

ImmCellTyper: an integrated computational pipeline for systematic mining of Mass Cytometry data to assist deep immune profiling

Reviewed Preprint

v2 • August 14, 2024

Revised by authors

Reviewed Preprint

v1 • April 5, 2024

Jing Sun , Desmond Choy, Nicolas Sompairac, Shirin Jamshidi, Michele Mishto, Shahram Kordasti 

Centre for Inflammation Biology and Cancer Immunology & Peter Gorer Department of Immunobiology, King's College London, SE1 1UL London, United Kingdom • School of Cancer and Pharmaceutical Sciences, King's College London, London, United Kingdom • Research group of Molecular Immunology, Francis Crick Institute, NW1 1AT London, United Kingdom • Haematology Department, Guy's Hospital, London, United Kingdom • Department of Clinical and Molecular Sciences, Università Politecnica delle Marche, Ancona, Italy

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Mass cytometry, also known as Cytometry by time-of-flight (CyTOF), is a cutting-edge high-dimensional technology for profiling marker expression at the single-cell level. This technology significantly advances clinical research in immune monitoring and the interrogation of immune cell populations. Nevertheless, the vast amount of data generated by CyTOF poses a daunting challenge for analysis. To address this, we describe ImmCellTyper (<https://github.com/JingAnyaSUN/ImmCellTyper>), a novel and robust toolkit designed for CyTOF data analysis. The analytical framework incorporates an in-house developed semi-supervised clustering tool named BinaryClust, which first characterises main cell lineages, followed by in-depth interrogation for population of interest using unsupervised methods.

BinaryClust was benchmarked with existing clustering tools and demonstrated superior accuracy and speed across two datasets comprising around 4 million cells, performing as good as manual gating by human experts. Furthermore, this computational pipeline provides a variety of visualization and analytical tools spanning from quality control to differential analysis, which can be tailored to user's specific needs, aiming to provide a one-stop solution for CyTOF data analysis. The general workflow consists of five key steps: 1) Batch effect evaluation and correction, 2) Data quality control and pre-processing, 3) Main cell lineage characterisation and quantification, 4) Extraction and in-depth investigation of cell type of interest; 5) Differential analysis of cell abundance and functional marker expression (supporting multiple study groups). Overall, ImmCellTyper integrates expert's biological knowledge in a semi-supervised fashion to accurately deconvolute well-defined main cell lineages, while also preserving the potential of unsupervised approaches to discover novel cell subsets and providing a user-friendly toolset to remove the analytical barrier for high-dimensional immune profiling.

eLife assessment

ImmCellTyper presents a **useful** toolkit for CyTOF data analysis, integrating BinaryClust for semi-supervised clustering and cell type annotation. The evidence supporting the findings is **convincing**, with appropriate and validated methodology. This tool will be helpful to researchers in immunology and cytometry, offering a robust solution for cell type identification and differential analysis.

<https://doi.org/10.7554/eLife.95494.2.sa2>

Introduction

Mass cytometry or cytometry by time-of-flight (CyTOF) is a powerful high-throughput single-cell technology which employs stable elemental isotopes, as the same manner of fluorophores, to detect cellular proteins of interest. This approach successfully tackles the panel multiplex challenges faced by traditional flow cytometry due to spectral overlap and permits simultaneous measurement of over 40 parameters on millions of cells. To date, CyTOF has been widely applied in basic and translational medical research, such as deep immunophenotyping and characterisation of novel refined cell subsets, immune monitoring of cell-adoptive therapy, and dissecting cell subpopulations from heterogeneous tumour samples¹. Nonetheless, the advantages of CyTOF come with the problem of handling high-dimensional dataset. Traditional gating strategy for flow cytometry, while still serving as the ‘gold standard’ for cell population identification in cytometry data, may not be an optimal option for CyTOF data due to its high-dimensional settings. The high-parametric resolution in CyTOF, aimed at revealing previously undiscovered cell subpopulations, leads to a significant increase in the complexity of gating schemas and hierarchical depth, which makes manual gating extremely labour intensive and time-consuming². Therefore, effective computational toolkits and pipelines are entailed for CyTOF data mining and analysis.

Efforts have been directed toward developing means of clustering algorithms to deconvolute a pool of live cell mixture into distinct cell populations, which facilitates CyTOF data analysis. For instance, unsupervised methods, which include flowSOM³, Phenograph⁴, X-shift⁵, spade⁶, and DensVM⁷ etc., often combine with dimension reduction techniques like t-SNE⁸, UMAP⁹, PCA etc., and require manual annotation of each cluster based on the marker expression patterns indicated by heatmaps. This approach works well to analyse populations in a data-driven manner, with all files concatenated and analysed all at once. Compared to manual gating, it has advantages in terms of convenience, efficiency, and relative unbiasedness from biological preconception, facilitating the detection of novel cell phenotypes for deep phenotyping. However, the unsupervised approaches are not always ideal and suitable for cytometry data. Mathematical clustering does not necessarily have biological meaning, leading to occasional inaccuracies. And several benchmarking studies suggested that the accuracy for these unsupervised tools may not be optimal¹⁰. Additionally, the technical uncertainty of results produced by different clustering approaches remains a conundrum. Even for the same unsupervised method, the discrepancy among different runs without setting a seed, reduces the reproducibility and may cause confusion, particularly for biologists with limited computational knowledge. Moreover, manual validation is also essential, as biological annotation is the inevitable step to provide biological relevant labels for the clusters. However, this process can be time-consuming and subjective, hindering the automation of pipelines. This problem is particularly pronounced for a large marker panel and samples with high heterogeneity, resulting in a higher number of clusters that need to be annotated.

Advances in artificial intelligence have accelerated the development of alternative clustering methods in a supervised manner for cell type inference. These methods consider the ‘ground truth’ or prior knowledge about the marker expression of each given cell types to automatically label each cell. Currently, a couple of semi-automatic methods have been developed, such as Linear Discriminant Analysis (LDA)¹¹, DGCyTOF¹², CyAnno¹³, DeepCyTOF¹⁴, Automated Cell-type Discovery and Classification (ACDC)¹⁵ and Semi-supervised Category Identification and Assignment (SCINA)¹⁶ etc. ACDC¹⁵ and SCINA¹⁶ use a matrix of pre-defined markers for each cell type to annotate the clusters showing the same signature. These methods assume that markers are either expressed or not expressed (binary), which limits their ability to distinguish cell subtypes with similar phenotypes, particularly non-canonical cell types that cannot be easily separated linearly. Alternatively, DeepCyTOF and LDA use a marker expression matrix extracted from manually gated cell types as a training dataset to build a machine learning model for cell type prediction. This approach has higher precision and accuracy compared with aforementioned methods¹⁰. Nonetheless, it can be labour intensive for preparation of the training set manually. Also, these methods are limited in their ability to predict novel cell subsets beyond the pre-gated set of cell types and lack a systematic and comprehensive way to assign the cells which were not identifiable under any of the gated cell types. New solutions have emerged with algorithms such as DGCyTOF¹² and CyAnno¹³, the former adopts a deep learning classification combined with hierarchical stable-clustering methods and an iteration calibration system to identify known cell types and assign novel subsets; while CyAnno is based on a machine-learning framework which allows the integrative modelling of both ‘gated’ and ‘ungated’ cells. Both methods have demonstrated high accuracy in their test datasets, but unfortunately are not widely used by the research community, possibly due to the issue for the hassle of training data preparation, the lack of user-friendliness and implementation challenges for bench researchers.

To address the common drawbacks of current semi-supervised and unsupervised clustering algorithms and preserve their strengths in discovering both canonical and non-canonical cell subsets respectively, we propose a strategy implemented in ImmCellTyper for cell classification named BinaryClust. By considering biologists’ prior biological knowledge and interpretation for canonical cell clusters in a semi-supervised manner, BinaryClust first automatically characterises the main cell lineages in a fast and accurate way. Subsequently, it extracts specific cell types of interest for further clustering using unsupervised algorithms to identify cell subsets including previously unreported non-conventional population. In addition, this R-implemented pipeline takes advantage of SingleCellExperiment class for data management, providing an easy-to-use and organized systematic workflow of CyTOF data handling. The whole pipeline includes quality control and batch effect correction, which helps to effectively pool datasets from different batches, and ensures the robustness for downstream analysis. Meanwhile, modules like dimension reduction, semi-supervised and unsupervised clustering (flowSOM and Phenograph), interactive data visualisation, and statistical testing for complex study design were also incorporated in this pipeline. Compared with existing integrated computational workflow, such as CATALYST¹⁷, CapX¹⁸, Cytokit¹⁹ and ImmunoCluster²⁰ which was developed and maintained by our team, this recently developed toolkit advanced further on coherence, functionality and user-friendliness. Overall, this approach has the potential to facilitate and smooth the investigation of CyTOF-based research.

Results

BinaryClust is comparable with manual gating in quantifying the abundances of main cell lineages

The core concept for BinaryClust is depicted in [Figure 1](#), where a simple user-designed cell type marker expression matrix is required and serves as a reference to characterise the main cell lineages, with positive markers indicated in ‘+’, negative markers in ‘-’, and irrelevant markers in

'A'. K-means (k=2) will be applied to divide the positive and negative cell population of each marker, then align it to the reference table to infer main cell types. This is followed by the extraction of population-of-interest for downstream clustering using unsupervised methods for subpopulation discovery.

To assess the performance of the automated cell type classification and prediction, we generated a test CyTOF dataset using peripheral blood mononuclear cell (PBMC) samples from seven patients with myeloproliferative neoplasm (MPN) and two healthy donors, employing a 37-marker deep immunophenotyping panel (**Supplementary Table 1** [↗](#)). Manual gating was performed by two independent experts which identified seven main cell lineages. These results act as the reference for evaluating the computer-aided methods. The hierarchical sequential gating strategy was explicitly illustrated in **Supplementary Figure 1** [↗](#). We evaluated the agreement between manual gating results and BinaryClust results regarding cell frequencies of each population. The mean value was calculated from manual gating results of two experts to compare with BinaryClust-generated results using Pearson correlation analysis. As shown in **Figure 2A** [↗](#), the two methods exhibit a strong correlation with coefficient (R^2) equals to 1, 1, 0.99, 0.96, 0.96, 0.99, 0.99 for CD4 T cells, CD8 T cells, dendritic cells, NK cells, monocytes, and gamma delta T cells, respectively (all $P < 0.0001$). And most of the data points remain close to the line of equality (red line, $R^2=1$), indicating a high degree of agreement. Meanwhile, the Bland-Altman plots in **Figure 2B** [↗](#) also suggest no consistent bias of manual gating versus BinaryClust across all the identified cell types. The good performance of BinaryClust was further validated in the influenza dataset published by our group²¹ [↗](#), which contains FCS files from 11 individuals with 6 main immune cell types detected, the average F-measure and ARI reached 0.98 and 0.91 respectively (**Supplementary Figure 2A-B** [↗](#)).

BinaryClust achieves high accuracy and speed compared with flowSOM and LDA

To further evaluate BinaryClust's performance, we compared it with the well-performing unsupervised algorithm flowSOM and supervised classifier Linear discriminant analysis (LDA). FlowSOM was run on the same MPN dataset with k value set as 20, followed by manual annotation and cluster merging to identify the same cell populations as in manual gating and BinaryClust. The cell frequencies derived from BinaryClust, flowSOM, and manual gating (expert1 and expert2) were compared using interaction plot (**Figure 3** [↗](#)). We observed that BinaryClust remains consistent with manual gating (all $P > 0.05$), whereas flowSOM identified significantly less Gamma Delta T cells and dendritic cells ($P < 0.001$ and $P = 0.006$, respectively) compared with the other three measurements. We also increased the initial k value to 40 to improve accuracy by over clustering then merging clusters³ [↗](#). We found that the average frequency for Gamma Delta T cells and dendritic cells increased to 1.5% and 4.65% respectively but still remained significantly different ($P < 0.05$) from the other two methods. It is also interesting to see that although the same gating strategy was applied, different experts obtained results with slight variation despite no statistical significance. The same findings were confirmed using boxplot as shown in **Supplementary Figure 3** [↗](#).

Subsequently, LDA was also tested on the same dataset. Since this method requires a training dataset from manual gating to build a model, we exported the cell events and cluster assignment from Cytobank, then equally partitioned all cells into training and test dataset.

Due to the difference in method implementation, we did not include LDA for cell abundance quantification comparison with other methods, since only half of the cells were used for prediction. To ensure an equitable comparison and further evaluate the accuracy, F-measure, and Adjusted Rand Index (ARI) were calculated using manual gating cluster IDs as reference. As indicated in **Table 1** [↗](#), for the MPN dataset, BinaryClust achieved high F-measure and ARI in all 7

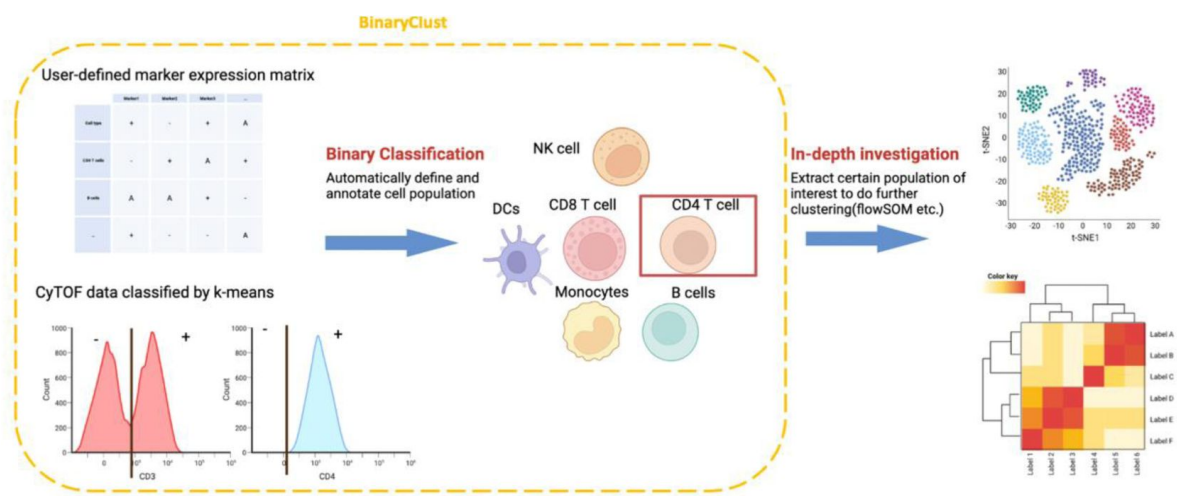


Figure 1

Schematic diagram of BinaryClust framework. Semi-supervised classification is first performed on selected markers in the user-defined marker expression matrix to classify and annotate major cell types. Population-of-interest can be further extracted and explored using unsupervised clustering methods followed by differential analysis. Figure was generated by BioRender (<https://www.biorender.com/>).

