Thinking Outside the Gate:

Single-Cell Assessments in Multiple Dimensions

Pia Kvistborg, Cécile Gouttefangeas, Nima Aghaeepour, Angelica Cazaly, Pratip K. Chattopadhyay, Cliburn Chan, Judith Eckl, Greg Finak, Sine Reker Hadrup, Holden T. Maecker, Dominik Maurer, Tim Mosmann, Peng Qiu, Richard H. Scheuermann, Marij J.P. Welters, Guido Ferrari, Ryan R. Brinkman, and Cedrik M. Britten

Supplemental Acknowledgements

Table S1. Freely Available Software for Computational Analysis of Flow/Mass Cytometry Data

Type/Name	Use	Reference	Technical Notes
Pre-processing			
EMMIX-JCM	Reduce batch data variability	PMID: 24983991	R package that uses joint clustering and matching to construct an overall batch template
FCSClean	Identify outlier events (i.e. fluorescence vs. time)	http://www.genepattern.org/	R software that removes spurious events based on time vs. fluorescence
FCSTrans	Manipulate file formats	PMC3932304	R package for FCS to .txt conversion
fdaNorm	Adjust data to account for batch effects	PMC3648208	R/Bioconductor software to adjust data to account for batch effects like laser drift
flowCore	Read/write, process (transform, compensate) of flow data. The basic flow infrastructure.	PMC2684747	R/Bioconductor core infrastructure for representing cell populations and the parent/child relationships amongst them
flowQ	Identify outlier samples (e.g. wells drying out, reagent issues)	PMC2768034	R/Bioconductor package that provides infrastructure to generate interactive HTML quality reports
flowStats	Advanced statistical methods and functions, specialized and general gating algorithms	http://www.bioconductor.org/	R/Bioconductor software that collects several algorithms together for normalization and gating
flowUtils	Import gates, transformation and compensation	http://www.bioconductor.org/	R/Bioconductor package to support Gating-ML specification to exchange gate coordinates between softwares
flowWorkspace	Import manually gated data from FlowJo workspaces, represent manual and automated gating hierarchies efficiently	PMC3992339	R/Bioconductor core infrastructure that makes manually gated data accessible to BioConductor's computational flow tools by importing pre-processed and gated data from FlowJo
guassNorm	Adjust data to account for batch effects	PMC3648208	R/Bioconductor software to adjust data to account for batch effects like laser drift
MIFlowCyt	Checklist for describing flow cytometry experiments	PMC2773297	Minimum information guideline to aid description of experiments in manuscripts
ncdfFlow	Advanced method for large dataset processing	PMC3992339	R/Bioconductor package that overcomes memory limitations when working with large datasets by

			storing FCS data in netCDF files on disk
plateCore	Analyze multiwell plates	PMC2777006	R/Bioconductor package that enable automated negative control-based gating and plate-based analysis
QUALIFIER	Identify outlier samples	PMC3499158	R/Bioconductor software that uses manual gates to perform an extensive series of statistical quality assessment checks on gated cell sub-populations
ReFlow	Manage and annotate single or multi-center FCS data	https://github.com/whitews/ReFlow Submitted manuscript	FCS data management framework that provides a modern web interface for harmonizing FCS annotation especially across multiple laboratories
Automated Gating			
ACCENSE	Unsupervised cell population identification	PMC3890841	R/Matlab software for dimensionality reduction with density-based partitioning
ASPIRE	Unsupervised cell population identification	PMID: 25248977	Matlab code that performs joint cell clustering and cluster matching in the presence of random effects
CCAST	Unsupervised cell population identification	PMID: 25078380	R package that combines any clustering algorithm with silhouette measures followed by recursive partitioning to generate a decision tree based gating strategy
FindingPopulations	Unsupervised cell population identification	PMID: 25170025	R script for cell population identification using Fluorescence-Minus-One controls
FLAME	Unsupervised cell population identification	PMC2682540	Multivariate finite mixtures of skew & tailed distributions
FLOCK	Unsupervised cell population identification	PMC3084630	Stand alone software for clustering using an adaptive multi-dimensional mesh to estimate local density followed by hierarchical merging of adjacent regions based on density differentials
flowCL	Annotate cell populations using formal ontologies	http://www.bioconductor.org/ PMID: 25481008	R/Bioconductor software to label cell populations with semantic identifier from the Cell Ontology library
flowClust	Unsupervised cell population identification	PMC2701419	R/Bioconductor software for clustering using t-mixture model with Box-Cox transformation with support for Bayesian priors
flowDensity	Supervised cell population identification to automate manual gating process	PMID: 25378466	R/Bioconductor software for supervised gating to match manual analysis for clinical trials and diagnosis
flowFP	Unsupervised cell population identification	PMC2777013	R/Bioconductor software for fingerprint generation via multivariate probability distribution

