable coverage of different characteristics of CyTOF data. These real datasets come from a variety of biological conditions. In particular, while CyTOF has been applied for peripheral blood by immunologists mostly, we included CyTOF data from a lot of solid tissue experiments. It is of interest to compare whether solid tissue and peripheral blood CyTOF data possess any differences that would render different DR methods more suitable. In our real datasets, we also included one Imaging CyTOF dataset created on breast cancer tissues (the BC cohort of **Sup. Table 2**). For simulation of CyTOF data, we developed Cytomulate (**method section**), the first formal simulation tool that is able to well mimic the behaviors of CyTOF data. Cytomulate is general-purpose and has considered unique features of CyTOF data, such as cell differentiation paths, batch effects, temporal signal-decay effects(18), bead events for internal control, and expressional outliers. To our knowledge, prior works like LAMBDA(19), have created specialized procedures to generate simulated CyTOF data in the context of model-based clustering to achieve automatic cell type identification. These simulations cannot be used to test

methods developed for other purposes (*e.g.*, batch normalization, DR). For each simulation, we used a real dataset as a reference, to match its dimensions and other statistical properties.

Overall, we found a large diversity across evaluation criteria and datasets, with SAUCIE, MDS, scvis and Umap being top-performers overall (**Fig. 2**). The performances of the DR methods on the real and simulated CyTOF data are very comparable, indicating the validity of our data simulation approach. We will discuss each evaluation criterion in more details (**Fig. 2-5**), after which we conclude with guidelines for method users and future perspectives for method developers. We developed the CyTOF playground webserver to help future users of DR methods visualize our benchmark results of these 535 CyTOF datasets, and choose the optimal DR methods for their own datasets according to prior benchmark results (**Fig. 1e**). We also offer our unified pipeline for homogenized input/output of the DR methods and the benchmark metrics (**Fig. 1e**), so that future users can easily execute *ad hoc* deployment of DR and the benchmark metrics for their own data. As another important deliverable of our work, we also packaged our CyTOF data simulation algorithm into Cytomulate, and share it through cytomulate.readthedocs.io (**Fig. 1e**) for researchers in the bioinformatics field to perform other CyTOF-related research.

Accuracy of DR methods

We found that accuracy performances of the DR methods were very variable across datasets, indicating that there is no 'one-size-fits-all' method that works well on every dataset (**Fig. 2** and **Fig. 3**). Umap is one of the best DR methods for scRNA-seq data, and it is good to see that Umap also performed well on CyTOF data. But it is somewhat surprising to see that less well-known or less used methods such as SAUCIE, MDS and scvis out-perform Umap on many datasets, and rank higher overall, in terms of accuracy. tSNE and DiffMap, which are also quite popular scRNA-seq DR methods, perform only moderately on CyTOF data. To rule out the potential bias in the different numbers of datasets from each cohort (each one of the 11 cohorts of **Sup. Table 2**), we aggregated the accuracy performance scores of the DR methods for the datasets from the same cohorts. We still observed the same variable performances of the DR methods across different CyTOF data cohorts (**Sup. File 1** and **Fig. 3a**). Also, comparing these different datasets (**Fig. 3a**), we surprisingly observed that tSNE, which is another popular DR method for scRNA-seq data, performs poorly on most CyTOF datasets, except for the Breast cancer (BC) Imaging CyTOF datasets. Also, on the Brain CyTOF dataset, where the single cells were extracted from solid tissues, MDS seems to perform poorly and the best method is scvis, not SAUCIE.

We also examined the different categories of the accuracy metrics, and have identified even stronger complementarity between the DR methods (**Fig. 3b**). For the purpose of this analysis, the performances of the metrics on all datasets were averaged. We found that SAUCIE tends to

perform generally well on most categories. But MDS seems to be the best in preserving local and global structures. Scvis and Umap are better at downstream analysis needs involving clustering and cell type assignment, and also achieving concordance with the matched scRNA-seq data. PHATE also has decent performances on many metrics. tSNE also has a better performance than SAUCIE in terms of achieving concordance with the matched scRNA-seq data. There have been some hot debates regarding the global and local structure preserving capabilities of tSNE and uMAP(20,21). We found, in the context of CyTOF data, that UMAP seems to be better than tSNE at preserving global structures and that they seem to be un-distinguishable in terms of local structure preserving performance.

To more accurately evaluate how the performances of the DR methods vary as a function of the characteristics of the CyTOF datasets. We generated 425 simulated CyTOF datatets by the Cytomulate tool's probabilistic-model-based simulation. We varied the numbers of cells, the numbers of protein markers, the numbers of independent cellular trajectories (#trees), and the number of cell types in the simulated CyTOF datasets. We benchmarked the accuracy of the DR methods on these simulated data with the same criteria (**Fig. 3c**). We also evaluated the performances of the DR methods for each specific category of the accuracy metrics (**Sup. Fig. 3**). MDS, SAUCIE, and scvis are again the top methods, followed by PHATE and Umap. MDS and scvis (subsampling) performs more poorly with small numbers of input cells, while SAUCIE, scvis (no subsampling), PHATE and Umap tend to work better with small numbers of input cells. Scvis (subsampling or no subsampling), PHATE and Umap tend to work better with datasets of more protein features. Scvis (subsampling or no subsampling), and Umap tend to work better with datasets of more cell types. The number of independent trajectories do not have an obvious influence on the performances of the DR methods.

The high variability of the DR methods between datasets of diverse characteristics could indicate some complementarity between the different methods. To test this, we calculated the likelihood of obtaining a top DR result when using more than one DR methods. We evaluated this on the real CyTOF data (**Fig. 2b**). We observed that using a single method (SAUCIE) for DR of CyTOF data can only generate "top" results ~35% of times. Adding a second method can guarantee a top result ~65% of times. Considering all top 5 methods together can lead to a close to 100% chance of obtaining a top result. We performed this analysis on the simulated CyTOF data (**Fig. 2c**) and the combination of all real and simulated CyTOF data (**Fig. 2d**), which yield similar observations. Altogether, this shows that there is considerable complementarity between the different methods and that users should try out a diverse set of methods on their data. Moreover, it also opens up the possibilities for new ensemble methods that utilize this complementarity.

Scalability of DR methods

While many early DR methods (especially those developed for scRNA-seq) were developed at a

time when profiling more than a thousand cells was exceptional, methods now have to cope with half a million of cells from CyTOF data, and maybe soon with even more. To assess their scalability, we ran each method on up- and downscaled versions of the Oetjen dataset(22), where the maximum number of cells is one million (**Fig. 4ab**). Runtime (**Fig. 4a**) and memory usage (**Fig. 4b**) seem to be highly correlated overall. In terms of both runtime and memory usage, we found that the scalability of most methods was overall very poor. Comparatively, SAUCIE, FA, ICA, PCA, NMF are most favorable in terms of runtime and memory usage in general. Umap and tSNE are in the next group of methods with slightly less efficient runtime and memory usage characteristics. The other methods are in the third tier with up to ~100 times more consumption of runtime and memory usage. The runtimes of these methods in our various experiments can be as much as 12 hours and the memory usage can be as large as 700GB.

Due to the scalability issue of many of these methods, we recommend downsampling the whole CyTOF datasets for some of these DR methods. In fact the original publications of several of these DR methods also recommend downsampling. For the purpose of observing the overall cell state space of the cells (such as major cell populations and their relationships), downsampling, up to a certain extent, will probably not negatively impact the conclusions to be drawn. But all downstream analyses will have to be limited to the downsampled cells and rare cell populations may be missed in the down-sampled subset, which is the detrimental effect of downsampling. So whether to adopt downsampling or not also depends on the purpose of the analyses. MDS, which has good accuracy performance, scales poorly with the number of cells (**Fig. 4ab**). We were not able to finish processing most of the experiments with large numbers of cells for MDS. Thus, in all our other benchmark studies, we applied MDS on sub-sampled CyTOF data only. Scvis also scales poorly with numbers of cells in terms of runtime. But it's runtime plateaus with large numbers of cells and never exceeds 12 hours. In addition, scvis's memory consumption is low, and very comparable with the first tier of methods (SAUCIE, *etc*). Therefore, in our all benchmark studies, we evaluated scvis on both the full CyTOF datasets and also down-sampled datasets.

Due to the many different implementations and variations of tSNE, we also specifically focused on these tSNE methods. We found that the Barnes-Hut tSNE shows much poorer performance than the openTSNE and fitSNE, in terms of computation time (**Sup. Fig. 4**), with up to ~70 differences. The memory usage is similar among these tSEN implementations, but still with several folds of differences. This indicates that the specific implementations of the DR algorithms can have a dramatic effect on the performances of the methods.

