Supplementary information

**Current trends in flow cytometry automated data analysis software**

Melissa Cheung1\*, Jonathan J. Campbell2, Liam Whitby3, Robert J. Thomas1,

Julian Braybrook2 and Jon Petzing1

1 Centre for Biological Engineering, Loughborough University, Loughborough, Leicestershire, UK.

2 National Measurement Laboratory, LGC, Queen’s Road, Teddington, UK.

3 UK NEQAS for Leucocyte Immunophenotyping, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK.

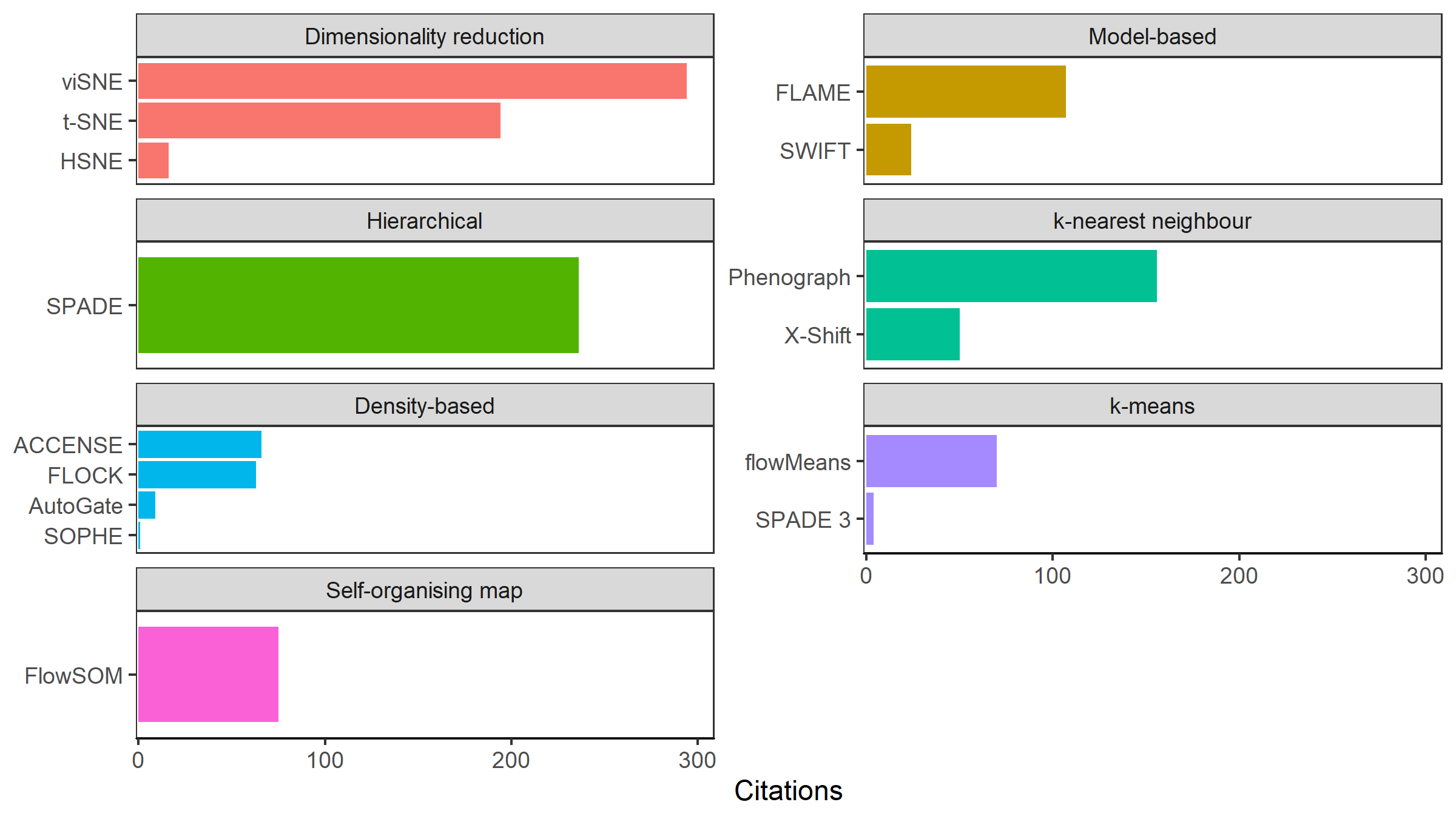
Supplementary Table 1. Software identified in literature survey, ranked according to number of citations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rank** | **Software name** | **Number of citations** | **Abbreviation** | **Purpose** | **Reference** |
| 1 | viSNE | 294 | visualization tool for high-dimensional single-cell data based on the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm | Visualisation of high-dimensional single-cell data via dimensionality reduction | (1) |
| 2 | SPADE | 236 | Spanning-tree progression analysis of density-normalized events | Visualisation of high-dimensional cytometry data by downsampling, clustering and a minimal spanning tree | (2) |
| 3 | t-SNE | 194 | t-Distributed Stochastic Neighbor Embedding | Dimensionality reduction for visualisation | (3) |
| 4 | Phenograph | 156 |  | Model cellular phenotypes | (4) |
| 5 | FLAME | 107 | Flow analysis with automated multivariate estimation | Identify cell populations by multivariate mixture modelling | (5) |
| 6 | Citrus | 87 | Cluster identification, characterization, and regression | Identification of stratifying cellular subpopulations | (6) |
| 7 | FlowSOM | 75 | Self-organizing map | Clustering data into self-organizing maps and visualisation by minimal spanning trees | (7) |
| 8 | DensVM | 70 | Density-based clustering aided by support vector machine | Cell population identification and classification | (8) |
| 8 | flowMeans | 70 |  | Cell population identification by k-Means based clustering | (9) |
| 10 | ACCENSE | 66 | Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding | Identification of cell subpopulations through t-SNE dimensionality reduction and density-based partitioning | (10) |
| 10 | Wanderlust | 66 |  | Developmental trajectory detection | (11) |
| 12 | FLOCK | 63 | FLOw Clustering without K | Cell population identification by density-based clustering | (12) |
| 13 | flowClust | 58 |  | Cell population identification by multivariate t-mixture modelling with Box-Cox transformed data | (13) |
| 14 | flowMerge | 53 |  | Cell population identification using flowClust and a cluster merging algorithm | (14) |
| 15 | X-Shift | 50 |  | Exploration of single-cell data by clustering (K-nearest neighbour density estimate) and visualisation by divisive marker trees and force-directed layouts | (15) |
| 16 | SamSPECTRAL | 48 |  | Cell population identification by spectral clustering and sampling | (16) |
| 17 | flowPeaks | 42 |  | Cell population identification by K-means clustering and density peak finding | (17) |
| 17 | OpenCyto | 42 |  | Mimicking manual gating based on hierarchical automated gating pipelines | (18) |
| 19 | Mixture model | 41 |  | Cell population identification by mixture modelling | (19) |
| 20 | HPDGMM | 31 | Hierarchical Dirichlet Process Gaussian Mixture Model | Rare event detection and cell subset alignment across multiple samples | (20) |
| 21 | flowDensity | 26 |  | Mimicking manual gating based on cellular density distributions | (21) |
| 22 | SWIFT | 24 | Scalable Weighted Iterative Flow-clustering Technique | Identification of rare cell populations based on Gaussian mixture model-based clustering | (22,23) |
