fortune Journals

Research Article

JOURNAL OF BIOINFORMATICS AND SYSTEMS BIOLOGY

ISSN: 2688-5107

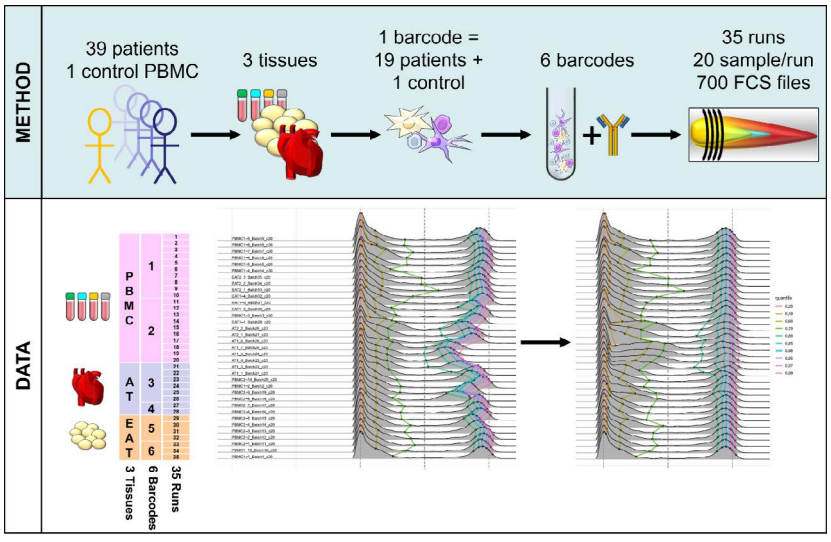
CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data

Samuel Granjeaud[1], Naoill Abdellaoui[2,3], Anne-Sophie Chrétien[1,4], Eloise Woitrain[5], Laurent Pineau[5], Sandro Ninni[5], Alexandre Harari[2,3], Marion Arnaud[2,3], David Montaigne[5], Bart Staels[5], David Dombrowicz[5], Olivier Molendi-Coste[5,6,*]

Abstract

Innovation in cytometry propelled it to an almost "omic" dimension technique during the last decade. The application fields concomitantly enlarged, resulting in generation of high-dimensional high-content data sets which have to be adequately designed, handled and analyzed. Experimental solutions and detailed data processing pipelines were developed to reduce both the staining conditions variability between samples and the number of tubes to handle. However, an unavoidable variability appears between samples, barcodes, series and instruments (in multicenter studies) contributing to "batch effects" that must be properly controlled. Computer aid to this aim is necessary, and several methods have been published so far, but configuring and carrying out batch normalization remains unintuitive for scientists with a purely biological university education. To address this challenge, we developed an R package called CytoBatchNorm that offers an intuitive and user-friendly graphical interface. Although the processing is based on the script by Schuyler et al., the graphical interface revolutionizes its use. CytoBatchNorm enables users to define a specific correction for each marker in a single run. It provides a visualization that guides you through quickly setting the correction for each marker. It allows corrections to be previewed and inter-marker effects to be checked as the settings are updated. CytoBatchNorm will help the cytometry community to adequately scale data between batches, reliably reducing batch effects and improving subsequent dimension reduction and clustering.

Visual Abstract



Affiliation:

[1]Centre de Recherche en Cancérologie de Marseille (CRCM), INSERM, U1068, CNRS, UMR7258, Institut Paoli-Calmettes, Aix-Marseille University, UM105, Marseille, France

[2]Ludwig Institute for Cancer Research, Lausanne Branch, Department of Oncology, University of Lausanne (UNIL) and Lausanne University Hospital (CHUV), Agora Cancer Research Center, Lausanne, Switzerland

[3]Center for Cell Therapy, CHUV-Ludwig Institute, Lausanne, Switzerland

[4]Immunomonitoring Department, Institut Paoli-Calmettes, Marseille, France

[5]University Lille, INSERM, CHU Lille, Pasteur Institute of Lille, U1011-EGID, Lille, France

[6]University Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US41 – UAR 2014 – PLBS, F-59000 Lille, France

***Corresponding author:**
Olivier Molendi-Coste, University Lille, INSERM, CHU Lille, Pasteur Institute of Lille, U1011-EGID, Lille, France; University Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US41 – UAR 2014 – PLBS, F-59000 Lille, France

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Received: October 18, 2024
Accepted: October 24, 2024
Published: December 27, 2024

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

Keywords: Mass cytometry; Flow cytometry; Batch effect; Batch effect correction; R package.