In the above experiments, we have a total of 34 protein markers in the CyTOF data (from the original Oetjen dataset), which is typical for most current CyTOF datasets. In consideration of the possibility that future iterations of the CyTOF technology could incorporate more protein markers, we up-sampled the Oetjen dataset to create up to 60 protein marker genes. Then we

performed the same benchmark study in terms of runtime (**Fig. 4c**) and memory usage (**Fig. 4d**). Our benchmark analyses show that most methods do not vary too much with respect to the numbers of protein markers in the CyTOF datasets, alleviating concerns for increased runtime/memory usage for future CyTOF data with more protein markers.

Finally, we also noticed during our DR benchmark practices that some of these DR methods will abort with errors due to increasing sample sizes. We evaluated whether the DR methods would abort for at least one of the benchmark experiments for a given number of cells (**Fig. 4e**). Our analyses show that Diffmap, GrandPrix, Isomap, KPCA Poly/RBF, LLE, MDS, and Spectral start to abort with errors with increasing number of cells. We also performed this analysis for error occurrences with a given number of protein markers (**Fig. 4f**). We found that MDS and Diffmap abort with at least one error in the benchmark experiments of any number of protein features.

Altogether, the scalability analysis indicated that the number of cells, but not number of protein markers, in the CyTOF data are an important factor in the choice of method. But unfortunately, CyTOF data inherently profiles more cells than scRNA-seq data. Therefore, our analyses indicate that method development should pay more attention to maintaining reasonable running times and memory usage, when considering CyTOF data.

Stability of DR methods

It is not only important that a method is able to infer an accurate model in a reasonable time frame, but also that it produces a similar model when given very similar input data. To test the stability of each method, we executed each method on 100 bootstrap samples of the Oetjen cohort's Sample A, and calculated the stability of the DR methods, across different bootstrap datasets. Stability is calculated by the variability of the rankings of the DR methods, in terms of the accuracy metrics, across the bootstrap samples. We found that the top methods that have the best stability performances are KPCA, NMF, ZIFA, and FA. The DR methods that have the best accuracy (**Fig. 2**) unfortunately do not possess the best stability performances (**Fig. 5a**). Comparatively, Umap and scvis are more stable. Interestingly, the stability of tSNE largely depends on the specific implementation. MDS, PHATE and SAUCIE have relatively least stability, suggesting future improvements to these methodologies are needed from method developers in this regard. We also calculated the stability of the DR methods, with respect to each accuracy metric category, and found a high level of complementarity between the DR methods. Scvis and PHATE are more stable for global structure preservation, MDS and tSNE are more stable for local structure preservation, UMap and scvis are more stable for downstream analysis needs, Umap and SAUCIE are more stable for concordance with scRNA-seq data.

Another important aspect of the DR methods related to the stability of computation is how much the results will vary as a function of their tuning parameters. Sometimes, it may be preferable

that the DR results do not vary too much even given different parameters. But in other times, the data analytics practitioners may prefer that parameter tuning will lead to more different results given different parameters, so that they can choose a parameter set that yields the optimal DR results. While we leave the choice of preference to the practitioners, we performed a small scale study of parameter tuning. For several top DR performers, we tuned their parameters, as is shown in **Sup. Table 4**. We included PHATE, SAUCIE, scvis, tSNE, and UMAP and left out MDS, as MDS has few tunable parameters. We found that (**Fig. 5b**) all tuned methods benefit from careful selection of settings. tSNE and SAUCIE noticeably benefit the most in terms of both rank improvement and percent improvement in accuracy. However, we do not recommend performing extensive parameter sweep for scvis, as its relative inefficiency outweighs the minimal gain over the default, which already performs well. We also present the specific configurations optimized for each major category of accuracy in **Sup. Table 4** so that users can use these settings as a starting point according to their own needs.

Usability of DR methods

While not directly related to the accuracy of DR, it is also important to assess the quality of the implementation and how user-friendly the DR methods are for a biological user. We evaluated each method (**Fig. 2**) using a transparent checklist of important scientific and software development practices, including documentation, automated testing and integration, regular update, cross platform compatibility, capable of accepting general matrix input, software license, and API support (**Sup. Table 3** and **Sup. File 2**). As shown in **Sup. Fig. 5**, we found that most methods fulfilled the basic criteria, such as the availability of a tutorial and elemental code quality criteria. However, installation issues seem to be quite general in bioinformatics and the DR field is no exception. It seems to be somewhat difficult to handle dependency issues for installation of SAUCIE and Grandprix. Installation of ZIFA requires manual intervention. Scvis is command line-only and thus cannot be flexibly invoked in R or Python, and the official implementation of diffusion map is in R only. The other methods are available in both R and Python. We observed no clear relationship between usability and method accuracy (**Fig. 2a**). Finally, we note that there are several methods that are capable of mapping new data onto existing DR embeddings (*e.g.* scvis, tSNE and Umap), which could be very convenient in handling continuously generated data. Overall, the two top DR methods of scvis and SAUCIE, according to the accuracy criteria, may not be very well known and their authors have not devoted sufficient resources to polish their software implementations. Our work suggests that the field should consider improving the usability of these two methods through various means, to benefit CyTOF researchers.

A uniform pipeline for implementing CyTOF DR methods and evaluation metrics

To facilitate open and user-friendly deployment and evaluation of DR methods, we established

the “CyTOF package” website to provide streamlined implementation of the DR methods and our evaluation metrics (**Fig. 1e**). As we illustrated in **Fig. 6a**, a new DR user of CyTOF data can either take a prior knowledge-driven approach of examining our DR benchmark results of the datasets evaluated in this study, shared conveniently through “CyTOF playground”, and then run their favorite DR methods using their own installation or the docker images we provide on “CyTOF package”. The user can also take an empirical approach of deploying the evaluation metrics (or a subset of our evaluation metrics according to preference or data availability) on their own CyTOF dataset, using the evaluation and DR code implementation that we offered through “CyTOF package”, before picking the best approach to move forward with. We also envision our system to be quite useful to future developers of CyTOF DR methods or other analytical methods for CyTOF data. We hope such a unified and streamlined evaluation-execution paradigm will become more widely adopted for various bioinformatics analytical scenarios so that method users can choose the best method and extract the most accurate signal from the biological data. We illustrate the usefulness of this framework here by comparing the DR results of several top-performing methods on the Levine32 dataset (**Fig. 6b**). Using our framework, this figure can be recreated using only a couple of lines of codes (shown in “CyTOF package”). In the future, this system could be extended to other analytical tasks of CyTOF data, such as cell typing and pseudotime inference, other types of data, such as scRNA-seq data, and easier downstream analyses.

DISCUSSION

In this work, we comprehensively reviewed and benchmarked commonly used DR methods for DR of CyTOF data across a large number of different datasets. Few attempts at tackling this issue have been made and the field seems to think in general that methodologies developed for scRNA-seq data can be directly applied on CyTOF data. This is problematic, as we showed that this assumption is not true in many cases, and there is a high degree of complementarity of DR methods according to different characteristics of the CyTOF datasets. SAUCIE and scvis are all-rounders that perform admirably across all categories without any noticeable flaws, whereas UMAP excels in downstream analyses with clustering workflows. MDS, on the other hand, suffers from poor scalability downstream performance, but its technical advantage in both global and local structure preservation offers a compelling choice as a validation tool.

New users to the field are confronted with an overwhelming choice of DR methods, without a clear idea of which would optimally solve their problem. To address this issue, we also developed a set of guidelines for choosing the best DR method depending on the context of the CyTOF datasets being analyzed. On the other hand, some researchers seem to analyze CyTOF in the same way as traditional flow cytometry data, using commercial software developed for flow cytometry. They ignore the fact that CyTOF is a much more advanced technique that offers capturing of a lot more protein channels, but has minimum spill-over effects(23). They also

ignore all the sophisticated algorithms developed in recent years, mainly in the field of scRNA-seq data analysis, that might be applied for analyzing CyTOF data to reveal complicated biological insights that cannot be afforded by standardized commercial software.

The fields of genomics and proteomics have been developing mostly independently, due to the inherent differences in the experimental technologies. The researchers in each field are generally not very familiar with the other field. Genomics characterize the DNA- and RNA-level perturbations of the biological systems, while proteomics characterize the protein-level perturbations, which are highly complementary. Therefore, the ability to combine both approaches will be extremely powerful. In recent years, there has been a growing trend to marry genomics with proteomics – proteogenomics. At the bulk tissue level, this is reflected in the generation of more and more patient cohorts that contain both genomics and proteomics data, such as the CPTAC consortium(24–27). At the single cell level, this trend is reflected by the growing generation of single cell datasets with both scRNA-seq and CyTOF data(9–11,28). The creation of such matched datasets calls for the development of dedicated analytical methodologies to integrate these two types of data. First, one should ensure that the conclusions drawn from each of the proteomics and genomics modalities are overall consistent, as we reviewed in this work for concordance between CyTOF and scRNA-seq data during DR. Second, robust and innovative statistical methods should be developed for the integrative analyses of CyTOF and scRNA-seq data, to maximize the potential of each technology.