| 23 | HSNE | 16 | Hierarchical Stochastic Neighbor Embedding | Visual exploration of the hierarchy in cytometry data | (24) |
| 24 | Misty Mountain | 15 |  | Cell population identification by density contour clustering | (25) |
| 25 | COMPASS | 14 | Combinatorial Polyfunctionality Analysis of Single Cells | Identification of cell subsets correlated with clinical outcomes | (26) |
| 25 | FlowFP | 14 | Fingerprinting for Flow Cytometry | Generation of multivariate distribution 'fingerprints' | (27) |
| 25 | immunoClust | 14 |  | Cell population identification by iterative model-based clustering | (28) |
| 28 | JCM | 12 | Joint Clustering and Matching | Cell population identification and matching across a batch of samples | (29) |
| 29 | flowType/RchyOptimyx | 11 |  | Cell population identification by partitioning and correlation with clinical outcomes | (30) |
| 30 | ASPIRE | 10 | Anomalous sample phenotype identification with random effects | Identification of anomalous samples with random effects | (31) |
| 30 | DeepCyTOF | 10 |  | Cell classification by deep learning | (32) |
| 32 | AutoGate | 9 |  | Sequential selection of cell subsets and visualisation | (33) |
| 33 | FloReMi | 8 | Flow Density Survival Regression Using Minimal Feature Redundancy | Survival time prediction | (34) |
| 34 | CCAST | 7 | Clustering, Classification and Sorting Tree | Isolation of homogenous subpopulations | (35) |
| 35 | flowLearn | 6 |  | Identification and quality checking of cell populations | (36) |
| 36 | ACDC | 5 | Automated Cell-type Discovery and Classification | Cell population discovery and classification | (37) |
| 37 | Competitive SWIFT | 4 | Scalable Weighted Iterative Flow-clustering Technique | Sample comparison by competitive clustering | (38) |
| 37 | SPADE 3 | 4 | Spanning-tree progression analysis of density-normalized events | Visualisation of high-dimensional cytometry data by downsampling, clustering and a minimal spanning tree | (39) |
| 39 | cytometree | 2 |  | Cell population identification based on a binary tree algorithm | (40) |
| 39 | DAFi | 2 | Directed Automated Filtering and Identification of cell populations | Cell population identification based on recursive data filtering and clustering | (41) |
| 39 | diffcyt | 2 | Differential discovery in high-dimensional cytometry via high-resolution clustering | Differential discovery analysis | (42) |
| 39 | FlowVIEW | 2 |  | Quantification of cell populations via a supervised learning approach | (43) |
| 39 | LDA | 2 | Linear discriminant analysis | Prediction of cell populations | (44) |
| 44 | ECLIPSE | 1 | Elimination of Cells Lying in Pattern Similar to Endogeneity | Identification of disease-specific cells | (45) |
| 44 | NPflow | 1 | Bayesian Nonparametrics for Automatic Gating of Flow-Cytometry Data | Cell population identification by model-based clustering | (46) |
| 44 | PSM with GemStone | 1 | Probability State Modeling | Cell population identification via a probability-based approach | (47) |
| 44 | SOPHE | 1 | Second order polynomial histogram estimators | Cell population identification by data binning | (48) |
| 48 | PHATE | 0 | Potential of heat diffusion for affinity-based transition embedding | Dimensionality reduction for visualisation | (49) |
| 48 | SIC | 0 | Subset Identification and Characterisation | Subset identification and characterisation pipeline | (50) |
| 48 | SigClust | 0 | Signature based Single-Cell Clustering | Cell population identification using phenotypic signatures | (51) |
| 48 | UMAP | 0 | Uniform Manifold Approximation and Projection | Dimensionality reduction for visualisation | (52) |