Abbreviations: AT=Atrial Tissue, CyTOF=Cytometry by Time Of Flight, EAT=Epicardial Adipose Tissue, PBMC=Peripheral Blood Mononuclear Cells, QC=quality control, DWF=Detailed Workflow.

Introduction

Innovation in cytometry propelled it to an almost "omic" dimension technique during the last decade, with the rising of both cytometry by time-of-flight (CyTOF, or mass cytometry) and spectral cytometry, allowing elaboration of panels up to 30-50 parameters analyzed at the single cell level[1]. The application fields concomitantly enlarged, with data sets that include more and more (i) samples - specifically in human clinical studies, (ii) tissues from a single donor, and (iii) stimulated conditions for a single sample, all leading to an increase in the number of "tubes" composing an experiment. This results in generation of high-dimensional high-content data which have to be adequately designed, handled and analyzed[2].

Experimental solutions and detailed samples processing pipelines were developed to reduce both the staining conditions variability between tubes[3] and the number of tubes to handle/acquire, the later mainly consisting in barcoding of samples[4,5]. However, barcoding has its intrinsic limits in the number of samples that can be barcoded together, and acquisition of a single barcode (composed of tens of millions of cells) is very long and usually achieved in many runs, notably in mass cytometry in regard with its limited acquisition rate speed. Thus, large datasets are usually obtained during acquisitions spanning days or weeks. Despite the maximal attention paid by experimenters in standardizing their protocols, sequential manipulation and instrument instability over time[3] introduce variability between samples, barcodes and runs. The raise of multi-center studies also introduces an instrument-dependent effect that needs to be corrected to enhance data consistency[6]. All these sources of variations in results baseline participate to the commonly used "batch effects" denomination and have to be controlled properly[7–11].

Batch effects can be addressed at two levels of the experimental protocol. The instrument-related batch effect is mainly dependent on tuning and quality control (QC) accuracy, specific instrument sensitivity and sensitivity variations along acquisitions. Normalization methods based on calibration beads were developed for both flow[12,13] and mass cytometry[5] to counteract this variability level, and EQbeads (Standard BioTools Inc, San Francisco, CA, US) were structurally included in mass cytometers design. However, these methods have their own limitations as illustrated below and by others[3,14]. Based on external controls (beads), they only aim to standardize the instrument. The experimental-related batch effects basically regroup all parameters of an experiment that can influence staining and detection efficiencies of samples, in addition to the instrument used for acquisition.

To evaluate signal stability over batches/days, solutions have been developed, such as the inclusion of a "control" aliquot (i.e. a standard sample) in all batches. This design served as a substrate for the development of algorithms to correct batch effects[9]. Methods and software packages published so far[15–23] circumvent batch variations in channels intensity using either an "all-events" or a "pre-gated population-specific" adjustment. Whatever the package/method considered, configuration as well as application of the normalization process across batches remains either obscure or unintuitive for scientists with "pure" academic background in biology (Table 1), which usually and historically lack training in computer science and mathematics.

Table 1: Main published packages for batch effect correction

Method	Based on	Define batches	Includes a clustering	Requires a standard sample	Align	Parameter to be chosen	Visual interface	Normalization preview
iMuBAC	harmony	✓	✓			auto	No	No
gaussNorm, fdaNorm	-	× [1]			detected peaks	auto	No	No
cyCombine	Combat	✓	✓			auto	No	No
CytoNorm	-	✓	✓	✓	101 quantiles	auto	No	No
cytofRUV	RUV	✓				depth	No	No
CytofBatchAdjust	-	✓		✓	1 percentile	percentile	No	No
CytoBatchNorm	CytofBatchAdjust	✓		✓	1 (or 2) percentiles	percentile	Yes	Yes

1: as no batch is defined, all samples are aligned; when batch information is available, the transform defined within each batch is applied to every sample of the batch, which removes differences between batches but keeps differences within each batch.