© 2024, BioRender Inc. Any parts of this image created with BioRender are not made available under the same license as the Reviewed Preprint, and are © 2024, BioRender Inc.

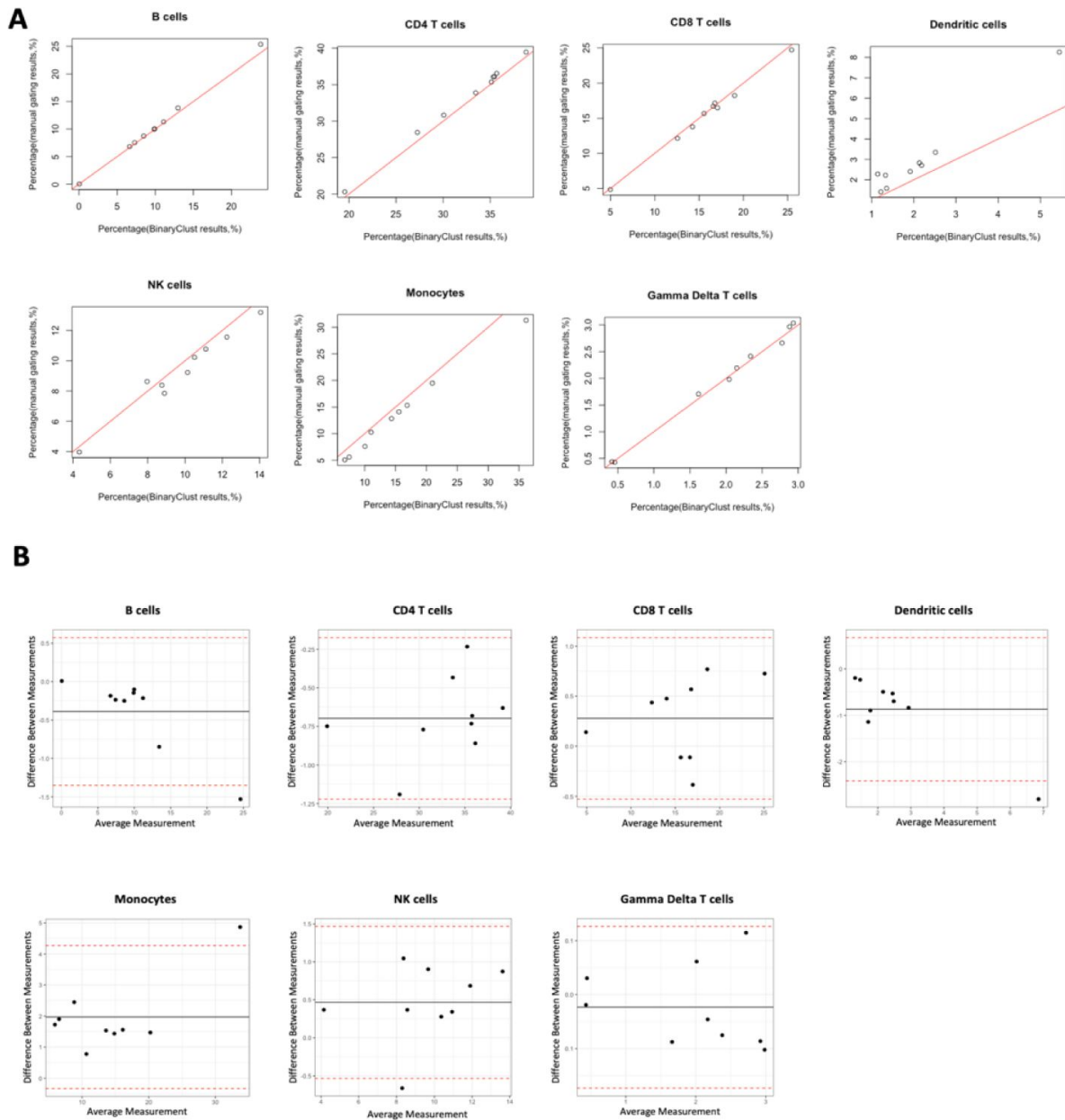


Figure 2

Agreement evaluation comparing manual gating and BinaryClust in MPN cohort (n=9). Manual gating of B cells, CD4 T cells, CD8 T cells, dendritic cells, NK cells, monocytes and gamma delta T cells were performed by two independent experts using Cytobank, and mean values of the population percentages were calculated to compare with BinaryClust results. Each dot represents one patient sample. (A) Scatter plot showing the correlation between the two methods, with the red line indicating perfect agreement (correlation coefficient = 1); (B) Bland-Altman plots of the two measurement methods among all the cell populations, with the black line suggesting the mean observed difference and red dotted lines indicating limits of agreement (1.96 x standard deviations).

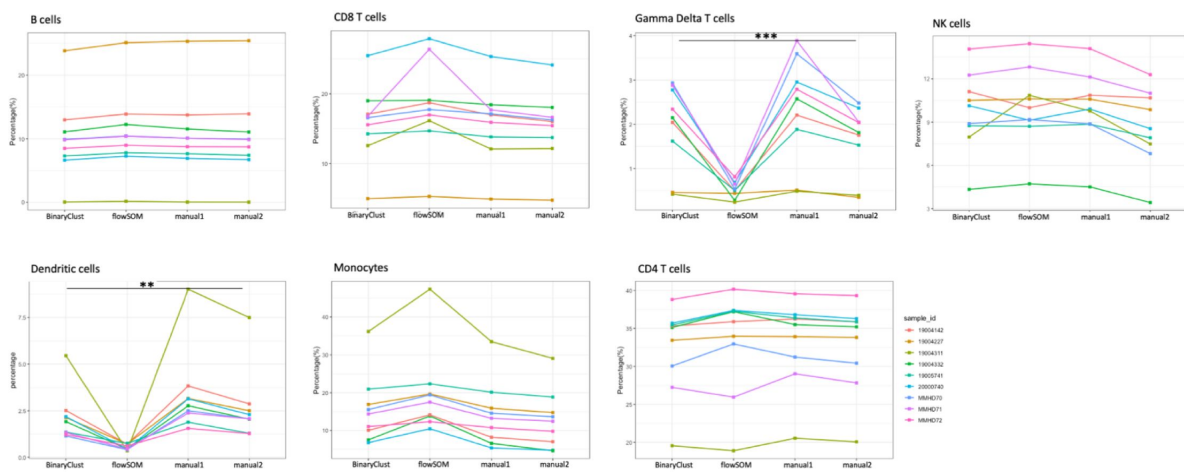


Figure 3

Comparison of manual gating (manual1 and manual2), BinaryClust and flowSOM clustering results in MPN cohort (n=9). Interaction plots showing the individual measurement (percentage) of each study participant with indicated colors by different methods across main cell lineages (B cells, CD8 T cells, Gamma Delta T cells, NK cells, Dendritic cells, monocytes and CD4 T cells); ANOVA was used for statistical testing, and significance was marked by asterisk. * P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

cell types with an average of 0.94 and 0.91 respectively. The performance remains excellent in influenza dataset as well (average F-measure = 0.98, **Supplementary Figure 2C** [↗](#)). LDA has equivalent prediction accuracy as BinaryClust with an average F-measure of 0.93 and ARI of 0.90. Both supervised methods outperformed flowSOM (average F-measure=0.75, average ARI=0.66), which is in line with previous benchmarking study on evaluating supervised and unsupervised clustering algorithms¹⁰ [↗](#). Notably, for cell types that constitute a substantial proportion in the pool like CD4 T cells, CD8 T cells, NK cells, monocytes, and B cells, flowSOM can identify them with high precision and sensitivity, while this is not the case for dendritic cells and gamma delta T cells which account for less than 3% of the whole population (F-measure = 0.51 and 0.15, respectively). Even so, the overall F-measure and ARI for flowSOM considering cell proportion reached 0.81 and 0.80.

It has been demonstrated that flowSOM and LDA are among the fastest clustering algorithms without compromising their performance. Here, we compared the speed amongst the three approaches in MPN, Influenza and COVID-19 datasets²² which contain 2,231,053, 210,933 and 3,862,628 cells respectively. Since LDA requires training data to be executed, we did not run it on COVID-19 dataset due to the absence of manual gating results. As shown in **Figure 4** [↗](#), BinaryClust exhibited the highest speed in both MPN and Influenza dataset but fell behind flowSOM slightly in COVID-19 dataset.

ImmCellTyper pipeline supports interactive data visualisation and comparison among study groups

BinaryClust is an important component of ImmCellTyper, a comprehensive integrated pipeline designed for systematic CyTOF data mining. Hence, we utilised the visualisation functions to further prove the reliability and robustness of BinaryClust.

BinaryClust inherently considers CyTOF makers as binary distributed which is in most of the case, but not always. Therefore, it is crucial to check the marker behaviour before running the pipeline. As shown in **Figure 5A** [↗](#), all markers selected for the classification matrix (**Figure 5B** [↗](#)) displayed a binary distribution implying the suitability for this pipeline. After clustering (**Figure 5C** [↗](#)), median marker expression heatmap (**Figure 5D** [↗](#)) can check the reliability of the results before proceeding for downstream analysis. In MPN dataset, we projected the cluster assignment resulted from BinaryClust and manual gating to UMAP (**Figure 5E** [↗](#)), along with the expression of the phenotypic markers (**Figure 5F** [↗](#) and **supplementary Figure 4A** [↗](#)). High similarity was observed between BinaryClust and manual gating, whereas slight difference was found on flowSOM results coloured UMAP on islands of CD8 T cells and CD4 T cells (**Supplementary Figure 4A** [↗](#)). FlowSOM appeared to classify cells that were close on spatial distance into the same cluster, in contrast to prior knowledge-based methods: BinaryClust and manual gating.

Application of ImmCellTyper pipeline to the COVID-19 dataset demonstrates its versatile functionalities for comprehensive data analysis

To showcase the analytical and visualisation functions of ImmCellTyper pipeline, we used the dataset published by Chevrier et al.²² [↗](#), which described the immune signature of mild and severe COVID-19 patients in comparison with healthy individuals. There are a total of 82 FCS files with a 40-plex marker panel focusing on innate immunity in this dataset. An initial marker expression check was carried out based on the user-defined matrix (**Figure 6B** [↗](#)) and displayed in **Figure 6A** [↗](#). BinaryClust was then performed and identified 12 cell populations as expected. We used t-distributed stochastic neighbour embedding (t-SNE) in this dataset for dimension reduction, coloured by cell types and faceted by disease conditions, which exhibited substantial immune

Table 1

Precision, recall, F-measure, and ARI of indicated clustering methods.

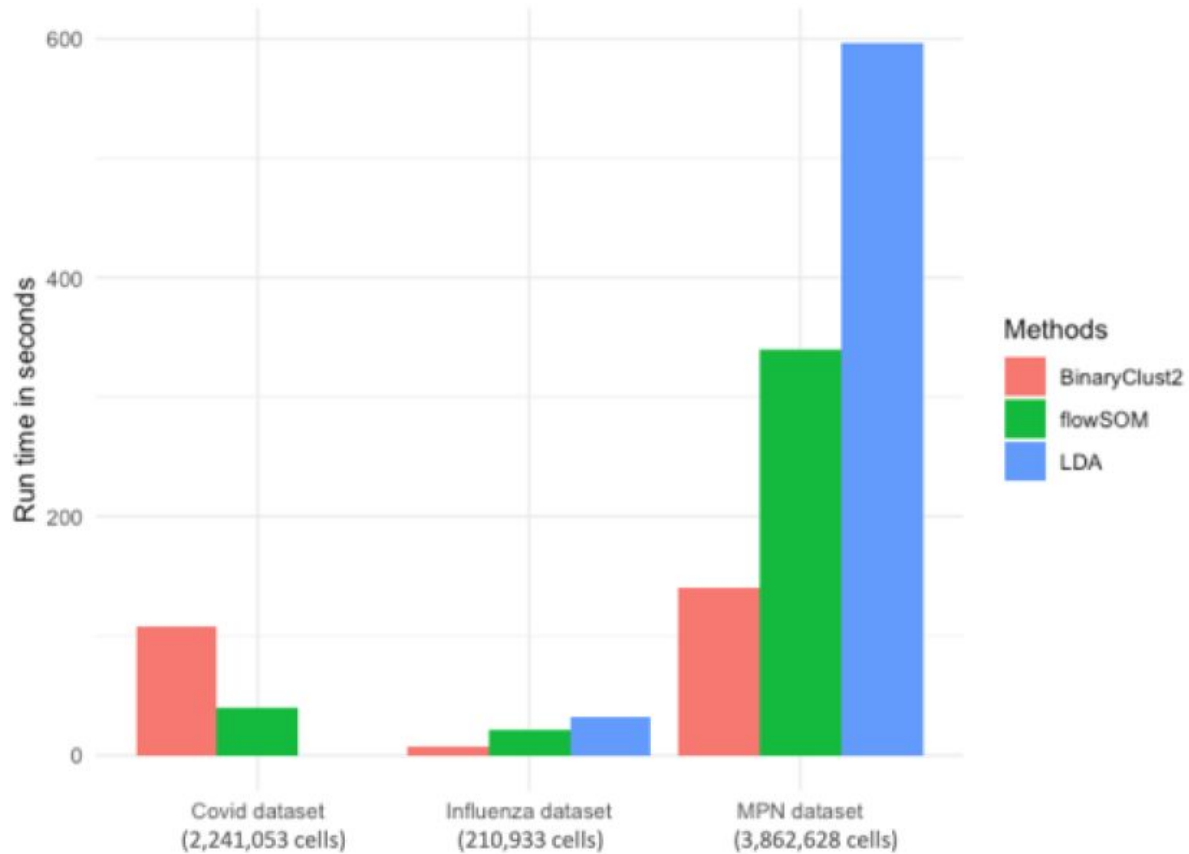
ARI, Adjusted Rand Index; LDA, Linear discriminant analysis