In this regard, CITE-Seq(29), which allows capture of the RNA transcriptomes and cell surface protein expression for the same cell, is unlikely to replace CyTOF (matched with regular scRNA-seq). CITE-seq can capture much less number of cells (a few thousand usually) than CyTOF. And CITE-seq can only profile the expression of protein surface markers, while CyTOF is capable of capturing the expression of both intra-cellular and cell surface proteins(30). It also suffers from drop-outs in the protein expression measurements by the UMI counting approach(31,32), while CyTOF is free from this caveat.

CyTOF possesses a much higher potential of being developed into a clinically applicable diagnostic/prognostic/predictive tool, compared with scRNA-seq, due to its cheaper per-cell cost, shorter workflow, and direct measurement of proteomics changes. Our study of the performances of DR methods on CyTOF data urges the field to rethink analytical strategies for CyTOF data and to develop state-of-the-art tools to address this key gap. This effort should ultimately lead to effective translational application of the CyTOF technology for personalized medicine.

MATERIALS AND METHODS

Real CyTOF datasets used in this study

We accessed a total of 11 cohorts of 110 real CyTOF datasets for this benchmark study (**Sup. Table 2**). These include the TuPro cohort(9), the Oetjen cohort(22), the CyAnno cohort(33), the

LG cohort(34), the Brain cohort(35), the BC cohort(36), the Levine32 cohort(37), the Levine13 cohort(37), the Samusik cohort(38), the Lung cancer cohort(39), and the Covid cohort. For all datasets, we either obtained their cell typing results from the original publications, or performed cell typing ourselves based on clustering of the cells according to their proteomics features followed by manual assignment according to expression of marker genes. We showed the manual cell typing results of the Lung cancer cohort, in **Sup. Fig. 6**, as an example. Matched scRNA-seq data are available for the TuPro and Oetjen datasets. For the BC dataset, there were too many different samples. To avoid biasing our benchmarking results, we selected the top 14 datasets with more than 5,000 cells.

CyTOF data simulation (Cytomulate)

We devised a probabilistic simulation model that captured the key features of the $arcsinh(./5)$ -transformed CyTOF data. Suppose that the desired resulting dataset contains $m = 1, \cdots, M$ protein markers, $n = 1, \cdots, N$ cell events, and $p = 1, \cdots, P$ cell types. Let the probability of observing the pth cell type be π_p. Then, for the nth cell event, we can construct a cell type indicator variable $\Pi_n \sim Categorical(\pi)$, where $\pi = [\pi_1, \cdots, \pi_P]$ is a P-by-1 probability vector.

Given the cell type p for the nth cell event, we assume that without cell differentiation or background noise the expression level for the mth marker is $X_m \sim N(\mu_m^p, (\sigma_m^p)^2)$. Since a cell type typically expresses a set of marker genes, to further mimic the characteristics of CyTOF data, we associate, with each cell type p, a set $A^{\phi,p} = \{m\colon \textit{the mth marker is not expressed in cell type } p\}$. Then, given $A^{\phi,p}$, we set $\mu_m^p = \sigma_m^p = 0, \forall m \in A^{\phi,p}$.

To simulate cellular lineage, we assume that the underlying cell lineage is composed of several non-overlapping cell differentiation trees where cell type p differentiates into z^p "children" $p^{c_1}, \cdots, p^{c_{z^p}}$. From the mth marker of each cell type p to the mth marker of each of its "children" p^c, we construct a differentiation path $G_m^{p \to p^c}: [0, 1] \mapsto R$ such that $G_m^{p \to p^c}(0) = 0$, and $G_m^{p \to p^c}(1) = \mu_m^{p^c} - \mu_m^p$, where μ_m^p and $\mu_m^{p^c}$ are the mean expression levels of the mth marker of cell type p and cell type p^c, respectively. We also assume that $G_m^{p \to p^c}$ is itself a stochastic process whose realization is denoted as $g_m^{p \to p^c}$.

With these setups, the simulated expression level for the nth cell event was calculated following these steps: (1) Generate a cell type p from $Categorical(\pi)$. (2) If the cell type p has "children",

we randomly select from them a cell type p^c. (3) For each marker m, we generate a sample x_m from $N(\mu_m^p, (\sigma_m^p)^2)$, a "pseudotime" t_m^p from $Beta(0.4, 1)$, and a background noise e from $N(0, \sigma^2)$. (4) The final expression level for the mth marker is then calculated as $Y_m^p = x_m + g_m^{p \to p^c}(t_m^p) + e$.

An example simulated CyTOF dataset generated by Cytomulate was showcased in **Sup. File 1**, which shows that Cytomulate-generated CyTOF data closely mimic real CyTOF data.

Development of the CyTOF DR playground webserver

We developed an online web tool for displaying the results of benchmarking the dimension reduction algorithms on CyTOF data. The technical stack includes HTML, CSS, JavaScript and DataTables. The visualization table to display the benchmarking results of the real and the simulation datasets in this web tool was created with the DataTables jQuery plugin. This table provides searching, sorting and pagination features and progressive bars to visualize the performances of the DR methods according to different metrics. There are two main features in this web tool. The first one is to search for DR results based on multiple input conditions (left column). And the other is to dynamically calculate the average of the performance scores for the user-chosen metrics and to rank the chosen DR methods based on the average scores (top row). This web tool is hosted on the Database for Actionable Immunology (DBAI) website (https://dbai.biohpc.swmed.edu/)(31,32,40).

Statistical analysis

All computations were performed in the R and Python programming languages. For all boxplots appearing in this study, box boundaries represent interquartile ranges, whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range, and the line in the middle of the box represents the median. All statistical tests were two-sided unless otherwise described.

Code availability

The CyTOF simulating model is available at: https://cytomulate.readthedocs.io/en/dev/. The CyTOF package webserver is available at github.com/kevin931/cytofDR. The CyTOF playground webserver is available at dbai.biohpc.swmed.edu/cytof-playground.

Data availability

The public CyTOF and scRNA-seq data used in this study were from citations(9,11,22,34–39,41–43) and described in **Sup. Table 2**, with data accession methods documented in this table. The in-house COVID-19 vaccinee CyTOF datasets were available from

DBAI (https://dbai.biohpc.swmed.edu/), with accession codes from CyTOF_00000000000001 to CyTOF_00000000000012.

FIGURES AND TABLE LEGENDS

Fig. 1 Overview of several key aspects of this benchmark study. (a) The datasets, DR methods, validation metrics, and deliverables of this study. (b) Synopsis of the real CyTOF datasets used for evaluation in this study. (c) The simulation algorithm for CyTOF data. (d) Diagrams explaining several of the most important validation metrics in the accuracy category. (e) Deliverables of this study, including, a complete set of guidelines for choosing DR methods based on data characteristics and user preferences (CyTOF playground), our streamlined pipeline for implementation of the DR methods and evaluation metrics (CyTOF package), and the Cytomulate tool.

Fig. 2 Overview of the benchmark study. (a) The DR methods reviewed in this study, their basic characteristics described in text, and their performances in terms of accuracy, scalability, stability, and utility. The accuracy was evaluated on the real CyTOF data and simulated CyTOF data, respectively. And their average accuracy performances are also reported. (b-d) Complementarity of the DR methods, evaluated on the real CyTOF datasets (b), the simulated CyTOF datasets (c), and the combined datasets (d). We define complementarity as the likelihood of obtaining a top-performing method for a given dataset by choosing a specific DR method. In practice, we tallied the top-ranking methods for all real and simulation datasets and calculated the observed proportion of each method being on top. We then select seven top methods according to overall accuracy as presented in (a) and show the increasing likelihood by using a larger set of these complementary methods.

Fig. 3 Detailed results on the accuracy performances of the DR methods. (a) The accuracy performances of the DR methods on all real CyTOF datasets, grouped by the 11 cohorts. Orange color indicates better performance, while the blue color indicates worse performance. All sub-metrics of the accuracy criterion were averaged. (b) The accuracy performances of the DR methods for each accuracy category. Results for all real CyTOF datasets were averaged, and the results for each detailed sub-metric of accuracy was shown. Orange color indicates better performance, while the blue color indicates worse performance. (c) The accuracy performances on the DR methods on the simulated CyTOF data. The performances were visualized as a function of the different characteristics of the CyTOF datasets. Orange color indicates better performance, while the blue color indicates worse performance. All sub-metrics of the accuracy criterion were averaged.

Fig. 4 Detailed results on the scalability performances of the DR methods. The runtime (a,c) and memory usage consumption (b,d) of the DR methods. In (a) and (b), the DR methods were evaluated based on increasing numbers of cells. In (c) and (d), the DR methods were evaluated based on increasing numbers of protein features. (e,f) Whether the DR methods abort with error

in at least one of the runtime/memory benchmarking analyses in each cell number group (e) and protein feature number group (f).