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

The CytofBatchAdjust R script released by Schuyler et al. in 2019[19] smartly proposed an "all-events" adjustment of peaks intensities for each channel of the control tube included in each batch to the peaks intensities of the control tube of a referent batch. This adjustment is performed channel by channel, independently of each other. It can be achieved with different options, scaling on quantiles or on a specified percentile, considering or not zero values and with or without arcsinh transformation of data. As depicted by the authors, the "quantile" adjustment is not adequate for normalization of cytometry data, because it sometimes creates artefacts that can only be identified on a bi-parametric plot. Indeed, when the distribution of events in the control sample differs even slightly between the reference batch and the other batches (e.g. the proportion of events in the positive and negative peaks), the "quantile" adjustment shifts events from one peak to the other. By contrast, the "percentile" method allows the user to choose a single percentile value (from 1 to 99) which identifies an intensity in each batch and linearly scales each batch in order to align those intensities to the intensity of the reference batch. Determining the best percentile for a given channel has so far been tricky and empirical, mainly because users lack an indicative visualization. If the percentile is chosen in a region where the distribution of events varies between the control tubes of different batches, the channel scaling will be aberrant in some batches, introducing computational biases in the downstream data analysis, which would only be identified after the correction has been applied.

To provide a comprehensive, transparent and interactive pipeline to reliably overcome the different levels of batch variability, we developed the CytoBatchNorm R package, which is adapted from CytofBatchAdjust code and methodology[19]. We illustrate its use with an experiment of 700 FCS files from 35 batches including 3 human tissues and benchmark it on two different data sets versus the two most-commonly used packages: CytofBatchAdjust[19] and CytoNorm[18]. This package is intended to be "ready-to-use" for experimenters without any R coding skills. We improved the CytofBatchAdjust R script by providing (i) an intuitive user-friendly interface, (ii) a visualization for each marker to help selecting the best percentile relative to the distribution of intensity across the control samples in each batch before proceeding with normalization, (iii) a comprehensive table to specify the percentile channel-by-channel, (iv) the ability to perform a bi-percentile adjustment, (v) dot-plots and output graphs to control the batch adjustment accuracy, (vi) code for the Windows system.

Methods

Dataset 1cxv

Datasets structures are summarized in Table 2

Table 2: Datasets structure

Datasets	Panel Size	Number of batches	Number of patients	Barcoding	Number of output files	Nature of control sample
Dataset 1	46	35	38	yes	700	Frozen healthy patient PBMC
Dataset 2	44	5	14	no	19	Frozen healthy patient PBMC

Sample collection

Atrial Myocardial Tissue (AT) obtained from the right atrial appendage before aortic cross-clamping and cardioplegia, Epicardic Adipose Tissue (EAT) and PBMC from 38 patients undergoing a Surgical Aortic Valve Replacement (SAVR) (POMI-AFclinical study NCT#03376165; PI: D Montaigne) as well as PBMC from a single healthy donor were analyzed by mass cytometry. Written informed consent was obtained from all patients before inclusion. Non-parenchymal cells fraction from AT and EAT were prepared from tissues immediately cleaned, minced and digested at 37°C for 45 min in 0.1% collagenase I and PBMC were obtained as previously described[24].

Staining protocol

Detailed antibody panel and staining protocol were described previously[24]. Briefly, thawed PBMCs ($3x10^6$ cells) and non-parenchymal cells from AT and EAT ($40x10^3$ to $1.4.10^6$ cells) were sequentially stained for viability, for extracellular targets sensitive to fixatives, barcoded, stained for extracellular markers not sensitive to fixatives, for intracellular targets and for DNA as summarized in Table 3. Six independent pools of 19 experimental samples (2 pools of PBMC, 2 pools of AT, 2 pools of EAT) added with one aliquot of the control PBMC sample from a single healthy donor were processed independently for barcoding (each pool for a total of 20 samples), as summarized in Graphical Abstract.

Acquisition

Cells were resuspended in Maxpar Water at $5x10^5$ cells/mL with 1:10 volume of Four-Element Calibration Beads (Standard BioTools Inc, San Francisco, CA, US) and analyzed on an Helios instrument (Standard BioTools Inc, San Francisco, CA, US). Runs were carried out for a maximum of 150 minutes at a maximum rate of 500 events/sec, and a quick tuning was performed between each run.