	Gated population	Counts	Cluster Cell Counts	True Positive	Precesion	Recall	F-measure	Average F-measure	ARI	Average ARI
Bina Clust2	CD4 T cells	1270041	1232832	1226712	1.00	0.97	0.98	0.94	0.95	0.91
	CD8 T cells	601995	601639	582402	0.97	0.97	0.97		0.95	
	NK cells	389261	380999	376561	0.99	0.97	0.98		0.97	
	Monocytes	576624	625735	568924	0.91	0.99	0.95		0.92	
	B cells	397279	380170	374501	0.99	0.94	0.96		0.95	
	Dendritic cells	135894	86012	84130	0.98	0.62	0.80		0.74	
	TCRgd T cells	87194	73599	73562	1.00	0.84	0.92		0.91	
flowSOM (k=20)	CD4 T cells	1270041	1268929	1226712	0.97	0.97	0.97	0.75	0.90	0.66
	CD8 T cells	601995	692374	571344	0.83	0.95	0.89		0.81	
	NK cells	389261	393646	365403	0.93	0.94	0.93		0.91	
	Monocytes	576624	792447	568284	0.72	0.99	0.85		0.73	
	B cells	397279	405886	390077	0.96	0.98	0.97		0.96	
	Dendritic cells	135894	20832	18254	0.88	0.13	0.51		0.22	
	TCRgd T cells	87194	19720	4967	0.25	0.06	0.15		0.08	
LDA	CD4 T cells	634733	665481	632700	0.95	1.00	0.97	0.93	0.93	0.90
	CD8 T cells	301138	327469	299309	0.91	0.99	0.95		0.92	
	NK cells	194470	202110	187259	0.93	0.96	0.94		0.92	
	Monocytes	288338	301387	282836	0.94	0.98	0.96		0.94	
	B cells	198582	211090	195908	0.93	0.99	0.96		0.94	
	Dendritic cells	68152	90456	61785	0.68	0.91	0.79		0.75	
	TCRgd T cells	43405	43064	39976	0.93	0.92	0.92		0.92	

ARI, Adjusted Rand Index; LDA, Linear discriminant analysis

Figure 4

Comparison of BinaryClust, flowSOM and LDA on speed. Bar chart showing runtime (in seconds) of the three methods in three different datasets.



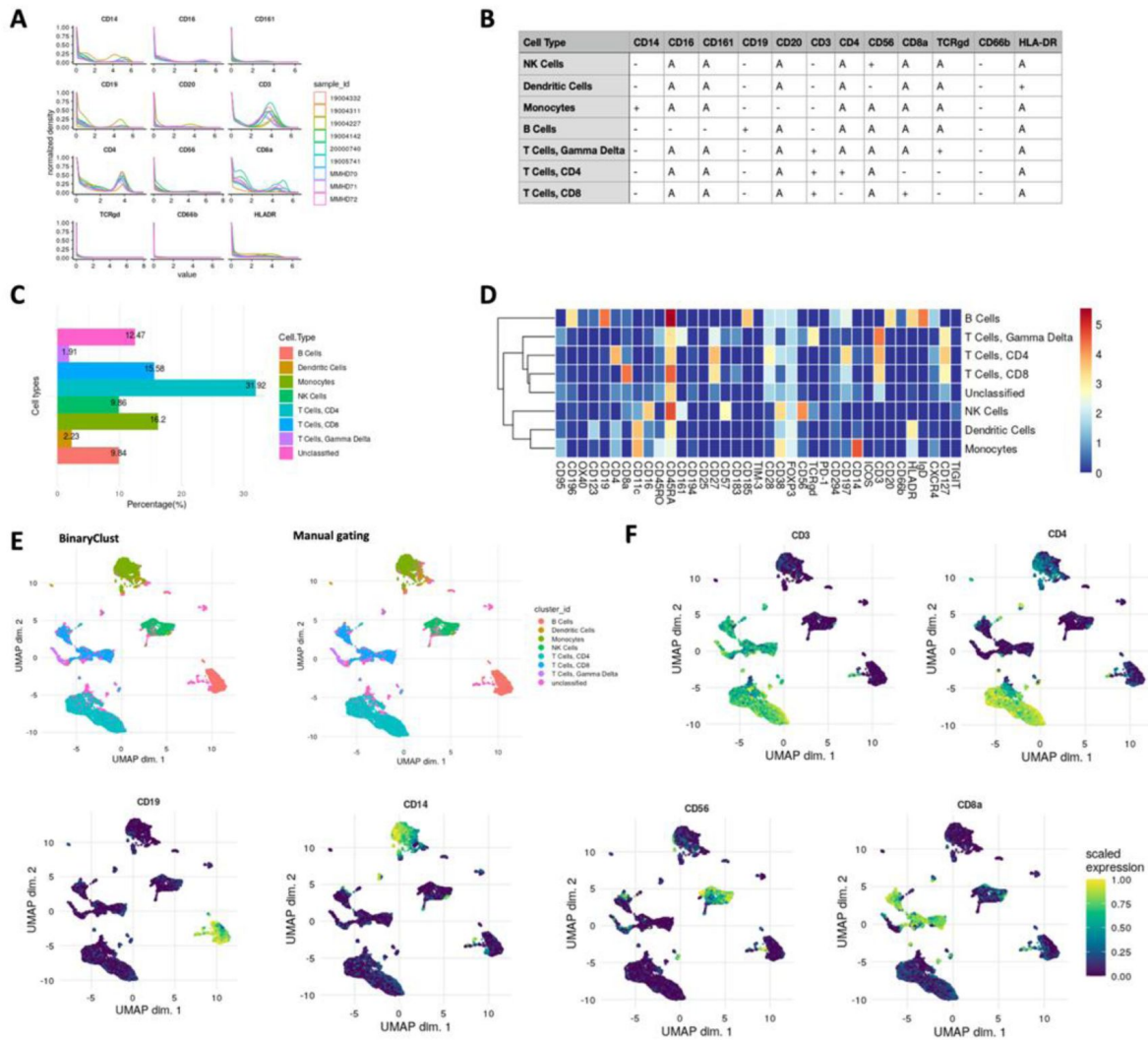


Figure 5

Cell type characterization and visualisation using ImmCellTyper pipeline in MPN dataset (n=9). (A) Intensity distribution of selected phenotypic markers used for BinaryClust classification, coloured by sample_id; (B) Pre-defined expression classification matrix for the MPN dataset, '+' indicates positive, '-' indicates negative and 'A' suggests 'any'; (C) Proportion of the main cell lineages of all cells in the concatenated FCS files after classification; (D) Median marker expression heatmap of BinaryClust classification results; (E) UMAP plot of random downsample of 2000 cells per patient coloured by main cell types based on BinaryClust classification (left) and manual gating results (right). (F) UMAP plots coloured by normalized expression of indicated markers (CD3, CD4, CD8a, CD20, CD19, CD14, and CD56) across 2000 cells per sample.

alteration among healthy control, mild and severe COVID-19 patients (**Figure 6C** [↗](#)). From the heatmap in **Figure 6D** [↗](#), we can have an overview of the marker expression of each population, which remains consistent with our initial definition indicated in the expression matrix.

The abundances of the identified cell types were quantified for each individual study participant in the format of stacked histogram (**Figure 7A** [↗](#)) and summarized in boxplot (**Figure 7B** [↗](#)). Compared with healthy volunteers, there are significant immune alterations in COVID-19 patients in B cells, Basophils, cDCs, Monocytes, NK cells, CD8 T cells, Neutrophils and pDCs (all $P < 0.05$), which depicts a similar trend to the original paper, despite slight discrepancies caused by different statistical methods used. Kruskal Wallis test was conducted for this dataset followed by multiple testing correction (Benjamini-Hochberg procedure) and Dunn's test for post hoc analysis, while Chevirer et al. used Mann-Whitney Wilcoxon test corrected by Holm method. We then selected 9 markers (IL-6, PD-L1, VISTA, IDO, TIM-3, TMEM173, GranzymeB, PPARg, Ki-67) as the state markers which can reflect the functional or proliferative status of the immune cell types. Notably, Granzyme B was observed to be significantly highly expressed in COVID-19 patients versus healthy control across all identified immune cell types (all $P < 0.05$), TMEM173 was substantially up-regulated in COVID-19 patients particularly in monocytes, neutrophils, and CD4 T cells; and PD-L1 expression remained low for all cells, indicating the immune system of COVID-19 patients were highly activated without exhaustion regardless of disease severity, which was not explored previously. Since monocytes and neutrophils were of particular interest in the original paper, and the panel also included specific markers for in-depth interrogation, we then extracted the two population and carried out Phenograph ($k=60$) to investigate the subclusters. As shown in **Figure 7D** [↗](#), Phenograph returned 16 and 14 subclusters respectively for monocytes and neutrophils, with heatmap provided in **Supplementary Figure 5** [↗](#).

Reanalysis of the COVID-19 dataset demonstrated the concordance of ImmCellTyper pipeline with original reports with additional findings, as well as its versatile interactive data visualisation functionalities.

Discussion

In this study, we present an analytical pipeline named ImmCellTyper for systematic exploration of CyTOF data. This pipeline addresses a comprehensive range of analytical needs encompassing data quality check, batch effects examination/correction, cell type identification, and downstream differential analysis accompanied by high-quality, publishable data visualisations. Furthermore, ImmCellTyper includes an in-house developed, knowledge-based, semi-supervised classifier BinaryClust with high accuracy and speed, and also integrates the well-performing and state-of-the-art unsupervised algorithms (Phenograph and flowSOM). By adopting the strategy of first obtaining the main cell types and then select specific cell types for in-depth interrogation with higher clustering resolution, ImmCellTyper combines the advantages of both supervised and unsupervised clustering algorithms for discovery of both major populations and refined subpopulations. Designed for ease of use, this pipeline features a clear and user-friendly workflow, and is compatible to the widely used pipeline CATALYST, aiming to provide a one-stop solution for CyTOF users.

For validating the robustness of the automated cell characterization function of ImmCellTyper: BinaryClust, we involved two independent datasets (MPN dataset and Influenza dataset) and compared with existing clustering tools flowSOM and LDA. Previous Benchmarking studies indicated that flowSOM was one of the best-performing unsupervised tools in precision, speed and stability, with F-measure ranges from 0.58~0.90 in various datasets¹⁰ [↗](#), and was widely used in high-impact research publications^{10,23} [↗](#). However, the results of flowSOM vary upon the k values set by the users which causes some uncertainty and impairs reproducibility of the results. In contrast, semi-supervised clustering methods like ACDC, DeepCyTOF¹⁴ [↗](#) and LDA

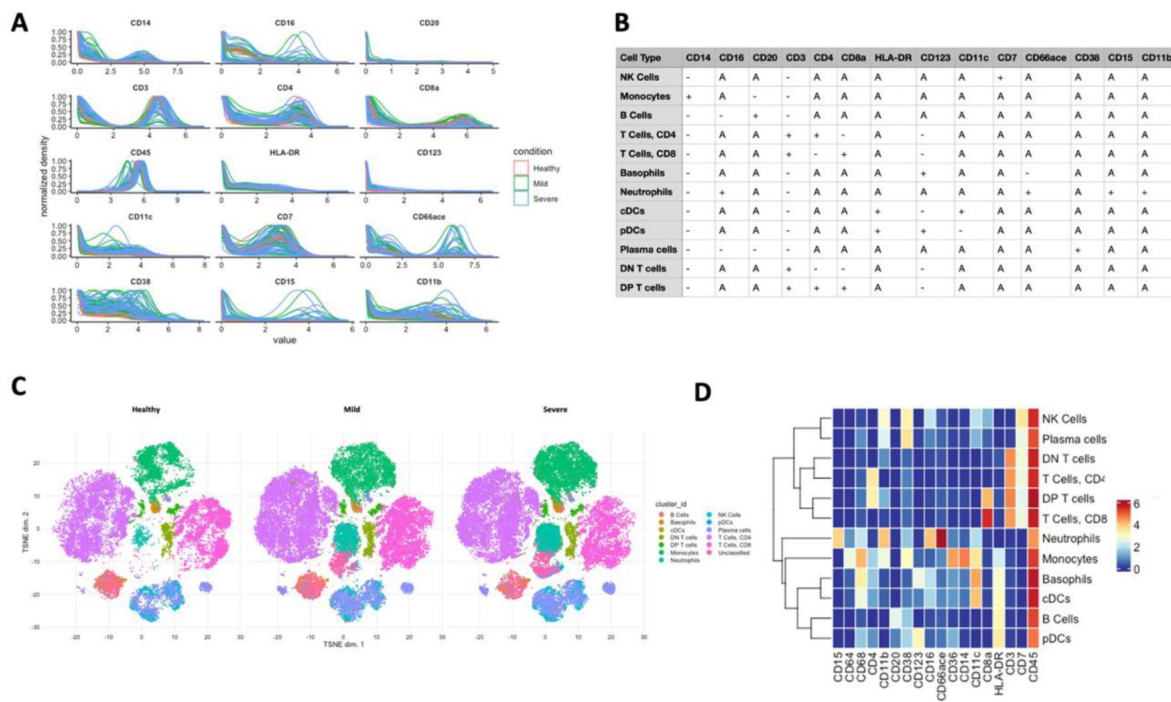


Figure 6

Applying ImmCellTyper pipeline on COVID-19 patient dataset (n=82) published by Chevrier et.al. (A) Marker intensity distribution of selected phenotypic markers used for BinaryClust classification, coloured by disease severity (n=22 healthy individuals, 28 mild COVID-19 patients and 38 severe COVID-19 patients); (B) Pre-defined marker expression classification matrix used for BinaryClust; (C) T-SNE plots, with 1000 cells per sample, were coloured by the main cell types generated by BinaryClust and faceted by different study groups; (D) The corresponding median marker expression heatmap of BinaryClust results for the COVID-19 dataset.