Fig. 5 Detailed results on the stability performances of the DR methods. (a) The stability of the DR methods. We calculated the stability of the DR methods, across different bootstrap datasets generated from the Oetjen cohort's Sample A. Stability is calculated by the variability of the rankings of the DR methods, in terms of the accuracy metrics (combined or separated), across the bootstrap samples. (b) Parameter tuning results. We performed a grid search of common settings for five of the top methods and evaluated their performances across all our metrics. For each method, we found the optimal setting for every major category of accuracy. The improvement in performance is measured by both raw rank difference and percent difference in terms of overall accuracy.

Fig. 6 Practical guidelines for DR method users. (a) A complete set of guidelines for users of DR methods to choose the best method for their CyTOF data. The user can either take a data-driven approach to run their dataset through all chosen DR methods based on their chosen evaluation metrics. Or the user can choose a prior-knowledge driven approach to query the best DR method based on similarity of the characteristics of their CyTOF dataset with those of the datasets we evaluated in this study. Our streamlined pipeline provides implementation of DR methods and the evaluation metrics, for users to benchmark and deploy the DR methods of choice. (b) Demonstration of how a common framework for DR methods facilitates broad applicability using one example CyTOF dataset. The DR plots were generated by our CyTOF Package pipeline.

Sup. Fig. 1 Number of yearly citations returned by Google Scholar, by searching for "CyTOF", "scRNA-seq", and "CyTOF & scRNA-seq".

Sup. Fig. 2 One example breast cancer Imaging CyTOF dataset. The per-cell staining of the Ecadhe (E-cadherin), Ki67, SMA (α-smooth muscle actin), and CD45 channels of one BC IMC sample were shown in the spatial context of the single cells.

Sup. Fig. 3 Performances of the DR methods on the simulation data. (a) Performances of the DR methods on the simulation data for each specific category of the accuracy metrics. (b) Performances of the DR methods on the simulation and real data for each major category of the accuracy metrics (scores of the specific categories averaged for each major category).

Sup. Fig. 4 Focusing on the runtime and memory usage performances of several tSNE algorithms, with increasing numbers of cells.

Sup. Fig. 5 Usability scoring for the DR methods. All detailed usability criteria were displayed, in the format of a heatmap.

Sup. Fig. 6 Manual cell type assignment of the CyTOF data from the Lung cancer cohort.

Sup. Table 1 All DR methods evaluated in this work

Sup. Table 2 CyTOF and scRNA-seq datasets used in this study

Sup. Table 3 Explanation of DR validation metrics employed in this study

Sup. Table 4 Tuning parameters tested for the DR methods. Related to **Fig. 5b**.

Sup. File 1 Additional analyses and discussions involving the CyTOF DR methods

Sup. File 2 Detailed explanation of the usability scoring criteria

FUNDING SOURCES

This study was supported by the National Institutes of Health (NIH) [CCSG 5P30CA142543/TW, NIH 1R01CA258584/TW, XW, U01AI156189/TW], Cancer Prevention Research Institute of Texas [CPRIT RP190208/TW, XW].

AUTHOR CONTRIBUTIONS

K.W. performed all DR benchmark analyses including acquiring of the real data. F.W. and K.W. created the interactive online tool for visualization DR results. Y.Y. and K.W. designed the simulation algorithm and generated the simulation data. T.W. and X.W. supervised the whole study. All authors wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Bibliography

1. Cheung RK, Utz PJ. Screening: CyTOF-the next generation of cell detection. Nat Rev Rheumatol. 2011 Jul 26;7(9):502–3.
2. Leipold MD, Obermoser G, Fenwick C, Kleinstuber K, Rashidi N, McNevin JP, et al. Comparison of CyTOF assays across sites: Results of a six-center pilot study. J Immunol Methods. 2018 Feb;453:37–43.
3. Behbehani GK. Applications of mass cytometry in clinical medicine: the promise and perils of clinical cytof. Clin Lab Med. 2017 Dec;37(4):945–64.
4. Li Y, Wang Y, Yao Y, Griffiths BB, Feng L, Tao T, et al. Systematic Study of the Immune Components after Ischemic Stroke Using CyTOF Techniques. J Immunol Res. 2020 Aug 28;2020:9132410.
5. Fernández-Zapata C, Leman JKH, Priller J, Böttcher C. The use and limitations of

single-cell mass cytometry for studying human microglia function. Brain Pathol. 2020 Nov 3;30(6):1178–91.

6. Lu T, Park S, Zhu J, Wang Y, Zhan X, Wang X, et al. Overcoming Expressional Drop-outs in Lineage Reconstruction from Single-Cell RNA-Sequencing Data. Cell Rep. 2021 Jan 5;34(1):108589.

7. Li WV, Li JJ. An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat Commun. 2018 Mar 8;9(1):997.

8. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods. 2018 Jul;15(7):539–42.

9. Stark SG, Ficek J, Locatello F, Bonilla X, Chevrier S, Singer F, et al. SCIM: universal single-cell matching with unpaired feature sets. Bioinformatics. 2020 Dec 30;36(Suppl_2):i919–27.

10. Ha MK, Kwon SJ, Choi J-S, Nguyen NT, Song J, Lee Y, et al. Mass Cytometry and Single-Cell RNA-seq Profiling of the Heterogeneity in Human Peripheral Blood Mononuclear Cells Interacting with Silver Nanoparticles. Small. 2020 May;16(21):e1907674.

11. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. Nat Immunol. 2019 Jul;20(7):928–42.

12. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, et al. Visualizing structure and transitions in high-dimensional biological data. Nat Biotechnol. 2019 Dec 3;37(12):1482–92.

13. Amodio M, van Dijk D, Srinivasan K, Chen WS, Mohsen H, Moon KR, et al. Exploring single-cell data with deep multitasking neural networks. Nat Methods. 2019 Nov;16(11):1139–45.

14. Pierson E, Yau C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. Genome Biol. 2015 Nov 2;16:241.

15. Ding J, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. Nat Commun. 2018 May 21;9(1):2002.

16. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics. 2015 Sep 15;31(18):2989–98.

17. Chang Q, Ornatsky OI, Siddiqui I, Loboda A, Baranov VI, Hedley DW. Imaging Mass Cytometry. Cytometry A. 2017 Feb 3;91(2):160–9.

18. Finck R, Simonds EF, Jager A, Krishnaswamy S, Sachs K, Fantl W, et al. Normalization of mass cytometry data with bead standards. Cytometry A. 2013 May;83(5):483–94.

19. Abe K, Minoura K, Maeda Y, Nishikawa H, Shimamura T. Model-based clustering for flow and mass cytometry data with clinical information. BMC Bioinformatics. 2020 Sep 17;21(Suppl 13):393.

20. Kobak D, Linderman GC. Initialization is critical for preserving global data structure in both t-SNE and UMAP. Nat Biotechnol. 2021 Feb 1;39(2):156–7.

21. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018 Dec 3;37:38–44.

22. Oetjen KA, Lindblad KE, Goswami M, Gui G, Dagur PK, Lai C, et al. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. JCI Insight. 2018 Dec 6;3(23).

23. Chevrier S, Crowell HL, Zanotelli VRT, Engler S, Robinson MD, Bodenmiller B. Compensation of signal spillover in suspension and imaging mass cytometry. Cell Syst. 2018 May 23;6(5):612-620.e5.

24. Satpathy S, Krug K, Jean Beltran PM, Savage SR, Petralia F, Kumar-Sinha C, et al. A proteogenomic portrait of lung squamous cell carcinoma. Cell. 2021 Aug 5;184(16):4348-4371.e40.

25. Wang L-B, Karpova A, Gritsenko MA, Kyle JE, Cao S, Li Y, et al. Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell. 2021 Apr 12;39(4):509-528.e20.

26. Krug K, Jaehnig EJ, Satpathy S, Blumenberg L, Karpova A, Anurag M, et al. Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. Cell. 2020 Nov 25;183(5):1436-1456.e31.

27. Satpathy S, Jaehnig EJ, Krug K, Kim B-J, Saltzman AB, Chan DW, et al. Microscaled proteogenomic methods for precision oncology. Nat Commun. 2020 Jan 27;11(1):532.

28. Maeda T, Wakasawa T, Shima Y, Tsuboi I, Aizawa S, Tamai I. Role of polyamines derived from arginine in differentiation and proliferation of human blood cells. Biol Pharm Bull. 2006 Feb;29(2):234–9.

29. Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods. 2019 May;16(5):409–12.

30. Lin D, Gupta S, Maecker HT. Intracellular cytokine staining on PBMCs using CyTOF™ mass cytometry. Bio Protoc. 2015 Jan 5;5(1).