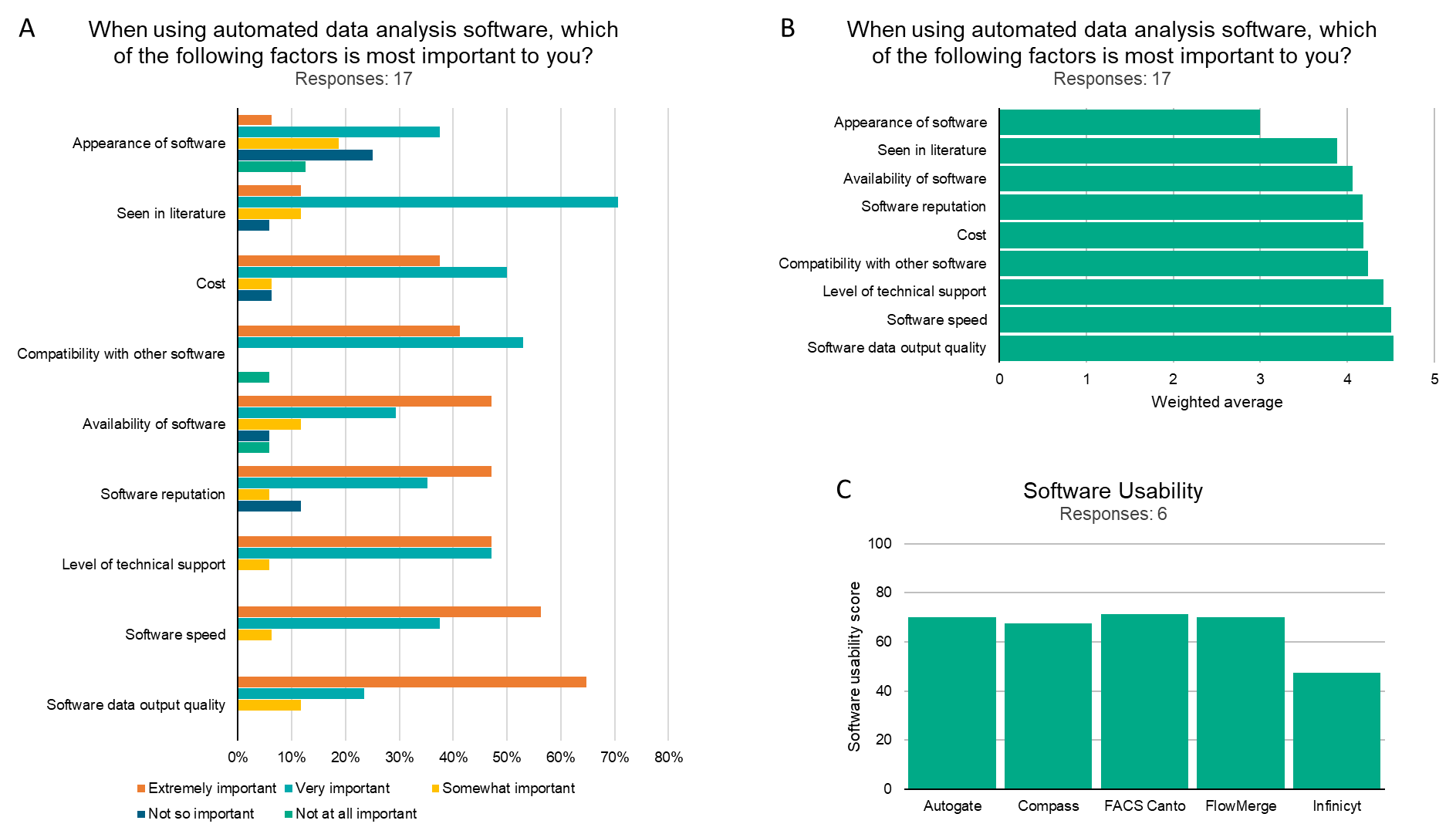
Supplementary Table 2. Survey questions and answer response choices.