flowMatch	Cell population matching	PMC3471348	R/Bioconductor software to match clusters across samples for producing robust meta-clusters
FlowMap-FR	Cell population matching	https://github.com/JoyceHsiao/flowMap-release	R/Bioconductor software to match cell population clusters across samples using the F-R statistic
flowMeans	Unsupervised cell population identification	PMC21182178	R/Bioconductor software for k-means clustering and merging using the Mahalanobis distance
flowMerge	Unsupervised cell population identification	PMC2798116	R/Bioconductor software that combines flowClust and entropy-based or mahalanobis distance-based cluster merging
flowPeaks	Unsupervised cell population identification	PMC3400953	R software for unsupervised clustering using k-means and mixture model
flowPhyto	Unsupervised cell population identification for marine biology	PMID: 21208987	R/Bioconductor software for analysis of marine biology data
flowQB	Identify bead populations for instrument setup	http://www.bioconductor.org/	R/Bioconductor software for analysis of detector efficiency and optical background for maching sensitivity analysis
flowTrans	Estimate parameters for data transformation	PMC3243046	R/Bioconductor infrastructure to optimize parameter choice for different transformations
flowType	Unsupervised cell population identification	PMC3998128	R/Bioconductor software for combinatorial gating of high-dimensional populations and correlative analysis against clinical outcomes
NetFCM	Semi-automated web-based method for flow cytometry data analysis	PMID: 25044796	Semi-automatic gating strategy that uses clustering and principal component analysis (PCA) together with other statistical methods to mimic manual gating approaches
NMF-curvHDR	Unsupervised cell population identification	PMC2832899	Density-based clustering and non-negative matrix factorization
OpenCyto	General framework to construct reproducible automated gating pipelines and simplifly data processing	http://opencyto.org/	R/Bioconductor infrastructure for hierarchical automated gating that maintains relationships amongst cell populations
py-fcm	Flow cytometry data management and Bayesian event modeling	PMID: 23874174	A collection of tools in Python for parsing and analysis of FCS files, with a focus on nonparametric Bayesian statistical mixture models for unsupervised cell population identification
SamSpectral	Supervised cell population identification	PMC2923634	R/Bioconductor software for efficient spectral clustering using density-based down-sampling
SPADE	Unsupervised cell population identification	PMC3196363	Matlab/Standalone/R/Bioconductor tool for density-based sampling, k-means clustering and minimum

			spanning trees
SWIFT	Unsupervised gating for rare cell populations	PMID: 24677621	Iterative weighted sampling procedure with splitting and merging to retain discrimination of extremely small subpopulations
X-cyt	Supervised cell population identification to automate manual gating process	PMC3839720	R script that partitions each sample with initialization by user template, then optimizes on the parameters via estimation-maximization
Data Analysis			
BioConductor	Open source software for bioinformatics	PMC545600	Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development
Citrus	Identify most important cell populations correlated with outcome of interest	PMID: 24979804	Regularized supervised learning algorithms to identify stratifying clusters that are the best predictors of a known experimental endpoint of interest
COMPASS	Identify combinatorial subsets of polyfunctional T cells	http://rglab.github.io/COMPASS/	R software that is multivariate extension of MIMOSA that jointly models all combinatorial polyfunctional cell subsets
GenePattern	Interface to multi-step analysis pipelines	PMC3717030	Web interface and cloud computing interface for over 30 R/BioConductor packages
MIMOSA	Identify responders and non-responders to stimulation in intracellular cytokine staining assay data	PMC3862207	R/Bioconductor software to detect antigen-specific changes in marginal or specific cell subsets
Post-processing			
CytoBank	Share and analyze data associated with flow and mass cytometry experiments	PMID: 20578106	Cloud-based solution for organizing, analyzing and visualizing single-cell data
flowPlots	Graphical displays with statistical tests for gated ICS flow cytometry data	http://www.bioconductor.org/	R/Bioconductor software that provides analysis plots and data class for gated flow cytometry data
flowViz	Visualization (e.g. histograms, dot plots, density plots, gating hierarchies and layouts)	PMC2768483	R/Bioconductor software that employs Trellis graphics and can be adapted to provide useful visualizations
FlowRepository	Share flow and mass cytometry data associated with peer-reviewed manuscripts,	PMID: 22887982	Open, public repository for flow cytometry data including FCS files and experiment settings

	annotated according to MIFlowCyt		
flowSOM	Enables the visualization of complex flow data	PMID: 25573116	R/Bioconductor software that uses Self-Organizing Maps
gEM/GANN	Extract feature sets from datasets	PMID: 25572884	Combines of WEKA and C++ feature selections program using Genetic Algorithm-Neural Network
ImmPort	Share and analyze genomic, proteomic and other data	PMID: 24791905	Production, analysis, archiving, and exchange of scientific data for the diverse community of life science researchers. Supported by NIAID/DAIT
RchyOptimyx	Identify most important cell populations correlated with outcome of interest	PMC3998128	R/Bioconductor software that optimizes cellular hierarchies to preserve correlation with external variables and summarize large datasets in simple plots
viSNE	Visualize high-dimensional data	PMC4076922	Matlab software that uses Stochastic Neighbor Embedding for visualizing high-dimensional data by giving each datapoint a location in a two or three-dimensional map

Published tools are classified in four categories, namely pre-processing, automated gating, data analysis, and post-processing.