31. Zhang Z, Xiong D, Wang X, Liu H, Wang T. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. Nat Methods. 2021 Jan 6;18(1):92–9.

32. Lu T, Zhang Z, Zhu J, Wang Y, Jiang P, Xiao X, et al. Deep learning-based prediction of the T cell receptor–antigen binding specificity. Nat Mach Intell. 2021 Sep 23;

33. Kaushik A, Dunham D, He Z, Manohar M, Desai M, Nadeau KC, et al. CyAnno: A semi-automated approach for cell type annotation of mass cytometry datasets. Bioinformatics. 2021 May 25;

34. Hawley D, Ding J, Thotakura S, Haskett S, Aluri H, Kublin C, et al. RNA-Seq and CyTOF immuno-profiling of regenerating lacrimal glands identifies a novel subset of cells expressing muscle-related proteins. PLoS ONE. 2017 Jun 29;12(6):e0179385.

35. Böttcher C, Schlickeiser S, Sneeboer MAM, Kunkel D, Knop A, Paza E, et al. Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. Nat Neurosci. 2019 Jan;22(1):78–90.

36. Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, et al. The single-cell pathology landscape of breast cancer. Nature. 2020 Feb;578(7796):615–20.

37. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir ED, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell. 2015 Jul 2;162(1):184–97.

38. Samusik N, Good Z, Spitzer MH, Davis KL, Nolan GP. Automated mapping of phenotype space with single-cell data. Nat Methods. 2016 Jun;13(6):493–6.

39. Gonugunta AS, von Itzstein MS, Mu-Mosley H, Fattah F, Farrar JD, Mobely A, et al. Humoral and cellular correlates of a novel immune-related adverse event and its treatment. J Immunother Cancer. 2021 Dec;9(12).

40. Zhu J, Gouru A, Wu F, Berzofsky JA, Xie Y, Wang T. BepiTBR: T-B reciprocity enhances B cell epitope prediction. iScience. 2022 Feb 18;25(2):103764.

41. Schulz D, Severin Y, Zanotelli VRT, Bodenmiller B. In-Depth Characterization of Monocyte-Derived Macrophages using a Mass Cytometry-Based Phagocytosis Assay. Sci Rep. 2019 Feb 13;9(1):1925.

42. Rundberg Nilsson A, Bryder D, Pronk CJH. Frequency determination of rare populations by flow cytometry: a hematopoietic stem cell perspective. Cytometry A. 2013 Aug;83(8):721–7.

43. Mosmann TR, Naim I, Rebhahn J, Datta S, Cavenaugh JS, Weaver JM, et al. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, part 2: biological evaluation. Cytometry A. 2014 May;85(5):422–33.

Fig. 1

(a)

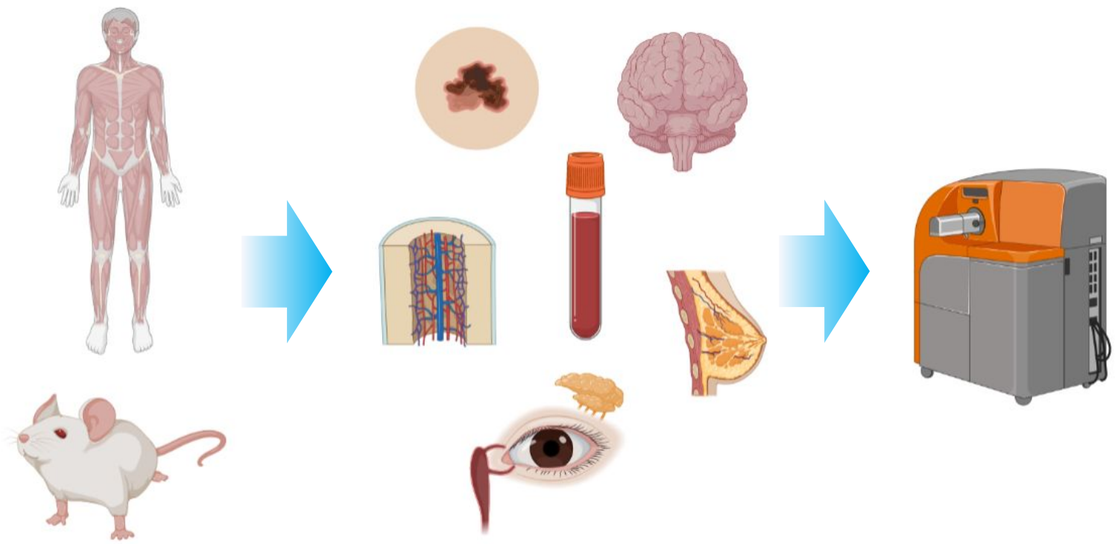
110 real and 425 simulated CyTOF datasets + 20 dimension reduction methods + 30 metrics

Accuracy + Scalability + Stability + Usability

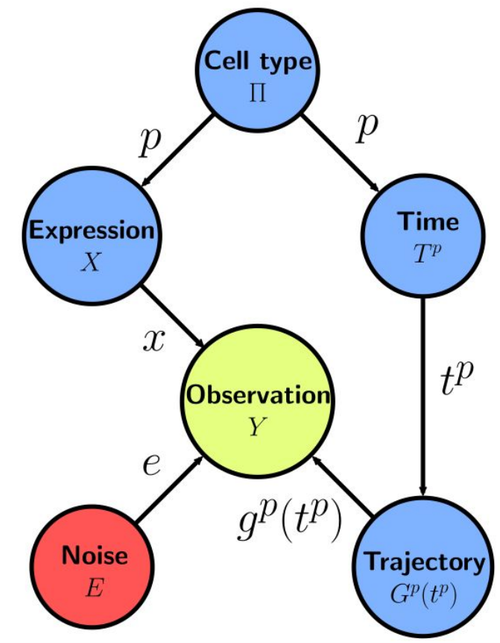
User guidelines + New possibilities for developers

Method/benchmark wrappers:
github.com/kevin931/cytofDR
Guideline app: dbai.biohpc.swmed.edu/cytof-playground
Simulation tool: cytomulate.readthedocs.io

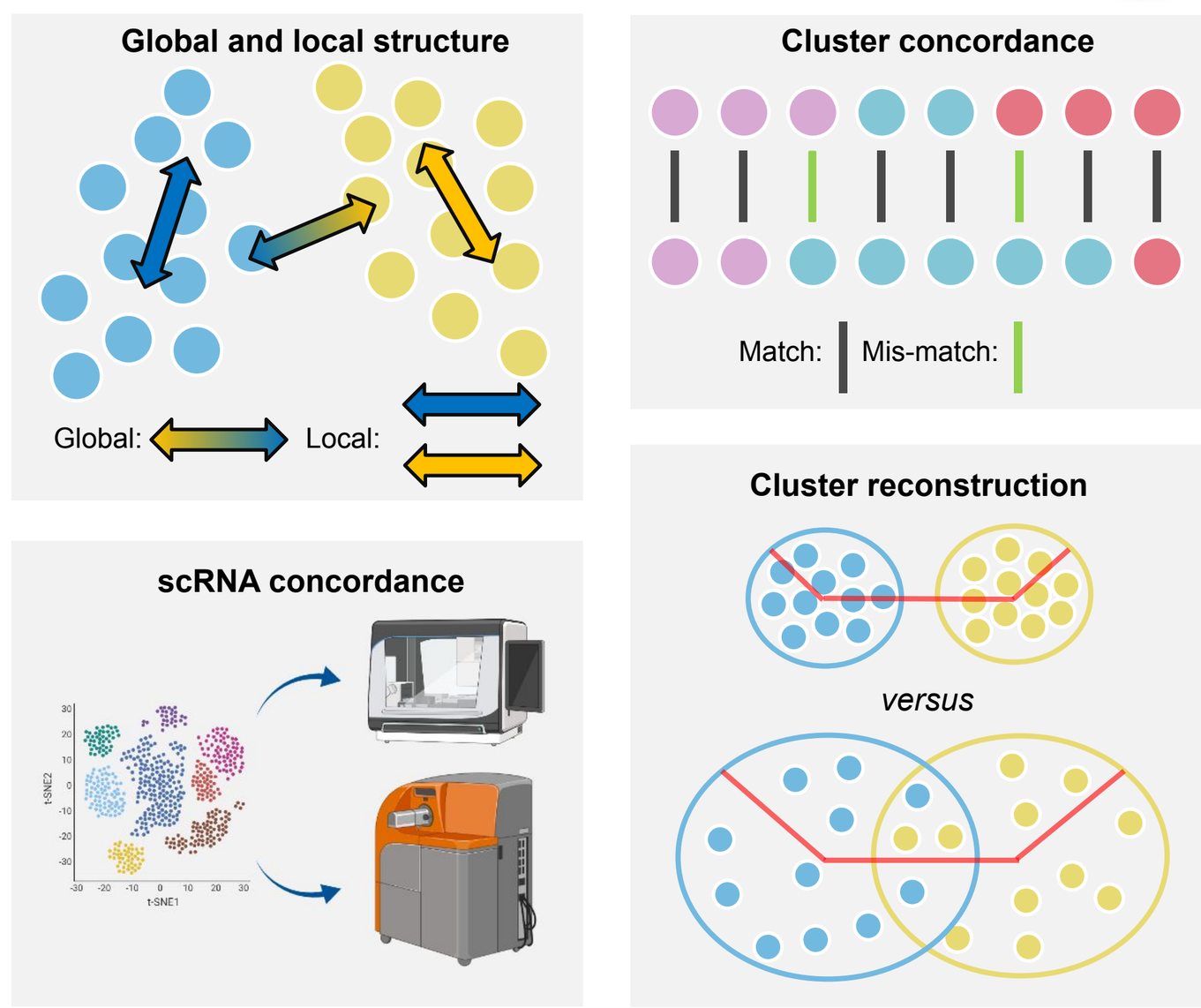
(b)