|  |  |
| --- | --- |
| Q1 | In a typical week, how many hours do you spend analysing (gating) flow cytometry data on a computer?   * Over 30 hours * 20-30 hours * 10-20 hours * 1-10 hours * Less than 1 hour |
| Q2 | How often do you use automated flow cytometry data analysis software to identify cell populations?   * Never – I only use manual gating to identify cell populations. * Rarely – I mainly use manual gating, but occasionally use automated tools. * Sometimes – I split my analysis equally between manual and automated cell population identification. * Usually – I mainly use automated tools, but occasionally use manual gating. * Always – I use automated tools for all my data analysis. |
| Q3 | Which software do you use for manual cell population identification? (Check all that apply)   * BD FACS Diva * BD FACS Canto * BD FACSuite * BD CellQuest * FCS Express * FlowJo * FlowLogic * Infinicyt * Kaluza * Navios * VenturiOne * WinList * Other (please specify) |
| Q4 | Which software do you use for automated cell population identification? (Check all that apply)  (Answer choices as in Supplementary Table 1) |
| Q5 | Which automated data analysis software are you aware of, but do not currently use? (Check all that apply)  (Answer choices as in Supplementary Table 1) |
| Q6 | When using automated data analysis software, which of the following factors is most important to you?  (Answer choices for each factor: Not at all important/ Not so important/ Somewhat important/ Very important/ Extremely important)   * Appearance of software * Availability of software * Compatibility with other software * Level of technical support * Seen in literature * Software data output quality * Software reputation * Software speed * Cost * Other (please specify) |
| Q7 | Please select the automated cell identification tool you are most familiar with.  (Answer choices as in Supplementary Table 1) |
| Q8 | Please mark your response about the software in Q7 to the following statements:  (Answer choices for each statement: Strongly disagree/ Disagree/ Neither agree nor disagree/ Agree/ Strongly agree)   1. I think that I would like to use this software frequently. 2. I found the software unnecessarily complex. 3. I thought the software was easy to use. 4. I think that I would need the support of a technical person to be able to use this software. 5. I found the various functions in this software were well integrated. 6. I thought there was too much inconsistency in this software. 7. I would imagine that most people would learn to use this software very quickly. 8. I found the software very awkward to use. 9. I felt very confident using the software. 10. I needed to learn a lot of things before I could get going with this software. |