Datasets 2

Datasets structures are summarized in Table 2

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

Table 3: Antibody panel from Dataset 1

Target	Clone	Antibody Isotype	Metal Tag	Barcoding fixative sensitivity	Intracellular	Hand coupling/ polymer	Staining Concentration	Kit Supplier	Kit reference	Antibody Supplier	Antibody reference	Antibody lot
CD45	HI30	Mouse IgG1, κ	89Y				1 µg/100µL			Std BioTools	3089003B	1151903
CD282	W15145C	Mouse IgG2a, κ	103Rh			Maxpar X8	2 µg/100µL	Std BioTools	201140A	Biolegend	392302	B250232
Pan-CEA	TET2	Mouse g G2bka	113In			Maxpar X8	2 µg/100µL	Std BioTools	201140A	Cliniscience	sc-59875	I2806
Connexin 43	578618	Mouse IgG2A	115In			Maxpar X8	2 µg/100µL	Std BioTools	201140A	R&D System	MAB7737	CGTK0118061
CD8	SK1	Mouse IgG1, κ	116Cd			Maxpar MCP9	2 µg/100µL	Std BioTools	201116A	Biolegend	344727	B265197
Perforin	dG9	Mouse IgG2b, κ	139La		X	LightningKit	2 µg/100µL	Expedeon	M139-0100	Biolegend	308102	B267038
CCR2	K036C2	Mouse IgG2a, κ	140Ce			Maxpar X8	2 µg/100µL	Std BioTools	201140A	Biolegend	357202	B240747
CD196/CCR6	G034E3	Mouse IgG2b, κ	141Pr	Yes			1 µg/100µL			Std BioTools	3141003A	3471805
CD19	HIB19	Mouse IgG1, κ	142Nd				1 µg/100µL			Std BioTools	3142001B	381907
CD127	A019D5	Mouse IgG1, κ	143Nd	Yes			1 µg/100µL			Std BioTools	3143012B	2501815
CD38	HIT2	Mouse IgG1, κ	144Nd	Yes			1 µg/100µL			Std BioTools	3144014B	2991817
CD163	GHI/61	Mouse IgG1, κ	145Nd				1 µg/100µL			Std BioTools	3145010B	2351602
IgD	IA6-2	Mouse BALB/c IgG2a, κ	146Nd				1 µg/100µL			Std BioTools	3146005B	671816
CD20	2H7	Mouse IgG2b, κ	147Sm				1 µg/100µL			Std BioTools	3147001B	2601808
CD34	581	Mouse IgG1, κ	148Nd	Yes			1 µg/100µL			Std BioTools	3148001B	1841814
CD194	L291H4	Mouse IgG1, κ	149Sm	Yes			1 µg/100µL			Std BioTools	3149029A	2991816
CD64	10.1	Mouse IgG1, κ	150Nd			Maxpar X8	2 µg/100µL	Std BioTools	201150A	Biolegend	305029	B245146
CD123	6H6	Mouse IgG1	151Eu				1 µg/100µL			Std BioTools	3151001B	2431808
TCRgd	11F2	Mouse BALB/c IgG1	152Sm	Yes			1 µg/100µL			Std BioTools	3152008B	441915
CD185	RF8B2	LOU/M IgG2b, κ	153Eu				1 µg/100µL			Std BioTools	3153020B	1621806
CD3	UCHT1	Mouse IgG1, κ	154Sm				1 µg/100µL			Std BioTools	3154003B	71917
CD27	L128	Mouse BALB/c IgG1	155Gd	Yes			1 µg/100µL			Std BioTools	3155001B	1731805
CD183	G025H7	Mouse IgG1, κ	156Gd	Yes			1 µg/100µL			Std BioTools	3156004B	2991818
Fox P3	259D/C7	Mouse BALB/c IgG1	157Gd		X	Maxpar X8	2 µg/100µL	Std BioTools	201157A	BD	560044	817636
CD33	WM53	Mouse IgG1, κ	158Gd				1 µg/100µL			Std BioTools	3158001B	1031707
CD11c	Bu15	Mouse IgG1, κ	159Tb				1 µg/100µL			Std BioTools	3159001B	2991809
CD28	CD28.2	Mouse IgG1, κ	160Gd				1 µg/100µL			Std BioTools	3160003B	3181807

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

Sample collection

In vitro-expanded tumor-infiltrating lymphocytes (TILs) were generated from tumor samples of 14 cancer patients, using an improved TIL culture method [25]. PBMCs from a single healthy donor were used as control sample. Samples were collected and biobanked from patients enrolled under protocols approved by the Lausanne university hospital (CHUV), Switzerland. Patients and healthy donors' recruitment, study procedures, and blood withdrawal were approved by regulatory authorities and all patients signed written informed consents.