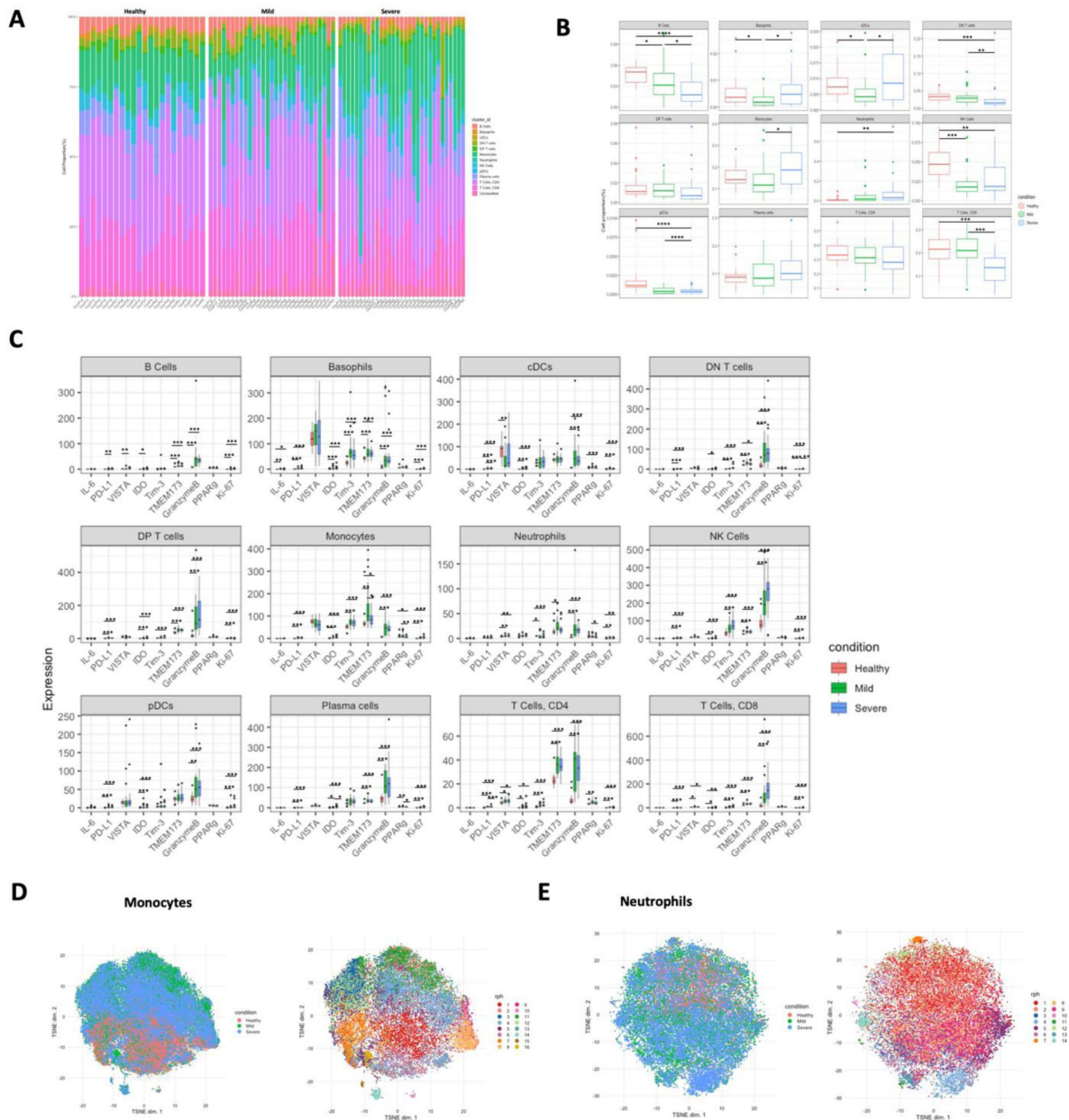


Figure 7

Quantification and statistical analysis comparing the study conditions in COVID-19 dataset (n=82). (A) Stacked histogram of main cell type composition per individual generated by BinaryClust, and grouped by study conditions (healthy, mild and severe); (B) Boxplots representing cell abundance frequencies among the study conditions, faceted by different main cell types; (C) State marker expression intensities with comparison of the study groups across the main cell types; (D) Clusters of monocytes and neutrophils were extracted from the whole cells for downstream interrogation. T-SNE plots with random downsample of 1000 monocyte cells and (E) neutrophils per sample were coloured by study conditions and Phenograph clustering results (k=60), respectively. Statistical significance was marked by asterisk. * P<0.05, **P<0.01, ***P<0.001, ****P<0.0001

outperformed the unsupervised methods in the same benchmarking study. The F-measure obtained by ACDC and LDA ranged from 0.78~0.99, which was significantly higher than unsupervised approaches including Accense, PhenoGraph, Xshift, k-means, flowMeans²⁴, FlowSOM and DEPECHE²⁵ (F-measure: 0.28~0.93)²³. In this study, we chose LDA because it has proven to be superior to ACDC and DeepCyTOF in terms of precision and speed^{10,11}. The excellent performance of LDA was demonstrated in MPN dataset of this study with F-measure reached 0.93 for around 4 million cells. BinaryClust was comparable in accuracy (F-measure=0.94) but much faster in speed. Both semi-supervised approaches outperformed flowSOM as expected. Although semi-supervised methods seem to perform better, the mainstream for CyTOF data analysis still relies on unsupervised methods. The reason for this might be that most of the semi-supervised methods rely on manual gating as reference and requires either the manually gated expression matrix or user-defined binary matrix as training data. This process takes extra time and efforts especially for defining subpopulations, and may introduce end-user bias. Additionally, the common restriction for all the supervised methods is that they have limited capability to reveal novel subsets as those are not predefined in the training pool, which poses a critical challenge for the pursuit of novel discoveries. Considering all these strength and limitations, as a semi-supervised method, BinaryClust possesses the advantages of incorporating prior biological knowledge, being reproducible across runs and accurate on cell type identification. Meanwhile, the input requirement is a simple matrix of cell population definition based on marker positivity, avoiding the hassle to manually gate example FCS files. With a deep understanding that certain CyTOF markers are expressed on a continuum rather than binary manner, this automated supervised approach will only be used for classification of main cell lineages. The test datasets in this study contained data from human PBMC samples which represented a heterogenous pool of immune cells, covering the major types of various immune populations, and demonstrated excellent performance of BinaryClust on classifying CD4 T cells, CD8 T cells, gamma delta T cells, NK cells, monocytes, dendritic cells, and B cells. We also recommend checking marker expression distribution before running the pipeline to ensure accurate results. After the above step, this pipeline supports cell population extraction for further dissection into subclusters using unsupervised approaches, which excel in detecting rare and refined subpopulations. Phenograph is a particularly effective tool for this purpose. By applying this strategy, our previous work with BinaryClust facilitated the evaluation of the impact of systematic anti-cancer agents on lymphocyte population in non-small cell lung cancer patients, as well as the characteristics of autologous T cell products after manufacturing²⁶.

Another purpose behind the development of the ImmCellTyper framework is to cater to general analysis needs, spanning from pre-processing to downstream differential analysis. There are existing well-established pipelines like CATALYST¹⁷, diffcyt²⁷, Cytokit¹⁹, and ImmunoCluster²⁰ etc., each of them has its own strength and limitations. One of the challenges existed in CyTOF data pre-processing is effectively integrating multiple batches, given the technical differences arising from experiments and instrumental acquisition, which could result in the separation of clustering across batches and potentially confounding the signal of interest. In ImmCellTyper pipeline, we provide the functionality for batch effects examination and incorporated two well-performing batch correction algorithms CytoNorm²⁸ and CytofRUV²⁹ into our framework prior to downstream analysis, which ensures the quality of the data and, to the best of our knowledge, makes it the first integrated pipeline with this module. For the downstream analysis, tools like diffcyt and CATALYST group markers into phenotypic and state makers, enabling the detection of differentially abundant cell clusters and differential expression of functional markers within each cell population²⁷. This well-recognised strategy has been extensively applied in various studies. But one limitation is that it directly classifies cells into high-resolution clusters, and cannot automatically merge subclusters into one major population with similar phenotypes²⁷. Our pipeline addresses this problem via a precise and convenient solution: using the semi-supervised classifier (BinaryClust) of ImmCellTyper. The two pipelines (CATALYST and ImmCellTyper) are well compatible, allowing users to leverage their functions together and providing a broader range of analytical options. On the other hand, Cytokit has the

advantages of the graphical user interface (GUI) for non-specialists, integration of a variety of clustering (DensVM⁷, ClusterX, FlowSOM, Phenograph), dimension reduction (PCA, ISOMAP, t-SNE) methods, and inference of the relatedness among cell populations, but it does not support complex study design, group comparison and statistical testing¹⁹. Several methods have been developed to fit with the high-dimensional features of CyTOF data for differential analysis including diffcyt, CellCnn³⁰, cydar³¹, citrus³², cyEMD³³ etc. Nonetheless, none of them is applicable to compare more than two study groups at once which involves multiple testing. Therefore, to accommodate study designs involving multiple groups, ImmCellTyper does statistics via first applying Kruskal Wallis test, followed by Benjamin-Hochberg (BH) procedure for multiple testing correction and post hoc analysis via Dunn's test or pairwise-Wilcoxon test. In cases where the comparison involves only two groups, Mann-Whitney test will be applied in terms of cell frequency, as the distribution of CyTOF data generally does not fit for normal distribution.

One limitation of this computational pipeline is that its semi-supervised cluster identification may not be well-suited for the direct identification of specific subpopulations which are defined by continuum markers. This is due to the method's reliance on the presumption that markers are binary distributed. In addition, for users with general interest in characterizing subclusters of each main cell lineage, it can be laborious to first perform BinaryClust then extract every population for unsupervised clustering. In such scenarios, BinaryClust can be used in parallel with other clustering methods or pipelines like CATALYST. This allows users to quickly obtain additional information about main cell types with the accuracy comparable to manual gating.

In summary, we introduce a novel open-source R-implemented strategy and a versatile toolbox for CyTOF data analysis. The future direction is to automate the data clean-up, compensation, and bead normalisation steps. Additionally, we also aim to implement the whole pipeline into Python, a more accessible programming language for bioinformatics novices and biologists who wish to perform high-dimensional data analysis independently.

Materials and methods

BinaryClust

Most of the CyTOF markers exhibit log-normal or bi-modal distribution with zero inflation after arcsinh transformation. Thus, we employ binary classification using k-means to group cells into negative and positive populations (k=2) for each marker indicated in the user-defined classification matrix. Here, k-means is an unsupervised clustering algorithm to cluster the data based on the Euclidean distance among points, which is calculated by the formula below:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

Then assign each data point into a cluster centroid which is denoted by c_i , and $\text{dist}()$ is the Euclidean distance:

$$\arg \min_{c_i \in C} (c_i, x)^2$$

By aligning the k-means results with the user-designed classification matrix, cell populations can be subsequently classified and annotated.

R package ImmCellTyper

This computational pipeline is implemented in the R package ImmCellTyper and publicly available on Github (<https://github.com/JingAnySun/ImmCellTyper>). Instructions for package installation and function usage can be found in the README file on the Github page. It is recommended to first

examine the unwanted non-biological variation across batches and perform additional batch normalisation if necessary. Afterwards, import the data into the second part of the pipeline for downstream analysis. Users are required to prepare all the FCS files, sample metadata containing the details and grouping information of each sample, panel metadata with the information of the antibody panel with metal tags used in the experiments, and cell type classification matrix with phenotypic marker expression in a binary manner of each cell lineage. All files need to be in the right format to use the pipeline.

ImmCellTyper workflow overview

ImmCellTyper pipeline is composed of seven steps corresponding to two separate sub-pipelines, as described in **Figure 8** and Github vignettes (<https://github.com/JingAnySun/ImmCellTyper/tree/main/vignettes>):

1. Sub-pipeline1 (corresponds to **Figure 8**, workflow step 2): Batch effects evaluation and correction. Batch effects occur when samples were collected and measured at different sites or time points, especially for large-scale studies. It is crucial to remove the unwanted variation which might interfere the true biological differences. After bead normalisation and data clean-up (**Figure 8**, workflow step 1), which can be performed using third-party platforms, such as CyTOF v7.0 system control software and Cytobank, users can systematically examine batch effects on two levels including marker behaviours and clustering results based on the method introduced by Trussart et al.²⁹ If needed, CytoNorm and CytofRUV, which are well-performing correction algorithms, can be used to align the existing batch effects.
2. Sub-pipeline2 (corresponds to **Figure 8**, workflow step 3-7): semi-supervised classification, differential analysis, and in-depth investigation. When data are cleaned and normalised, they can be imported into the second part of the pipeline, constructed into a SingleCellExperiment (SCE) object, and undergo semi-supervised classification to identify the major cell types and test the differential frequencies or state marker expression among study groups. After that, if the users have certain interests of specific cell types and pre-design the panel for that, or in another circumstance, the initial statistics draw the user's attention into a certain cell type, intuitively, further clustering using unsupervised tools like flowSOM or Phenograph should be conducted with an increased cluster resolution and deeper investigation for cell subsets, after extracting the cell population of interest. ImmCellTyper pipeline has the same data storage and infrastructure as CATALYST, therefore all the functions in CATALYST can be seamlessly used in ImmCellTyper, tailored to user's analytical needs. We do not elaborate on the basic functions of CATALYST, which can be found in the tutorial vignettes of the package (<https://github.com/HelenaLC/CATALYST>).

Batch correction algorithms

The batch effect correction algorithms embedded in the function of 'batchNorm' comprises CytoNorm, as described by Van Gassen et al.²⁸ and CytofRUV by Trussart et al.²⁹ Both algorithms rely on anchors (reference samples/technical replicates) across batches to perform normalisation. CytoNorm uses flowSOM clustering to first identify clusters prior to a population-specific transformation on the reference samples by computing the quantile values and aligning them with splines, whereas CytofRUV applies remove unwanted variation III (RUV-III) to CyTOF data by estimating and eliminating the non-biological variation of the pseudo-replicates.