(c)



(d)



(e)

CYTOF PACKAGE

CYTOF PLAYGROUND

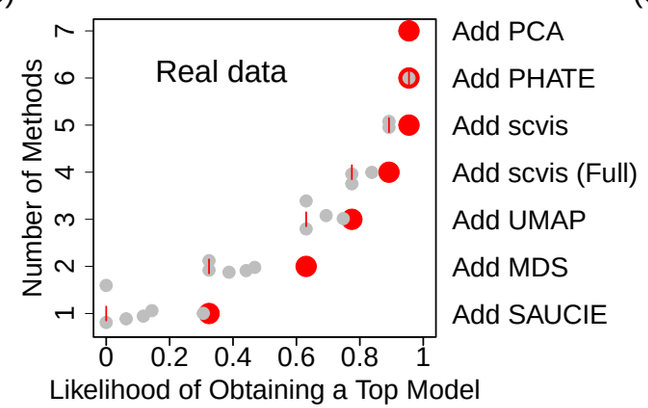
CYTOMULATE
MADE FOR CYTOF

Fig. 2

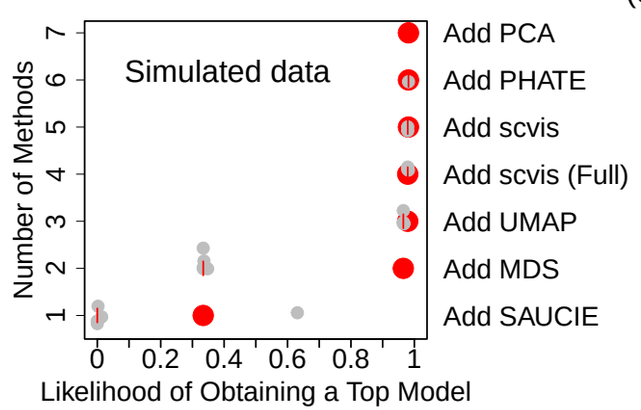
(a)

Method	Platform	Package	Purpose	Mapping	Accuracy: Overall	Accuracy: Real	Accuracy: Simulation	Scalability	Stability	Usability
SAUCIE	Python	SAUCIE	Single Cell	No						
MDS	Python	sklearn	General	No						
UMAP	Python	umap-learn	General/Single Cell	Yes						
scvis (Full)	Python	scvis	scRNA	Yes						
scvis	Python	scvis	scRNA	Yes						
PHATE	Python	phate	Single Cell	No						
PCA	Python	sklearn	General	Yes						
KPCA Poly	Python	sklearn	General	No						
Isomap	Python	sklearn	General	Yes						
tSNE (sklearn)	Python	sklearn	General	No						
GrandPrix	Python	grandprix	Single Cell	No						
tSNE	Python	openTSNE	General/Single Cell	Yes						
KPCA RBF	Python	sklearn	General	No						
DiffMap	R	destiny	General/Single Cell	No						
ICA	Python	sklearn	General	Yes						
FA	Python	sklearn	General	No						
ZIFA	Python	ZIFA	scRNA	No						
NMF	Python	sklearn	General	No						
Spectral	Python	sklearn	General	No						
LLE	Python	sklearn	General	Yes						

(b)



(c)



(d)