Supplementary figures



Supplementary Figure 1. Software citations by computational method. Citations by algorithm class, showing unsupervised methods with graphical user interfaces (GUIs) only.



Supplementary Figure 2. Results of a survey of clinical laboratories on the use of automated flow cytometry software.

References

1. Amir EAD, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, et al. ViSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol. 2013;31(6):545–52.

2. Qiu P, Simonds EF, Bendall SC, Gibbs KD, Bruggner R V, Linderman MD, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol. 2011;29(10):886–93.

3. Maaten L van der, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9(Nov):2579–605.

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

5. Pyne S, Hu X, Wang K, Rossin E, Lin T-I, Maier LM, et al. Automated high-dimensional flow cytometric data analysis. Proc Natl Acad Sci. 2009 May 26;106(21):8519–24.

6. Bruggner R V., Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. Proc Natl Acad Sci. 2014;111(26):E2770–7.

7. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytom Part A. 2015;87(7):636–45.

8. Becher B, Schlitzer A, Chen J, Mair F, Sumatoh HR, Teng KWW, et al. High-dimensional analysis of the murine myeloid cell system. Nat Immunol. 2014;15(12):1181–9.

9. Aghaeepour N, Nikolic R, Hoos HH, Brinkman RR. Rapid cell population identification in flow cytometry data. Cytom Part A. 2011;79 A(1):6–13.

10. Shekhar K, Brodin P, Davis MM, Chakraborty AK. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). Proc Natl Acad Sci. 2014;111(1):202–7.

11. Bendall SC, Davis KL, Amir EAD, Tadmor MD, Simonds EF, Chen TJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. Cell. 2014;157(3):714–25.

12. Qian Y, Wei C, Eun-Hyung Lee F, Campbell J, Halliley J, Lee JA, et al. Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data. Cytom Part B Clin Cytom. 2010;78B(S1):S69–82.

13. Lo K, Hahne F, Brinkman RR, Gottardo R. flowClust: a Bioconductor package for automated gating of flow cytometry data. BMC Bioinformatics. 2009 May 14;10(1):145.