Workflow steps	Description
Data clean-up and pre-processing	Bead normalisation is first performed on the raw FCS files to correct signal fluctuation of the CyTOF instrument, then FCS files are imported to cytobank for clean-up to exclude doublets and debris etc.
Batch effect evaluation and correction	Samples across different batches will be evaluated on both marker expression level and clustering level for batch effects; If needed, batch correction will be performed using function batchNorm, which provides two well-performing algorithms CytoNorm and CyTOFRUV.
Data transformation and SCE object construction	FCS files and relevant metadata for(samples and panel) will be integrated into one sce object, and FCS data will be transformed using co-factor 5.
Binary classification	BinaryClust will be subsequently performed based on the classification matrix designed by the user.
Differential analysis and population extraction	Differential cell abundance analysis and statistical comparison will be conducted. Then the user can quickly gate certain population of interest for down-stream in-depth analysis.
In-depth interrogation of population of interest	Unsupervised clustering such as flowSOM and Phenograph can be implemented for the extracted population for further clustering.
Differential analysis and statistical comparison	In this pipeline we support multiple study group statistical analysis(n>2) with multiple testing correction and post hoc analysis.

Figure 8

Overall schematic outline of the ImmCellTyper workflow with description for each step.

Agreement evaluation

To assess the agreement among manual gating, BinaryClust classification and flowSOM clustering, we used correlation, interaction plot and Bland-Altman analysis. Correlation evaluates the relationship between two variables which doesn't mean concordance, but if two methods agree, surely, they should be highly correlated. We compared the results of cell population frequency generated by experts and BinaryClust using Pearson correlation, calculated the correlation coefficient and P value, with the line of equality indicating perfect agreement (red solid line, $R^2=1$) in the plots to help gauge the degree of agreement between the two methods; Bland-Altman plot refers to a dot plot of the difference between two variables (y-axis) against the mean of them (x-axis), as described by J.Martian Bland and Douglas G. Altman in 1986, which represents a graphical magnitude of bias (average of difference) with 95% confidence interval. The math formula for the limit of agreement is as below:

Limits of agreement = mean difference observed \pm 1.96 \times standard deviation

We also used Interaction plot to display the interaction effects of the three methods including manual gating, BinaryClust classification and flowSOM clustering on the measurement of cell frequencies to evaluate the agreement.

F-measure

F-measure is a method to assess the accuracy of the clustering method compared to gold standard, which is manual gating results in this study. It stands for the harmonic mean of the precision and recall values, which can be calculated using below formula:

$$F - measure = 2x \frac{Precision \times Recall}{Precision + Recall}$$

Here, precision represents the positive predictive value: the proportion of true positive instances divided by the instances classified as positive by the clustering algorithm; recall evaluates the sensitivity, which is the number of true positive events correctly identified by the algorithm among all events that belong to the cluster. F-measure ranges from 0 to 1, where 1 suggests perfect performance and 0 indicates poor precision and recall.

Adjusted Rand Index

Adjusted rand index is a widely used method to measure the similarity between two clustering results. We used it to assess the agreement between test clustering algorithm and the gold standard labels which is derived from manual gating. ARI is defined as the following formula based on a contingency matrix (where n_{ij} , a_i , b_j are values from the contingency table):

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}$$

ARI ranges from -1 to 1, with 1 indicates a perfect match, 0 means not relevant, and -1 indicates complete mismatch between two clustering results.

Differential analysis

In order to perform the differential analysis of cell abundances and state marker expression amongst study groups, we consider the number of study groups into two conditions in function 'StatTest': 1) $n=2$, Mann-Whitney Wilcoxon analysis will be applied as cell frequency/marker

expression doesn't fit normal distribution; 2) $n=3$, Kruskal-Wallis test will be first performed followed by multiple testing correction using BH procedure and post-hoc analysis (Dunn's test or pairwise Wilcoxon test).

Sample collection and preparation for CyTOF

PBMC samples from MPN patients were requested and obtained from biobank at Guy's Hospital, under a protocol approved by the KCL Biobank Access Committee (REC18/EE/0025). Healthy volunteers were recruited at Guy's hospital with informed consent and ethical approval by King's College Research Ethics Committee (HR-17/18-5960 MOD-20/21-5960) in accordance with the Declaration of Helsinki. All identifiable information of study participants were securely stored in a trusted research environment managed by members of the team.

Venous blood samples from healthy volunteers were collected in BD vacutainer EDTA tubes, and PBMCs were isolated and purified using Ficoll-Hypaque density gradient centrifugation. In brief, blood was carefully layered onto Ficoll and spined at 460g for 20 minutes at room temperature (RT) without brake; Upon observing a clear separation of blood components, PBMCs were then carefully isolated and washed three times with RPMI 1640 medium to remove other contaminants. Subsequently, the purified PBMCs were cryopreserved in liquid nitrogen at a density of 1×10^7 cells per vial.

CyTOF Antibody staining

Cryopreserved cells were quickly thawed at 37°C in water bath, suspended in pre-warmed RPMI 1640 medium, washed three times, and underwent Fc-blocking using human TruStain FcX (Biolegend) for 10 mins at RT. Given that the antibody panel employs indirect detection of CD95 using anti-CD95-APC and anti-APC-106Cd, cells were initially stained with anti-CD95-APC for 30 minutes in a dark place, followed by two washes with 2ml cell staining buffer (CSB).

After the removal of supernatant, cells were resuspended in 300µl CSB and transferred into the dip tube containing lyophilized antibody mix of Maxpar direct immune profiling assay (MDIPA, Standard BioTools). Additional antibodies used to study T cell activation, migration, and exhaustion status (anti-TIM3, anti-PD-1, anti-ICOS, anti-TIGIT, anti-OX40) were also added.

Cells were incubated with the antibody mix for 30 mins at RT in compliance with manufacturer's instructions, washed twice and fixed using a freshly prepared 1.6% paraformaldehyde (PFA, Thermo Fisher Scientific) solution for 10 mins at RT. In the end, cells were washed with CSB twice to remove residual PFA and stained with 125nM Cell-ID intercalator-Iridium in 1 ml Maxpar Fix and Perm Buffer (Standard BioTools) overnight at 4°C.

The stained cells were frozen down using freezing media (fetal bovine serum containing 10% DMSO) and stored at -80°C freezer before data acquisition.

On the day for CyTOF acquisition, cryopreserved samples were thawed, washed twice with 1ml CSB, followed by additional two washes with 1ml cell acquisition solution (CAS, Standard BioTools), and centrifuged at 800g for 5 minutes. Cells were then filtered through a 40 µm cell strainer to avoid blockage and cell count was determined by Countess automated cell counter (Invitrogen). EQ four-element calibration beads (Standard BioTools) were added to a final concentration of 0.5×10^6 cells/ml to adjust signal fluctuation of the instrument. CyTOF acquisition was performed on Helios mass cytometer system. For each batch of cell staining and run, technical replicates from the same healthy donor were included to evaluate and correct batch effects.

The full antibody panel including the metal tag, clone, and supplier is listed in **Supplementary Table 1** [↗](#).

CyTOF data pre-processing

FCS files were first processed for bead normalisation using CyTOF v7.0 system control software (Standard BioTools) to correct signal drift during acquisition. Subsequently, the files were imported into Cytobank (Cytobank Inc.) for data cleaning, with a detailed procedure illustrated in **Supplementary Figure 6** [↗](#). The aim is to remove non-events including debris, doublets, normalisation (EQ) beads and other undesired events like dead cells. The CD45+ cell population was pre-gated using Cytobank for down-stream analysis.

Manual gating

Manual gating for major cell populations was performed using Cytobank platform, and gatingML files were exported and converted into gatingSet object to extract labels of each cell via the R packages flowWorkspace, openCyto and CytoML. These data were used as reference for benchmarking computer-aided algorithms.

Dataset availability

In this study, we tested ImmCellTyper pipeline on the MPN cohort with 7 MPN patients and 2 healthy volunteers, influenza cohort with 11 patients and the COVID-19 cohort with 59 COVID-19 patients and 23 healthy volunteers. The FCS files (after clean-up and gating of CD45 population) and metadata of the MPN cohort were deposited in Zenodo (doi: 10.5281/zenodo.10076940); The influenza cohort was published by our lab²¹ [↗](#), and the data were stored in Zenodo (doi: 10.5281/zenodo.7982165); The COVID-19 dataset was previously published by Chevrier et al.²² [↗](#), and the FCS files can be retrieved from Mendeley Data (<https://data.mendeley.com/datasets/vyy8ttw7n9/1> [↗](#)).

Author contributions

Jing Sun, Conceptualization, Data curation, Software, Formal analysis, Supervision, Validation, Visualisation, Methodology, Writing – original draft, Writing – review and editing, Resources, Experiments; Desmond Choy, Conceptualization, Software, Supervision, Methodology, Resources, Investigation; Nicolas Sompairac, Technical supervision, Resources, Data curation, Writing – review and editing; Shirin Jamshidi, Technical supervision, Resources, Writing – review and editing; Shahram Kordasti, Conceptualization, Resources, Funding acquisition, Supervision, Writing – review and editing, Project administration; Michele Mishto: Writing – review and editing

Competing interests

The authors declare that there are no competing interests exist.

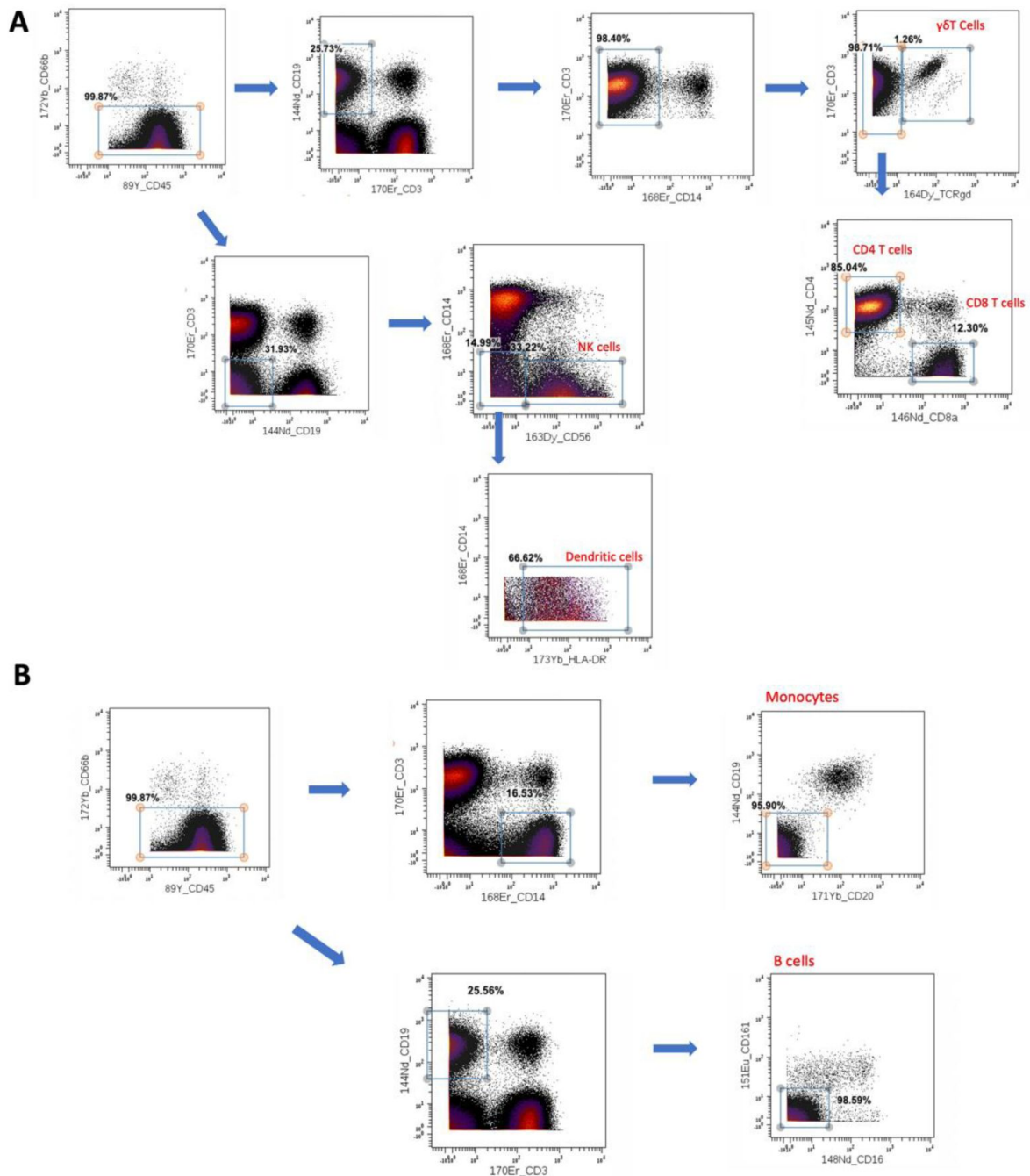
Acknowledgements

This work is supported in part by: Blood Cancer UK [Ref. 22009] and CRUK City of London Centre (CoL) Award [CTRQQR-2021/100004] to MM and SK; JS was supported by the K-CSC scholarship.

We thank BRC flowCore at Guy's Hospital for technical support, Biobank at King's College London for patient samples; We are thankful for Cynthia Bishop, Katrina Todd, Rosa Andres Ejarque, Elena Torre, Robert Page, Rita Antunes Dos Reis and Mishto lab members at King's College London,

Daniel Dancer, Thomas Adejumo at Standard Biotech for technical assistance and sample management, as well as all study participants for their kind support for this research. Special thanks to Eliezer Van Allen and Jihye Park at Dana Farber Cancer Institute to review and comment this work.