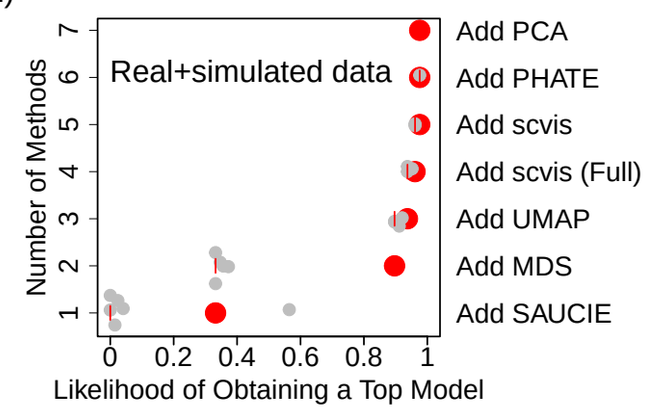


Fig. 3

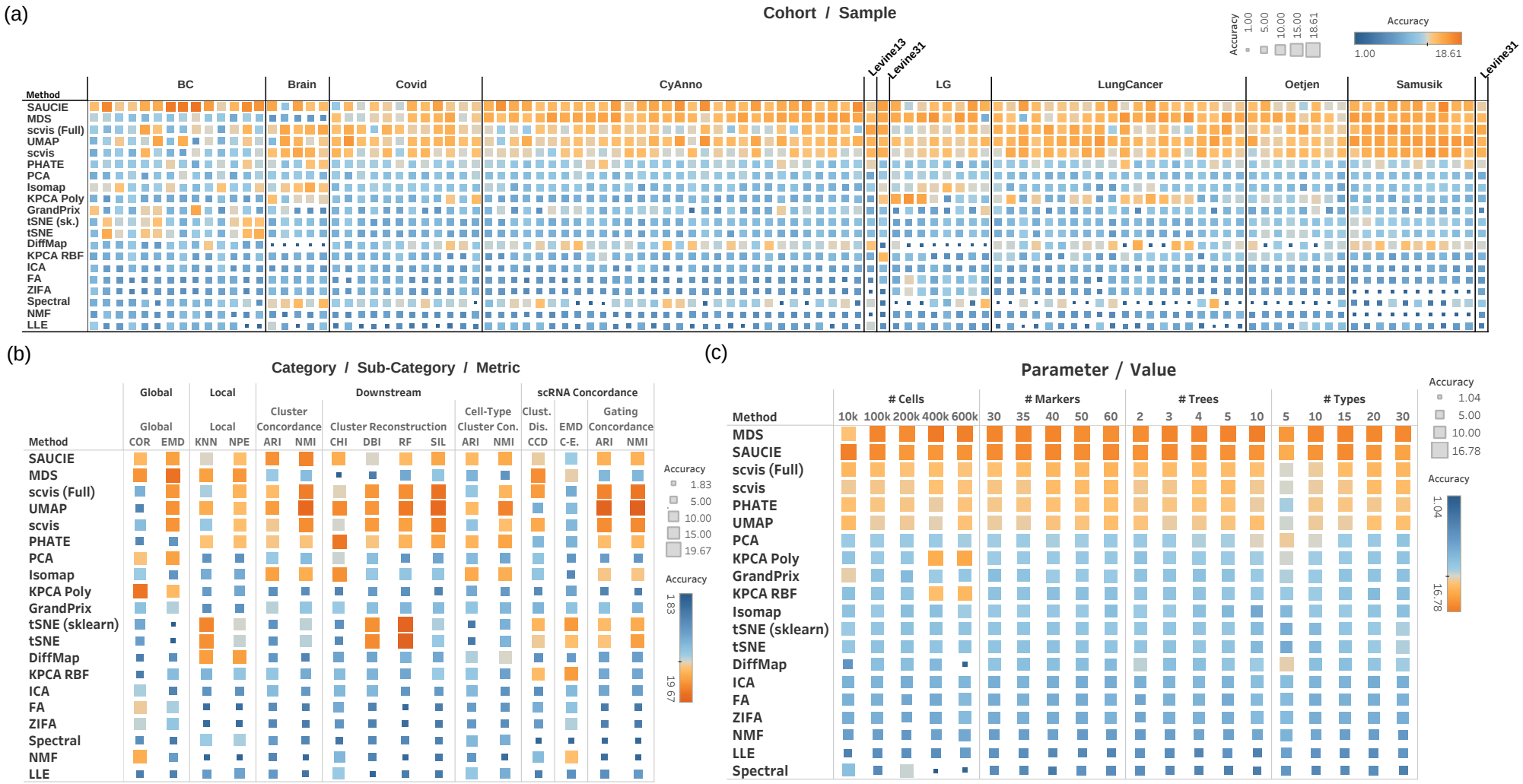

(a)
Cohort / Sample
Accuracy
1.00
5.00
10.00
15.00
18.61
Accuracy
1.00
18.61
Method
BC
Brain
Covid
CyAnno
Levine13
Levine31
LG
LungCancer
Oetjen
Samusik
Levine31
SAUCIE
MDS
scvis (Full)
UMAP
scvis
PHATE
PCA
Isomap
KPCA Poly
GrandPrix
tSNE (sk.)
tSNE
DiffMap
KPCA RBF
ICA
FA
ZIFA
Spectral
NMF
LLE
(b)
Category / Sub-Category / Metric
Global
Local
Downstream
scRNA Concordance
Global
Local
Cluster Concordance
Cluster Reconstruction
Cell-Type Cluster Con.
Clust. Dis.
EMD
Gating Concordance
Method
COR
EMD
KNN
NPE
ARI
NMI
CHI
DBI
RF
SIL
ARI
NMI
CCD
C-E.
ARI
NMI
SAUCIE
MDS
scvis (Full)
UMAP
scvis
PHATE
PCA
Isomap
KPCA Poly
GrandPrix
tSNE (sklearn)
tSNE
DiffMap
KPCA RBF
ICA
FA
ZIFA
Spectral
NMF
LLE
Accuracy
1.83
5.00
10.00
15.00
19.67
Accuracy
1.83
19.67
(c)
Parameter / Value
Accuracy
1.04
5.00
10.00
16.78
Accuracy
1.04
16.78
Method
Cells
10k
100k
200k
400k
600k
Markers
30
35
40
50
60
Trees
2
3
4
5
10
Types
5
10
15
20
30
MDS
SAUCIE
scvis (Full)
scvis
PHATE
UMAP
PCA
KPCA Poly
GrandPrix
KPCA RBF
Isomap
tSNE (sklearn)
tSNE
DiffMap
ICA
FA
ZIFA
NMF
LLE
Spectral

Fig. 4

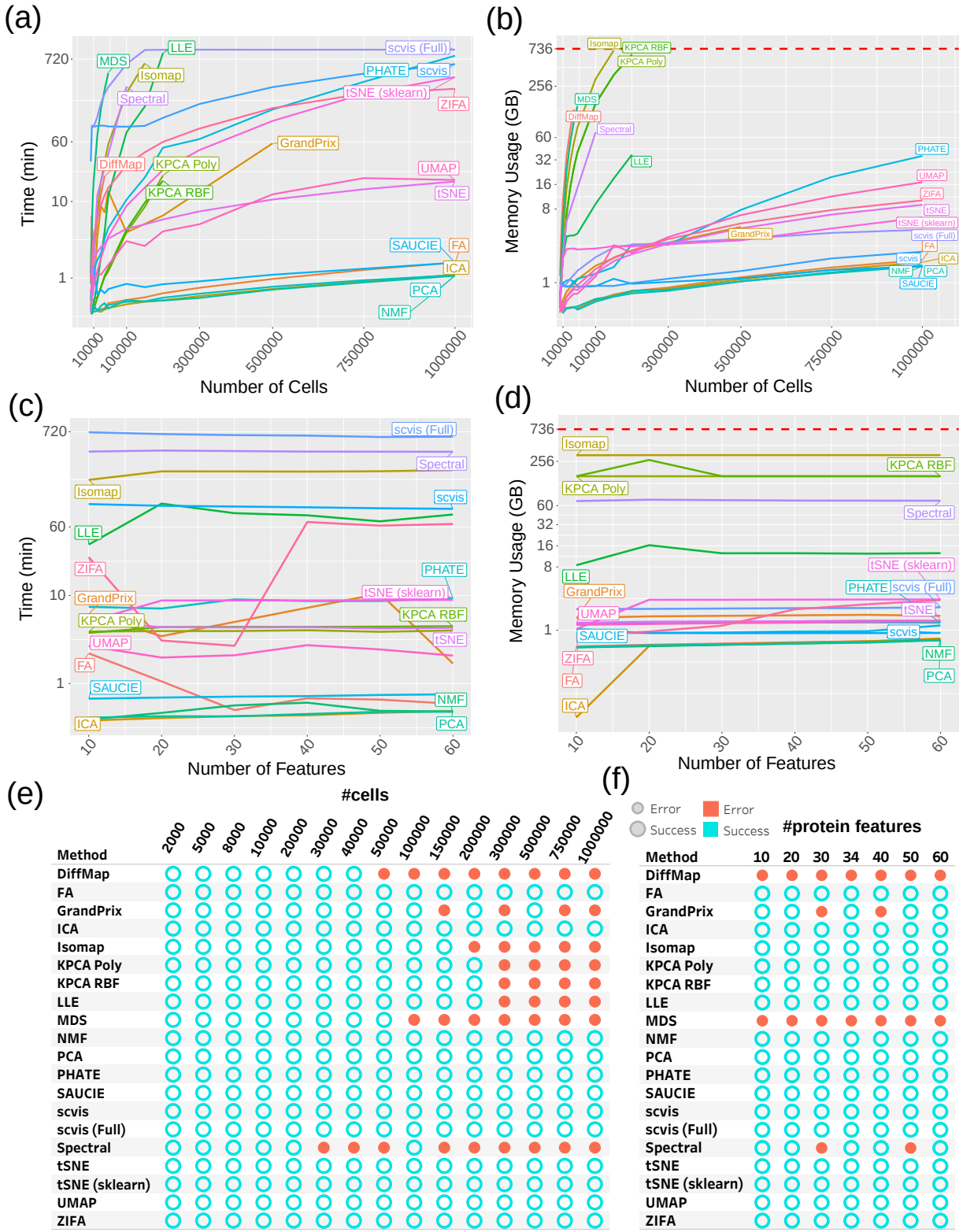
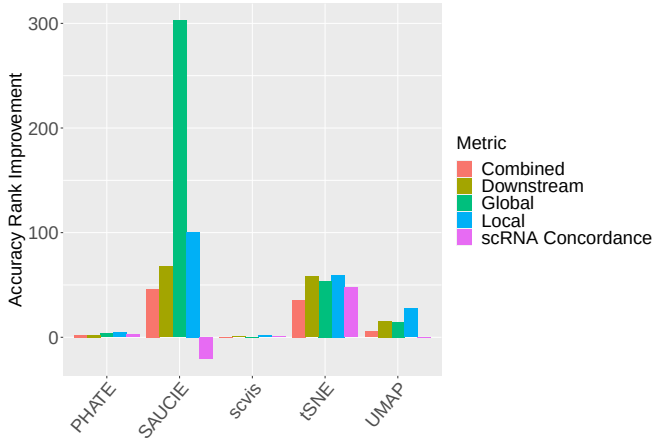

(a)
Time (min)
Number of Cells
LLE
MDS
Isomap
Spectral
scvis (Full)
PHATE
scvis
tSNE (sklearn)
ZIFA
GrandPrix
DiffMap
KPCA Poly
KPCA RBF
UMAP
tSNE
SAUCIE
FA
ICA
PCA
NMF
(b)
Memory Usage (GB)
Number of Cells
(c)
Time (min)
Number of Features
(d)
Memory Usage (GB)
Number of Features
(e)
#cells
Method
DiffMap
FA
GrandPrix
ICA
Isomap
KPCA Poly
KPCA RBF
LLE
MDS
NMF
PCA
PHATE
SAUCIE
scvis
scvis (Full)
Spectral
tSNE
tSNE (sklearn)
UMAP
ZIFA
(f)
Error
Success
Error
Success
#protein features

Fig. 5

(a)

Metric	KPCA RBF	NMF	KPCA Poly	ZIFA	FA	ICA	UMAP	PCA	scvis (Full)	scvis	tSNE	MDS	PHATE	SAUCIE	Isomap	tSNE Sk.	GrandP.	Spectral	LLE	diffmap
Overall	20.00	19.00	18.00	17.00	16.00	15.00	14.00	13.00	12.00	11.00	10.00	9.00	7.00	8.00	6.00	5.00	4.00	3.00	2.00	1.00
Global	20.00	17.00	19.00	11.00	16.00	7.00	5.00	12.00	13.00	18.00	10.00	9.00	15.00	2.00	6.00	14.00	4.00	8.00	3.00	1.00
Local	19.00	16.00	20.00	14.00	13.00	10.00	8.00	11.00	5.00	6.00	15.00	18.00	7.00	9.00	12.00	17.00	3.00	4.00	2.00	1.00
Downstream	9.00	16.00	20.00	11.00	8.00	7.00	18.00	4.00	19.00	17.00	14.00	13.00	10.00	15.00	6.00	12.00	5.00	3.00	2.00	1.00
scRNA Con.	19.00	20.00	13.00	14.00	12.00	17.00	18.00	16.00	10.00	5.00	7.00	8.00	4.00	15.00	6.00	3.00	9.00	11.00	2.00	1.00