14. Finak G, Bashashati A, Brinkman R, Gottardo R. Merging Mixture Components for Cell Population Identification in Flow Cytometry. Adv Bioinformatics. 2009;2009:1–12.

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

16. Zare H, Shooshtari P, Gupta A, Brinkman RR. Data reduction for spectral clustering to analyze high throughput flow cytometry data. BMC Bioinformatics. 2010;11.

17. Ge Y, Sealfon SC. Flowpeaks: A fast unsupervised clustering for flow cytometry data via K-means and density peak finding. Bioinformatics. 2012;28(15):2052–8.

18. Finak G, Frelinger J, Jiang W, Newell EW, Ramey J, Davis MM, et al. OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis. PLoS Comput Biol. 2014;10(8).

19. Boedigheimer MJ, Ferbas J. Mixture modeling approach to flow cytometry data. Cytom Part A. 2008;73(5):421–9.

20. Cron A, Gouttefangeas C, Frelinger J, Lin L, Singh SK, Britten CM, et al. Hierarchical Modeling for Rare Event Detection and Cell Subset Alignment across Flow Cytometry Samples. Altan-Bonnet G, editor. PLoS Comput Biol. 2013 Jul 11;9(7):e1003130.

21. Malek M, Taghiyar MJ, Chong L, Finak G, Gottardo R, Brinkman RR. FlowDensity: Reproducing manual gating of flow cytometry data by automated density-based cell population identification. Bioinformatics. 2015;31(4):606–7.

22. Naim I, Datta S, Rebhahn J, Cavenaugh JS, Mosmann TR, Sharma G. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 1: Algorithm design. Cytom Part A. 2014;85(5):408–21.

23. 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. Cytom Part A. 2014;85(5):422–33.

24. Van Unen V, Höllt T, Pezzotti N, Li N, Reinders MJT, Eisemann E, et al. Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. Nat Commun. 2017;8(1).

25. Sugar IP, Sealfon SC. Misty Mountain clustering: application to fast unsupervised flow cytometry gating. BMC Bioinformatics. 2010;11(1):502.

26. Lin L, Finak G, Ushey K, Seshadri C, Hawn TR, Frahm N, et al. COMPASS identifies T-cell subsets correlated with clinical outcomes. Nat Biotechnol. 2015;33(6):610–6.

27. Rogers WT, Moser AR, Holyst HA, Bantly A, Mohler ER, Scangas G, et al. Cytometric fingerprinting: Quantitative characterization of multivariate distributions. Cytom Part A. 2008;73(5):430–41.

28. Sörensen T, Baumgart S, Durek P, Grützkau A, Häupl T. immunoClust-An automated analysis pipeline for the identification of immunophenotypic signatures in high-dimensional cytometric datasets. Cytom Part A. 2015;87(7):603–15.

29. Pyne S, Lee SX, Wang K, Irish J, Tamayo P, Nazaire MD, et al. Joint modeling and registration of cell populations in cohorts of high-dimensional flow cytometric data. PLoS One. 2014;9(7).

30. O’Neill K, Jalali A, Aghaeepour N, Hoos H, Brinkman RR. Enhanced flowType/RchyOptimyx: a bioconductor pipeline for discovery in high-dimensional cytometry data. Bioinformatics. 2014;30(9):1329–30.

31. Dundar M, Akova F, Yerebakan HZ, Rajwa B. A non-parametric Bayesian model for joint cell clustering and cluster matching: Identification of anomalous sample phenotypes with random effects. BMC Bioinformatics. 2014;15(1):1–15.

32. Li H, Shaham U, Stanton KP, Yao Y, Montgomery RR, Kluger Y. Gating mass cytometry data by deep learning. Bioinformatics. 2017;33(21):3423–30.

33. Meehan S, Walther G, Moore W, Orlova D, Meehan C, Parks D, et al. AutoGate: automating analysis of flow cytometry data. Immunol Res. 2014;58(2–3):218–23.