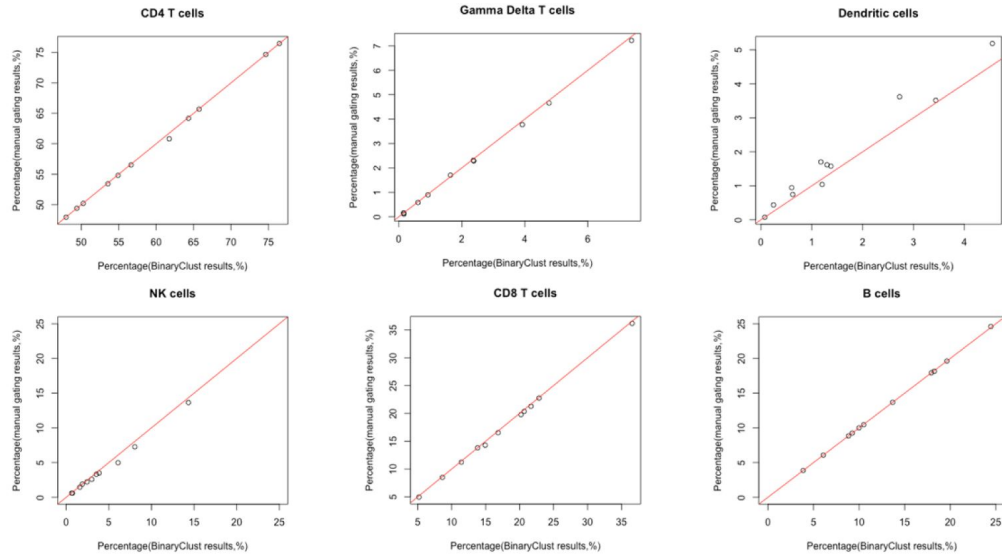
Supplementary Figures



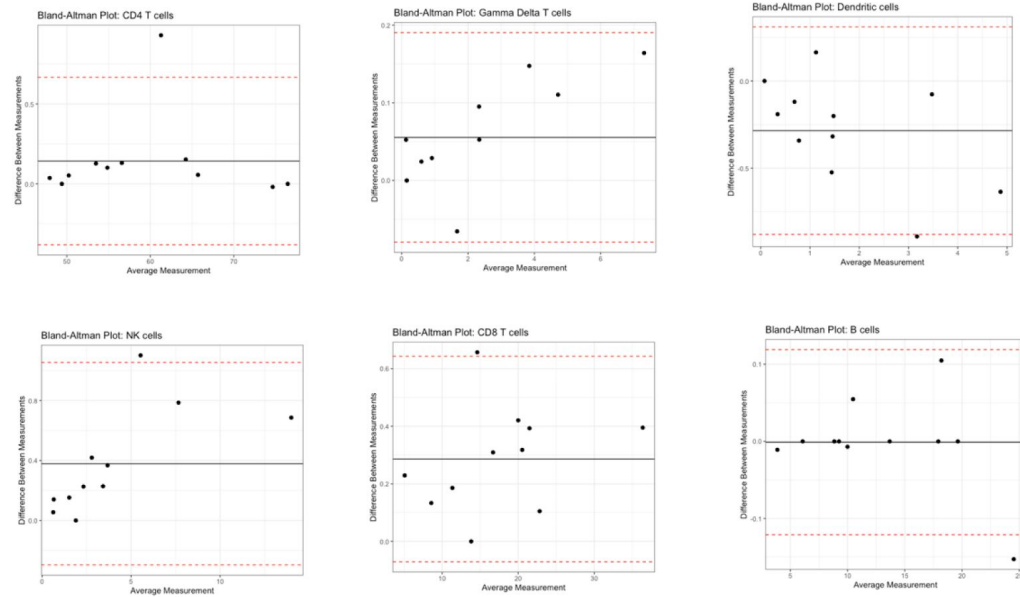
Supplementary Figure 1

Manual hierarchical gating strategy for main cell lineages from human PBMC samples (MPN dataset, n=9). All .FCS files were cleaned up to remove doublets, normalisation beads and debris (please refer to the methods for standard clean-up procedure), and pre-gated for CD45+ leukocytes. (A) Serial bi-axial scatter plots representing the gating diagram for T cell subsets (CD4 T cells, CD8 T cells and gamma-delta T cells), NK cells and dendritic cells based on the indicated phenotypic markers; (B) Serial bi-axial scatter plot indicating the gating strategy to isolate monocytes and B cells from leukocytes. All manual gating was done using Cytobank platform (<https://premium.cytobank.org/cytobank/>).

A



B

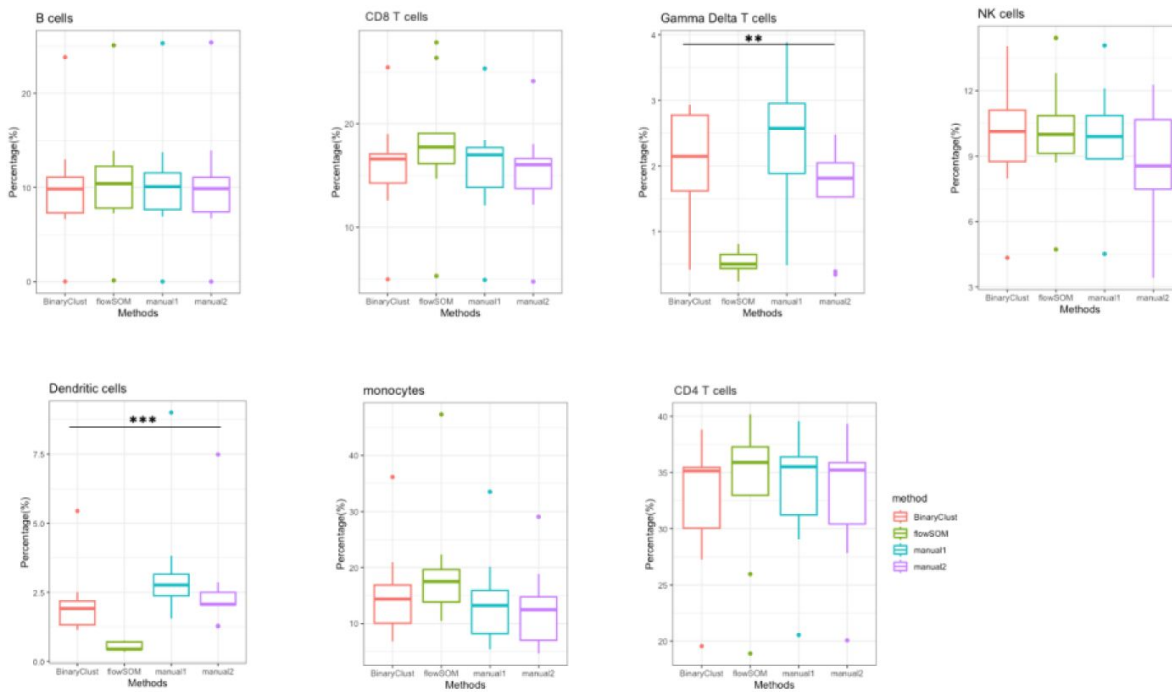


C

	Gated population	Counts	Cluster Cell Counts	True Positive	Precision	Recall	F-measure	Average F-measure	ARI	Average ARI
BinaryClust2	CD4 T cells	109498	109647	109404	1.00	1.00	1.00	0.98	0.95	0.91
	CD8 T cells	55449	56132	55435	0.99	1.00	0.99		0.95	
	NK cells	10759	11506	10748	0.93	1.00	0.97		0.97	
	B cells	16892	16881	16862	1.00	1.00	1.00		0.95	
	Dendritic cells	5090	4426	4417	1.00	0.87	0.93		0.74	
	TCRgd T cells	6848	7026	6837	0.97	1.00	0.99		0.91	

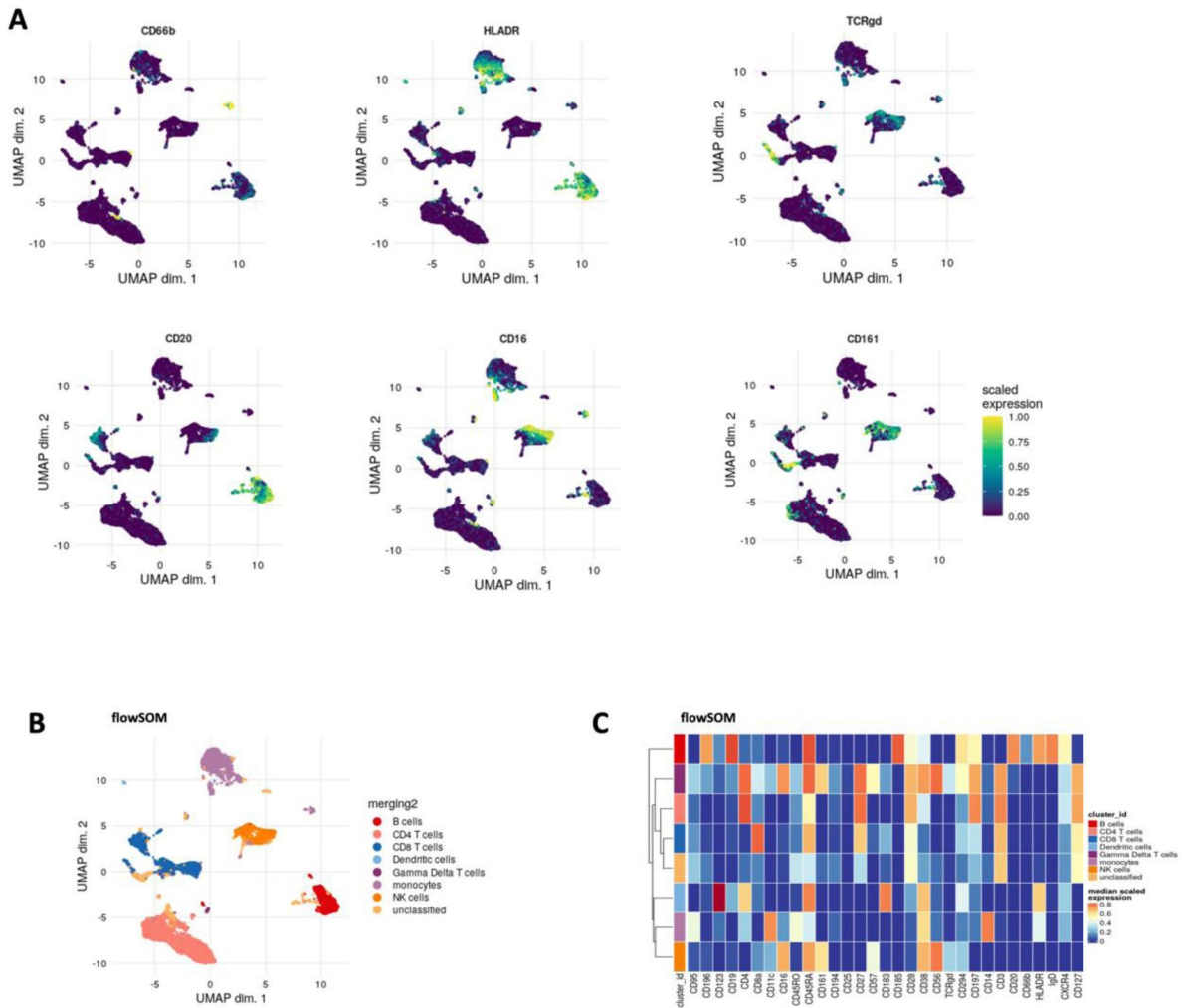
Supplementary Figure 2

Agreement evaluation between ImmCellTyper and manual gating in influenza dataset (n=11). Manual gating was performed using CytoBank and exported manually. (A) Correlation plots between ImmCellTyper results and manual gating results concerning percentages in CD4 T cells, Gamma Delta T cells, dendritic cells, NK cells, CD8 T cells and B cells, with red line indicating perfect agreement (correlation coefficient =1); (B) Bland-Altman plots of the two measurements in the indicated populations, with black line suggesting mean difference between measurements and dotted red line indicating limits of agreement (1.96 x standard deviations). (C) Calculation of precision, recall, F-measure for ImmCellTyper method in comparison to manual gating in the indicated cell populations.



Supplementary Figure 3

Boxplots of the indicated cell percentages generated by different methods. Statistical significance was marked by asterisk. * P<0.05, **P<0.01, ***P<0.001.

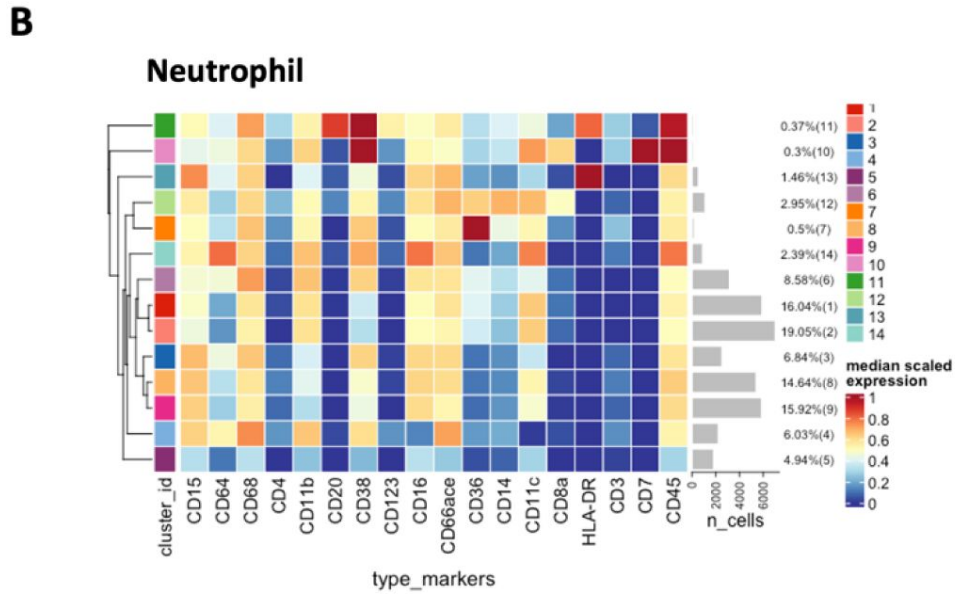
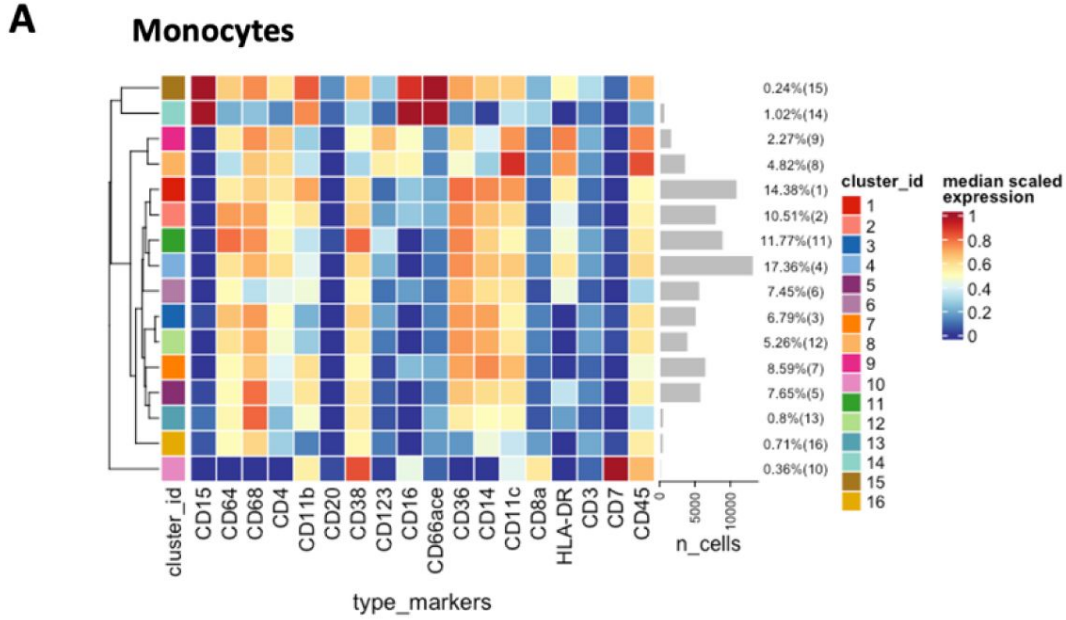