(b)



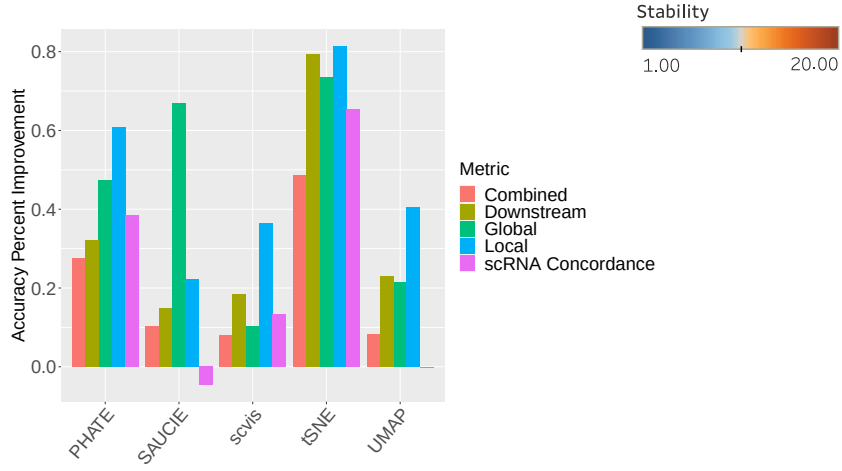
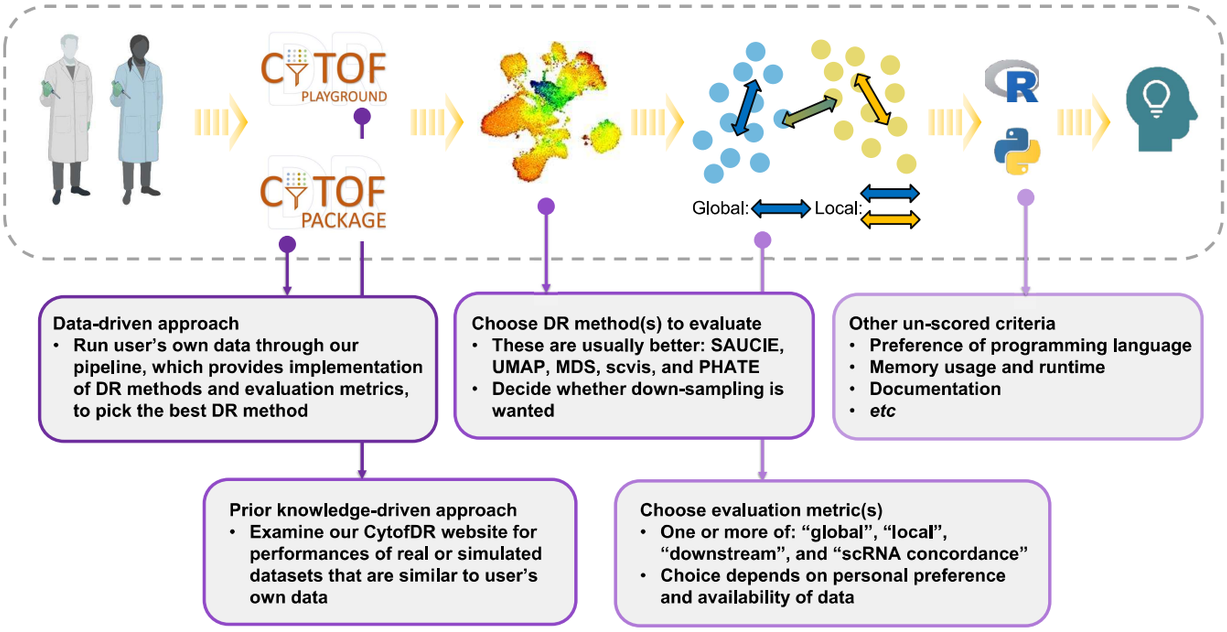
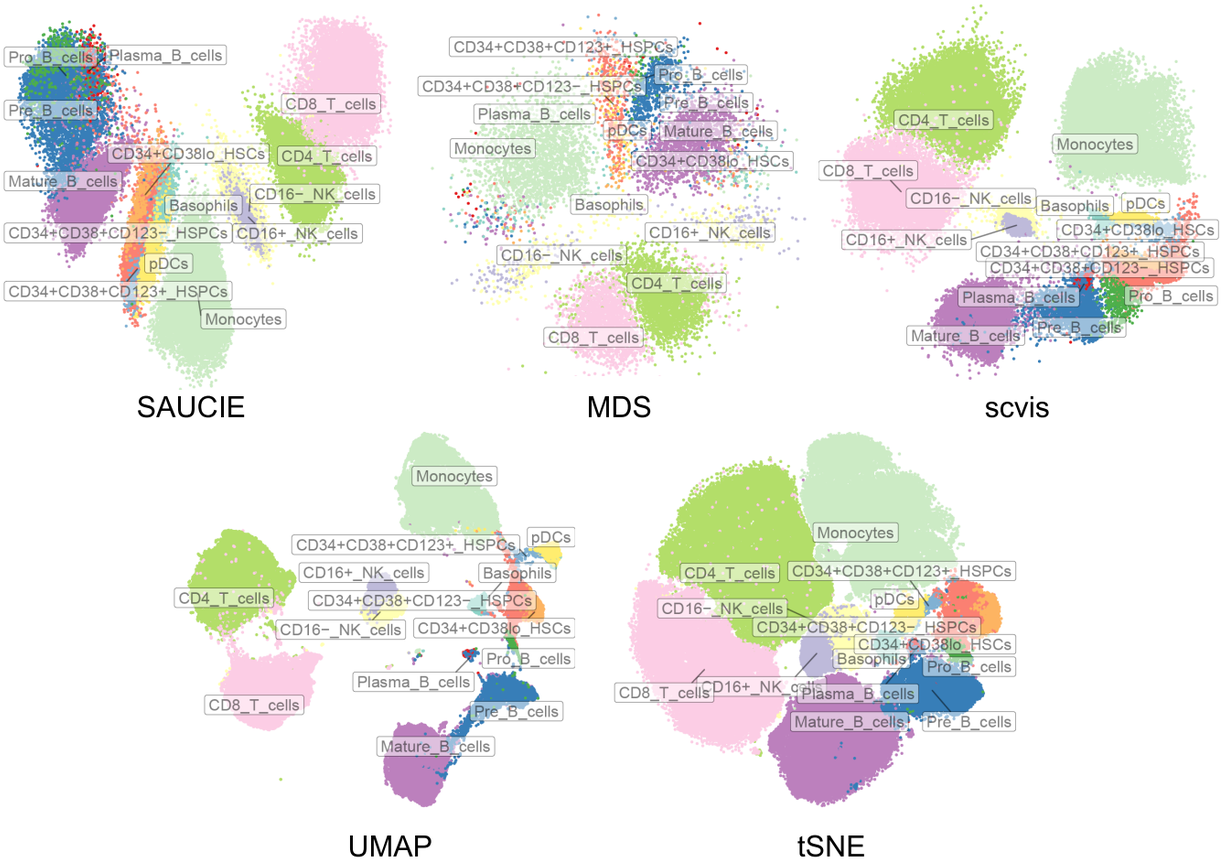


Fig. 6

(a)


CYTOF PLAYGROUND
CYTOF PACKAGE
Global:
Local:
Data-driven approach
• Run user's own data through our pipeline, which provides implementation of DR methods and evaluation metrics, to pick the best DR method
Choose DR method(s) to evaluate
• These are usually better: SAUCIE, UMAP, MDS, scvis, and PHATE
• Decide whether down-sampling is wanted
Other un-scored criteria
• Preference of programming language
• Memory usage and runtime
• Documentation
• etc
Prior knowledge-driven approach
• Examine our CytofDR website for performances of real or simulated datasets that are similar to user's own data
Choose evaluation metric(s)
• One or more of: "global", "local", "downstream", and "scRNA concordance"
• Choice depends on personal preference and availability of data

(b)


Pro_B_cells
Plasma_B_cells
Pre_B_cells
CD8_T_cells
CD34+CD38lo_HSCs
CD4_T_cells
Mature_B_cells
CD16−_NK_cells
Basophils
CD34+CD38+CD123−_HSPCs
CD16+_NK_cells
pDCs
CD34+CD38+CD123+_HSPCs
Monocytes
SAUCIE
MDS
scvis
UMAP
tSNE