34. Van Gassen S, Vens C, Dhaene T, Lambrecht BN, Saeys Y. FloReMi: Flow density survival regression using minimal feature redundancy. Cytom Part A. 2016;89(1):22–9.

35. Anchang B, Do MT, Zhao X, Plevritis SK. CCAST: A Model-Based Gating Strategy to Isolate Homogeneous Subpopulations in a Heterogeneous Population of Single Cells. PLoS Comput Biol. 2014;10(7):13–7.

36. Lux M, Brinkman RR, Chauve C, Laing A, Lorenc A, Abeler-Dörner L, et al. FlowLearn: Fast and precise identification and quality checking of cell populations in flow cytometry. Bioinformatics. 2018;34(13):2245–53.

37. Lee HC, Kosoy R, Becker CE, Dudley JT, Kidd BA. Automated cell type discovery and classification through knowledge transfer. Bioinformatics. 2017;33(11):1689–95.

38. Rebhahn JA, Roumanes DR, Qi Y, Khan A, Thakar J, Rosenberg A, et al. Competitive SWIFT cluster templates enhance detection of aging changes. Cytom Part A. 2016;89(1):59–70.

39. Qiu P. Toward deterministic and semiautomated SPADE analysis. Cytom Part A. 2017;91(3):281–9.

40. Commenges D, Alkhassim C, Gottardo R, Hejblum B, Thiébaut R. cytometree: A binary tree algorithm for automatic gating in cytometry analysis. Cytom Part A. 2018;93(11):1132–40.

41. Lee AJ, Chang I, Burel JG, Lindestam Arlehamn CS, Mandava A, Weiskopf D, et al. DAFi: A directed recursive data filtering and clustering approach for improving and interpreting data clustering identification of cell populations from polychromatic flow cytometry data. Cytom Part A. 2018;93(6):597–610.

42. Weber LM, Nowicka M, Soneson C, Robinson MD. diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering. Commun Biol. 2019;2(1).

43. Reiter M, Rota P, Kleber F, Diem M, Groeneveld-Krentz S, Dworzak M. Clustering of cell populations in flow cytometry data using a combination of Gaussian mixtures. Pattern Recognit. 2016;60:1029–40.

44. Abdelaal T, van Unen V, Höllt T, Koning F, Reinders MJT, Mahfouz A. Predicting Cell Populations in Single Cell Mass Cytometry Data. Cytom Part A. 2019;95(7):769–81.

45. Folcarelli R, van Staveren S, Bouman R, Hilvering B, Tinnevelt GH, Postma G, et al. Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm. Sci Rep. 2018;8(1):1–18.

46. Hejblum BP, Alkhassim C, Gottardo R, Caron F, Thiébaut R. Sequential Dirichlet process mixtures of multivariate skew t-distributions for model-based clustering of flow cytometry data. Ann Appl Stat. 2019;13(1):638–60.

47. Wong L, Hill BL, Hunsberger BC, Bagwell CB, Curtis AD, Davis BH. Automated analysis of flow cytometric data for measuring neutrophil CD64 expression using a multi-instrument compatible probability state model. Cytom Part B - Clin Cytom. 2015;88(4):227–35.

48. Zaunders J, Jing J, Leipold M, Maecker H, Kelleher AD, Koch I. Computationally efficient multidimensional analysis of complex flow cytometry data using second order polynomial histograms. Cytom Part A. 2016;89(1):44–58.

49. 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;37(12):1482–92.

50. Meehan S, Kolyagin GA, Parks D, Youngyunpipatkul J, Herzenberg LA, Walther G, et al. Automated subset identification and characterization pipeline for multidimensional flow and mass cytometry data clustering and visualization. Commun Biol. 2019;2(1):1–12.

51. Pouyan MB, Nourani M. Identifying Cell Populations in Flow Cytometry Data Using Phenotypic Signatures. IEEE/ACM Trans Comput Biol Bioinforma. 2017;14(4):880–91.

52. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37(1):38–47.