Supplementary Figure 4

(A) UMAP plots of normalized expression of indicated markers (CD66b, HLADR, TCRgd, CD20, CD16, CD161) across 2000 cells per sample in MPN dataset; (B) FlowSOM clustering was performed on the same dataset to compare with BinarClust, k=20 was chosen followed by manual annotation of each cluster. UMAP plot was projected with merged flowSOM clusters with biological annotation (downsample 2000 cells per sample); (B) The corresponding median marker expression heatmap after flowSOM clustering and annotation.

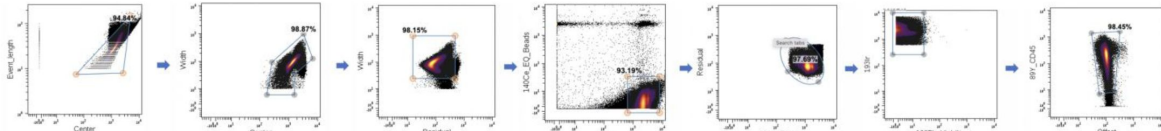
Supplementary Figure 5

Marker expression heatmap of (A) monocytes and (B) neutrophils generated by Phenograph clustering.



Supplementary Figure 6

Clean-up procedure of CyTOF data using Cytobank.



Antibody	Clone	Metal tag	Manufacturer
Anti-CD45	HI30	89Y	Standard BioTools
Live/dead Indicator	N/A	103Rh	Standard BioTools
Anti-CD95	DX2	APC	Biolegend
Anti-APC	APC003	106Cd	Standard BioTools
Anti-CD196	G034E3	141Pr	Standard BioTools
Anti-OX40	ACT35	142Nd	Standard BioTools
Anti-CD123	6H6	143Nd	Standard BioTools
Anti-CD19	HIB19	144Nd	Standard BioTools
Anti-CD4	RPA-T4	145Nd	Standard BioTools
Anti-CD8a	RPA-T8	146Nd	Standard BioTools
Anti-CD11c	Bu15	147Sm	Standard BioTools
Anti-CD16	3G8	148Nd	Standard BioTools
Anti-CD45RO	UCHL1	149Sm	Standard BioTools
Anti-CD45RA	HI100	150Nd	Standard BioTools
Anti-CD161	HP-3G10	151Eu	Standard BioTools
Anti-CD194	L291H4	152Sm	Standard BioTools
Anti-CD25	BC96	153Eu	Standard BioTools
Anti-CD27	O323	154Sm	Standard BioTools
Anti-CD57	HCD57	155Gd	Standard BioTools
Anti-CD183	G025H7	156Gd	Standard BioTools
Anti-CD185	J252D4	158Gd	Standard BioTools
Anti-TIM-3	F38-2E2	159Tb	Standard BioTools
Anti-CD28	CD28.2	160Gd	Standard BioTools
Anti-CD38	HB-7	161Dy	Standard BioTools
Anti-CD56	NCAM16.2	163Dy	Standard BioTools
Anti-TCRgd	B1	164Dy	Standard BioTools
Anti-PD-1	EH12.2H7	165Ho	Standard BioTools
Anti-CD294	BM16	166Er	Standard BioTools
Anti-CD197	G043H7	167Er	Standard BioTools
Anti-CD14	63D3	168Er	Standard BioTools
Anti-ICOS	C398.4A	169Tm	Standard BioTools
Anti-CD3	UCHT1	170Er	Standard BioTools
Anti-CD20	2H7	171Yb	Standard BioTools

Supplementary Table 1

CyTOF antibody panel for MPN cohort

Anti-CD66b	G10F5	172Yb	Standard BioTools836
Anti-HLA-DR	LN3	173Yb	Standard BioTools
Anti-IgD	IA6-2	174Yb	Standard BioTools
Anti-CXCR4	12G5	175Lu	Standard BioTools
Anti-CD127	A019D5	176Yb	Standard BioTools
Anti-TIGIT	MBSA43	209Bi	Standard BioTools

Supplementary Table 1 (continued)

References

1. Spitzer M.H., Nolan G.P (2016) **Mass Cytometry: Single Cells, Many Features** *Cell* **165**:780–791 <https://doi.org/10.1016/j.cell.2016.04.019>
2. Kimball A.K., Oko L.M., Bullock B.L., Nemenoff R.A., van Dyk L.F., Clambey E.T. (2018) **A Beginner's Guide to Analyzing and Visualizing Mass Cytometry Data** *J Immunol* **200**:3–22 <https://doi.org/10.4049/jimmunol.1701494>
3. Van Gassen S., Callebaut B., Van Helden M.J., Lambrecht B.N., Demeester P., Dhaene T., Saeyns Y. (2015) **FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data** *Cytometry A* **87**:636–645 <https://doi.org/10.1002/cyto.a.22625>
4. Levine J.H. *et al.* (2015) **Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis** *Cell* **162**:184–197 <https://doi.org/10.1016/j.cell.2015.05.047>
5. Samusik N., Good Z., Spitzer M.H., Davis K.L., Nolan G.P (2016) **Automated mapping of phenotype space with single-cell data** *Nat Methods* **13**:493–496 <https://doi.org/10.1038/nmeth.3863>
6. Qiu P., Simonds E.F., Bendall S.C., Gibbs K.D., Bruggner R.V., Linderman M.D., Sachs K., Nolan G.P., Plevritis S.K (2011) **Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE** *Nat Biotechnol* **29**:886–891 <https://doi.org/10.1038/nbt.1991>
7. Becher B. *et al.* (2014) **High-dimensional analysis of the murine myeloid cell system** *Nature Immunology* **15**:1181–1189 <https://doi.org/10.1038/ni.3006>
8. van der Maaten L., Hinton G. (2008) **Visualizing data using t-SNE** *Journal of Machine Learning Research* **9**:2579–2605
9. McInnes L., Healy J (2018) **UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction** *ArXiv* <https://doi.org/10.48550/arXiv.1802.03426>
10. Liu X., Song W., Wong B.Y., Zhang T., Yu S., Lin G.N., Ding X (2019) **A comparison framework and guideline of clustering methods for mass cytometry data** *Genome Biology* **20** <https://doi.org/10.1186/s13059-019-1917-7>
11. Abdelaal T., van Unen V., Höllt T., Koning F., Reinders M.J.T., Mahfouz A. (2019) **Predicting Cell Populations in Single Cell Mass Cytometry Data** *Cytometry A* <https://doi.org/10.1002/cyto.a.23738>
12. Cheng L., Karkhanis P., Gokbag B., Liu Y., Li L (2022) **DGCyTOF: Deep learning with graphic cluster visualization to predict cell types of single cell mass cytometry data** *PLOS Computational Biology* **18** <https://doi.org/10.1371/journal.pcbi.1008885>
13. Kaushik A., Dunham D., He Z., Manohar M., Desai M., Nadeau K.C., Andorf S (2021) **CyAnno: a semi-automated approach for cell type annotation of mass cytometry datasets** *Bioinformatics* **37**:4164–4171 <https://doi.org/10.1093/bioinformatics/btab409>

14. Li H., Shaham U., Stanton K.P., Yao Y., Montgomery R.R., Kluger Y (2017) **Gating mass cytometry data by deep learning** *Bioinformatics* **33**:3423–3430 <https://doi.org/10.1093/bioinformatics/btx448>
15. Lee H.C., Kosoy R., Becker C.E., Dudley J.T., Kidd B.A (2017) **Automated cell type discovery and classification through knowledge transfer** *Bioinformatics* **33**:1689–1695 <https://doi.org/10.1093/bioinformatics/btx054>
16. Zhang Z. *et al.* (2019) **SCINA: A Semi-Supervised Subtyping Algorithm of Single Cells and Bulk Samples** *Genes (Basel)* **10** <https://doi.org/10.3390/genes10070531>
17. Nowicka M., Krieg C., Crowell H.L., Weber L.M., Hartmann F.J., Guglietta S., Becher B., Levesque M.P., Robinson M.D (2017) **CytoF workflow: differential discovery in high-throughput high-dimensional cytometry datasets** *F1000Res* **6** <https://doi.org/10.12688/f1000research.11622.3>
18. Marsh-Wakefield F. *et al.* (2019) **Mass Cytometry Discovers Two Discrete Subsets of CD39(-)Treg Which Discriminate MGUS From Multiple Myeloma** *Front Immunol* **10** <https://doi.org/10.3389/fimmu.2019.01596>
19. Chen H., Lau M.C., Wong M.T., Newell E.W., Poidinger M., Chen J (2016) **Cytofkit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline** *PLoS Comput Biol* **12** <https://doi.org/10.1371/journal.pcbi.1005112>
20. Opzoomer J.W. *et al.* (2021) **ImmunoCluster provides a computational framework for the nonspecialist to profile high-dimensional cytometry data** *Elife* **10** <https://doi.org/10.7554/eLife.62915>
21. Alimam S. *et al.* (2021) **Altered immune response to the annual influenza A vaccine in patients with myeloproliferative neoplasms** *Br J Haematol* **193**:150–154 <https://doi.org/10.1111/bjh.17096>
22. Chevrier S. *et al.* (2021) **A distinct innate immune signature marks progression from mild to severe COVID-19** *Cell Rep Med* **2** <https://doi.org/10.1016/j.xcrm.2020.100166>
23. Liu P., Liu S., Fang Y., Xue X., Zou J., Tseng G., Konnikova L (2020) **Recent Advances in Computer-Assisted Algorithms for Cell Subtype Identification of Cytometry Data** *Front Cell Dev Biol* **8** <https://doi.org/10.3389/fcell.2020.00234>
24. Aghaeepour N., Nikolic R., Hoos H.H., Brinkman R.R (2011) **Rapid cell population identification in flow cytometry data** *Cytometry A* **79**:6–13 <https://doi.org/10.1002/cyto.a.21007>
25. Theorell A., Bryceson Y.T., Theorell J (2019) **Determination of essential phenotypic elements of clusters in high-dimensional entities-DEPECHE** *PLoS One* **14** <https://doi.org/10.1371/journal.pone.0203247>
26. O'Brien Gore C. *et al.* (2023) **Pre-treatment with systemic agents for advanced NSCLC elicits changes in the phenotype of autologous T cell therapy products** *Mol Ther Oncolytics* **31** <https://doi.org/10.1016/j.omto.2023.100749>
27. Weber L.M., Nowicka M., Sonesson C., Robinson M.D (2019) **diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering** *Communications Biology* **2** <https://doi.org/10.1038/s42003-019-0415-5>

28. Van Gassen S., Gaudilliere B., Angst M.S., Saeys Y., Aghaeepour N. (2020) **CytoNorm: A Normalization Algorithm for Cytometry Data** *Cytometry A* **97**:268–278 <https://doi.org/10.1002/cyto.a.23904>
29. Trussart M., Teh C.E., Tan T., Leong L., Gray D.H., Speed T.P (2020) **Removing unwanted variation with CytofRUV to integrate multiple CyTOF datasets** *Elife* **9** <https://doi.org/10.7554/eLife.59630>
30. Arvaniti E., Claassen M (2017) **Sensitive detection of rare disease-associated cell subsets via representation learning** *Nature Communications* **8** <https://doi.org/10.1038/ncomms14825>
31. Lun A.T.L., Richard A.C., Marioni J.C (2017) **Testing for differential abundance in mass cytometry data** *Nat Methods* **14**:707–709 <https://doi.org/10.1038/nmeth.4295>
32. Bruggner R.V., Bodenmiller B., Dill D.L., Tibshirani R.J., Nolan G.P (2014) **Automated identification of stratifying signatures in cellular subpopulations** *Proc Natl Acad Sci U S A* **111**:E2770–2777 <https://doi.org/10.1073/pnas.1408792111>
33. Arend L., Bernett J., Manz Q., Klug M., Lazareva O., Baumbach J., Bongiovanni D., List M (2022) **A systematic comparison of novel and existing differential analysis methods for CyTOF data** *Briefings in Bioinformatics* **23** <https://doi.org/10.1093/bib/bbab471>

Editors

Reviewing Editor

Seunghee Hong

Yonsei University, Seoul, Korea, the Republic of

Senior Editor

Satyajit Rath

Indian Institute of Science Education and Research (IISER), Pune, India

Reviewer #3 (Public Review):

Summary:

ImmCellTyper is a new toolkit for Cytometry by time-of-flight data analysis. It includes BinaryClust, a semi-supervised clustering tool (which takes into account the prior biological knowledge), designed for automated classification and annotation of specific cell types and subpopulations. ImmCellTyper also integrates a variety of tools to perform data quality analysis, batch effect correction, dimension reduction, unsupervised clustering, and differential analysis.

Strengths:

The proposed algorithm takes into account the prior knowledge.

The results on different benchmark indicates competitive or better performance (in terms of accuracy and speed) depending on the method.

<https://doi.org/10.7554/eLife.95494.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:**Reviewer #1 (Public Review):***Summary:*

This manuscript presented a useful toolkit designed for CyTOF data analysis, which integrates 5 key steps as an analytical framework. A semi-supervised clustering tool was developed, and its performance was tested in multiple independent datasets. The tool was compared to human experts as well as supervised and unsupervised methods.

Strengths:

The study employed multiple independent datasets to test the pipeline. A new semi-supervised clustering method was developed.

Weaknesses:

The examination of the whole pipeline is incomplete. Lack of descriptions or justifications for some analyses.

We thank the reviewer's overall summary and comments of this manuscript. In the last part of the results, we showcased the functionalities of ImmCellTyper in covid dataset, including quality check, BinaryClust clustering, cell abundance quantification, state marker expression comparison within each identified cell types, cell population extraction, subpopulation discovery using unsupervised methods, and data visualization etc. We added more descriptions in the text based on the reviewer's suggestions.

Reviewer #2 (Public Review):*Summary:*

The authors have developed marker selection and k-means (k=2) based binary clustering algorithm for the first-level supervised clustering of the CyTOF dataset. They built a seamless pipeline that offers the multiple functionalities required for CyTOF data analysis.

Strengths:

The strength of the study is the potential use of the pipeline for the CyTOF community as a wrapper for multiple functions required for the analysis. The concept of the first line of binary clustering with known markers can be practically powerful.

Weaknesses:

The weakness of the study is that there's little conceptual novelty in the algorithms suggested from the study and the benchmarking is done in limited conditions.

We thank the reviewer's overall summary and comments of this manuscript. While the concept of binary clustering by k-means is not novel, BinaryClust only uses it for individual markers to identify positive and negative cells, then combine it with the pre-defined matrix for cell type identification. This has not been introduced elsewhere. Furthermore, ImmCellTyper streamlines the entire analysis process and enhances data exploration on multiple levels. For instance, users can evaluate functional marker expression level/cellular

abundance across both main cell types and subpopulations; Also, this computational framework leverages the advantages of both semi-supervised and unsupervised clustering methods to facilitate subpopulation discovery. We believe these contributions warrant consideration as advancements in the field.

As for the benchmarking, we limited the depth only to main cell types rather than subpopulations. The reason is because we only apply BinaryClust to identify main cell types; For the cell subsets discovery, unsupervised methods integrated in this pipeline has already been published and widely used by the research community. Therefore, it does not seem to be necessary for additional benchmarking.

Reviewer #3 (Public Review):

Summary:

ImmCellTyper is a new toolkit for Cytometry by time-of-flight data analysis. It includes BinaryClust, a semi-supervised clustering tool (which takes into account prior biological knowledge), designed for automated classification and annotation of specific cell types and subpopulations. ImmCellTyper also integrates a variety of tools to perform data quality analysis, batch effect correction, dimension reduction, unsupervised clustering, and differential analysis.

Strengths:

The proposed algorithm takes into account the prior knowledge.

The results on different benchmarks indicate competitive or better performance (in terms of accuracy and speed) depending on the method.

Weaknesses:

The proposed algorithm considers only CyTOF markers with binary distribution.

We thank the reviewer's overall summary and comments of this manuscript. Binary classification can be considered as an imitation of human gating strategy, as it is applied to each marker. For example, when characterizing the CD8 T cells, we aim for CD19-CD14-CD3+CD4- population, which is binary in nature (either positive and negative) and follows the same logic as the method (BinaryClust) we developed. Results indicated that it works very well for well-defined main cell lineages, particularly when the expression of the defining marker is not continuous. However, the limitation is for subpopulation identification, because a handful makers behave in a continuum manner, so we suggest unsupervised method after BinaryClust, which also brings another advantage of identifying unknown subsets beyond our current knowledge, and none of the semi-supervised tools can achieve that. To address the reviewer's concern, we considered the limitation of binary distribution, but it does not profoundly affect the application of the pipeline.

Recommendations for the authors:

Reviewer #1 (Recommendations For The Authors):

Many thanks for the reviewers' comments and suggestions, please see below the point-to-point response:

(1) The style of in-text reference citation is not consistent. Many do not have published years.

The style of the reference citation has been revised and improved.

| (2) The font size in the table of Figure 1 is too small, so is Figure 2.

The font size has been increased.

| (3) Is flowSOM used as part of BinaryClust? How should the variable running speed of BinaryClust be interpreted, given that it is occasionally slower and sometimes faster than flowSOM in the datasets?

To answer reviewer's question, flowSOM is not a part of BinaryClust. They are separate clustering methods that have been incorporated into the ImmCellTyper pipeline. As described in Figure 1, BinaryClust, a semi-supervised method, is used to classify the main cell lineages; while flowSOM, an unsupervised method, is recommended here for further subpopulation discovery. So, they operate independently of each other. To avoid confusions, we slightly modified Figure 1 for clarification.

Regarding the variability in running speed in Figure 4. The performance of algorithms can indeed be influenced by the characteristics of the datasets, such as size and complexity. The differences observed between the covid dataset and the MPN dataset, such as marker panel, experimental protocol, and data acquisition process etc., could account for this variation. Our explanation is that flowSOM suits better the data structure of covid dataset, which might be the reason why it is slightly faster to analyse compared to the MPN dataset. Moreover, for the covid dataset, the runtime for both BinaryClust and flowSOM is less than 100s, and the difference is not notable.

| (4) In the Method section ImmCellTyper workflow overview, it is difficult to link the description of the pipeline to Figure 8. There are two sub-pipelines in the text and seven steps in the figure. What are their relations? Some steps are not introduced in the text, such as Data transformation and SCE object construction. What is co-factor 5?

Figure 8 provides an overview of the entire workflow for CyTOF data analysis, starting from the raw fcs file data and proceeding until downstream analysis (seven steps). But the actual implementation of the pipeline was divided into two separate sections, as outlined in the vignettes of the ImmCellTyper GitHub page (<https://github.com/JingAnyasun/ImmCellTyper/tree/main/vignettes>).

Users will initially run 'Intro_to_batch_exam_correct' to perform data quality check and identify potential batch effects, followed by 'Intro_to_data_analysis' for data exploration. We agree with the reviewer that the method for this section is a bit confusing, so we've added more description for clarification.

In processing mass cytometry data, arcsine transformation is commonly applied to handle zero values, skewed distributions, and to improve visualization as well as clustering performance. The co-factor here is used as a parameter to scale down the data to control the width of the linear region before arcsine transformation. We usually get the best results by using co-factor 5 for CyTOF data.

| (5) For differential analysis, could the pipeline analyze paired/repeated samples?

For the statistical step, ImmCellTyper supports both two-study group comparison using Mann-Whitney Wilcoxon test, and multiple study group comparison (n>2) using Kruskal Wallis test followed by post hoc analysis (pairwise Wilcoxon test or Dunn's test) with multiple testing correction using Benjamini-Hochberg Procedure.

Certainly, this pipeline allows flexibilities, users can also extract the raw data of cell frequencies and apply suitable statistical methods for testing.

(6) In Figure 2A, the range of the two axes is different for Dendritic cells, which could be misleading. Why the agreement is bad for dendritic cells?

The range for the axes is automatically adapted to the data structure, which explains why they may not necessarily be equal. The co-efficient factor for the correlation of DCs is 0.958, compared to other cell types (> 0.99), it is relatively worse but does not indicate poor agreement.

Moreover, the abundance of DCs is much less than other cell types, comprising approximately 2-5% of whole cells. As a result, even small differences in abundance may appear to as significant variations. For example, a difference of 1% in DC abundance represents a 2-fold change, which can be perceived as substantial.

Overall, while the agreement for DCs may appear comparatively lower, it is not necessarily indicative of poor performance, considering both the coefficient factor and the relative abundance of DCs compared to other cell types.

(7) In the Results section BinaryClust achieves high accuracy, what method was used to get the p-value, such as lines 212, 213, etc.?

The accuracy of BinaryClust was tested using F-measure and ARI against ground truth (manual gating), the detailed description/calculation can be found in methods. For line 212 and 213, the p-value was calculated using ANOVA for the interaction plot shown in Figure 3. We've now added the statistical information into the figure legend.

(8) The performance comparison between BinaryClust and LDA is close. The current comparison design looks unfair. Given LDA only trained using half data, LDA may outperform BinaryClust.

It is true that LDA was trained using half data, which is because this method requires manual gating results as training dataset to build a model, then apply the model to the rest of the files to label cell types. Here we used 50% of the whole dataset as training set. We are of course very happy to implement any additional suggestions for a better partition ratio.

(9) There are 5 key steps in the proposed workflow. However, not every step was presented in the Results.

Thanks for the comments. The results primarily focused on demonstrating the precision and performance of BinaryClust in comparison with ground truth and existing tools. Additionally, a case study showcasing the application/functions of the entire pipeline in a dataset was also presented. Due to limitation in space, the implementation details of the pipeline were described in the method section and github documentations, which users/readers can easily access.

Reviewer #2 (Recommendations For The Authors):

The tools suggested by the authors could be potentially useful to the community. However, it's difficult to understand the conceptual novelty of the algorithms suggested here. The concept of binary clustering has been described before (<https://doi.org/10.1186/s12859-022-05085-z>, <https://doi.org/10.1152/ajplung.00104.2022>), and it mainly utilizes

k-means clustering set to generate binary clusters based on selected markers. Other algorithms associated with the package are taken from other studies.

We acknowledge the reviewer's comment regarding the novelty of our method. While the concept of binary clustering by k-means has been previously described to transcriptome data, our approach applies it to CyTOF data analysis, which has not been introduced elsewhere. Furthermore, ImmCellTyper streamlines the entire analysis process and enhances data exploration on multiple levels. For instance, users can evaluate functional marker expression level/cellular abundance across both main cell types and subpopulations; Also, as stated in the manuscript, this computational framework leverages the advantages of both semi-supervised and unsupervised clustering methods to facilitate subpopulation discovery. We believe these contributions warrant consideration as advancements in the field.

In addition, the benchmarking of clustering performance, especially to reproduce manual gating and comparison to tools such as flowSOM is not comprehensive enough. The result for the benchmarking test could significantly vary depending on how the authors set the ground truth (resolution of cell type annotations). The authors should compare the tool's performance by changing the depth of cell type annotations. Especially, the low abundance cell types such as gdT cells or DCs were not effectively captured by the suggested methods.

Thanks for the comment. We appreciate the reviewer's concern. However, as illustrated in figure 1, our approach uses BinaryClust, a semi-supervised method, to identify main cell types rather than directly targeting subpopulations. The reason is because semi-supervised method relies on users' prior definition thus is limited to discover novel subsets. In the ImmCellTyper framework, unsupervised method was subsequently applied for subset exploration following the BinaryClust step.

Regarding benchmarking, we focused on testing the precision of BinaryClust for main cell type characterization, because it is what the method is used for in the pipeline, and we believe this is sufficient. As for the cell subsets discovery, the unsupervised methods we integrated has already been published and widely used by the research community. Therefore, it does not seem to be necessary for additional benchmarking.

Moreover, as shown in Figure 3 and Table 1, our results indicated that the F-measure for DCs and gdT cells in BinaryClust is 0.80 and 0.92 respectively, which were very close to ground truth and outperformed flowSOM, demonstrating its effectiveness.

We hope these clarifications address the reviewer's concern.

Minor comments:

(1) In Figure 4, it's perplexing to note that BinaryClust shows the slowest runtime for the COVID dataset, compared to the MPN dataset, which features a similar number of cells. What causes this variation? Is it dependent on the number of markers utilized for the clustering? This should be clarified/tested.

Thanks for the comment, but we are not sure that we fully understand the question. As shown in figure 4 that BinaryClust has slightly higher runtime in MPN dataset than covid dataset, which is reasonable because and the cell number in MPN dataset is around 1.6 million more than covid dataset.

(2) Some typos are noted:

*- DeepCyTOF and LDA use a maker expression matrix extracted → "marker"?**

Corrected.

| - *Datasets(Chevrier et al.)which* → *spacing**

Corrected.

| - *This is due to the method's reliance* → *spacing**

Corrected.

Reviewer #3 (Recommendations For The Authors):

Is it possible to accommodate more than two levels within the clustering process, i.e., can the proposed semi-supervised clustering tool be extended to multi-levels instead of binary?

Thanks for the comments. Binary classification can be considered as an imitation of human gating strategy, as it is applied to each marker. For example, when characterizing the CD8 T cells, we aim for CD19-CD14-CD3+CD4- population, which is binary in nature (either positive and negative) and follows the same logic as the method (BinaryClust) we developed. Results indicated that it works very well for well-defined main cell lineages. However, the limitation is for subpopulation identification, because a handful of makers behave in a continuum manner, so we would suggest unsupervised method after BinaryClust, which also brings another advantage of identifying unknown subsets beyond our current knowledge, and none of the semi-supervised tools can achieve that. To answer the reviewer's question, it is possible to set the number to 3,4,5 rather than just 2, but considering the design and rationale of the entire framework (as describe in the manuscript and above), it doesn't seem to be necessary.

| *Could you please comment on why on the COVID dataset, BinaryClust was slower as compared to flowSOM?*

Thanks for the question. The performance of algorithms can indeed be affected by the characteristics of the datasets, such as their size and complexity. The covid and MPN datasets differ in various aspects including marker panel, experimental protocol, and data acquisition process, among others, which would account for the observed variation in speed. So, our explanation is flowSOM suits better for the structure of covid dataset than MPN dataset. Additionally, for covid dataset, both BinaryClust and flowSOM have runtimes of less than 100s, and the difference between the two isn't particularly dramatic.

| *Minor errors:*

| *Line#215 "(ref)" reference is missing*

Added.

| *Figure 3, increase the font of the text in order to improve readability.*

Increased.

| *Line#229 didn't --> did not.*

Corrected

| *Line#293 repetition of the reference.*

The repetition is due to the format of the citation, which has been revised.

<https://doi.org/10.7554/eLife.95494.2.sa0>