Staining protocol

Frozen TILs ($1x10^6$ to 5.10^6 cells) and PBMCs ($3x10^6$ cells) were thawed, rested overnight, and labelled with 44 metal-coupled antibodies (Standard BioTools Inc, San Francisco, CA, US & in house). Cells were first stained for viability then stained for extracellular targets (Standard BioTools 400276 Protocol).

Cells were then fixed using Cytofix fixation buffer (BD Biosciences) and permeabilized using Phosflow Perm Buffer III solution (BD Biosciences). Next, intracellular staining was performed (antibody incubation: 30 min at RT). For DNA staining, cells were next incubated overnight at 4°C with cell intercalation solution following the manufacturer's protocol (Standard BioTools 400276 Protocol).

Acquisition

Cells were finally washed and resuspended in a Maxpar Cell Acquisition Solution containing EQ Four Element Calibration Beads (Standard BioTools Inc, San Francisco, CA, US) at a cell concentration of 10^6 cells/mL, immediately prior to CyTOF data acquisition. Five runs at different days were realized using an Helios Mass Cytometer (Standard BioTools Inc, San Francisco, CA, US) at a maximum rate of 500 events/sec.

Data pre-processing

Raw mass cytometry data (Datasets 1 and 2) were first normalized with the calibration EQ-bead passport pre-loaded in the CyTOF Software version 7 (Standard BioTools 400276 Protocol) and then debarcoded following the manufacturer's instructions (Dataset 1). Data pre-processing was performed using Cytobank (Beckman Coulter, Indianapolis, IN, USA) for Dataset 1 or FlowJo 10 (Becton Dickinson, Franklin Lakes, NJ, US) for Dataset 2.

Batch effect normalization

Dataset 1

The 6 barcoded pools of samples all included an aliquot of a control PBMC sample from a single healthy donor (C20). Those 6 barcoded pools were acquired on an Helios instrument through a total of 35 different runs dispatched as follows: PBMC pool 1 – 10 runs; PBMC pool 2 – 10 runs; AT pool 1 – 6 runs; AT pool 2 – 2 runs; EAT pool 1 – 4 runs; EAT pool 2 – 3 runs (Graphical Abstract). Those 35 runs were thereafter considered as 35 independent batches in order to compensate for possible signal drift over duration of acquisition of a single barcoded pool.

The C20 included in each barcoded run was gated on nucleated - single - biological - non beads - $CD45^+$ live events (Supplemental Figure 1) and used as an anchor for application of the modified CytofBatchAdjust R code published by Schuyler et al.[19] as exposed in the RESULTS section. Batch-adjusted FCS files from each single sample were concatenated to reconstitute original samples.

Dataset 2

Five batches of samples, all including an aliquot of a control PBMC sample from a single healthy donor, were acquired on an Helios instrument over a total of 5 different runs (1 run per batch). The PBMC from the healthy donor included in each batch was gated on nucleated - single - biological - non beads - $CD45^+$ live events (Supplemental Figure 1) and used as an anchor for application of the modified CytofBatchAdjust R code published by Schuyler et al.[19]. Dataset 2 served for benchmarking against CytoNorm[18], which was performed as a plugin supplied by FlowJo software (Becton Dickinson, Franklin Lakes, NJ, US) following the provider's instructions.

Spillover compensation

Dataset 1

For compensation matrix calculation, Comp Beads (Becton Dickinson, Franklin Lakes, NJ, US) were single stained in CSB with 1µg of each antibody for 30 minutes at RT. Exception was made for the anti-CADM1 antibody labelled with 196Pt which is a chicken IgY, that is not captured by Comp Beads, which was replaced by an anti-CD8 (clone SK1, Mouse IgG1, κ) conjugated with the same batch of 196Pt. After two washes with CSB and two wash with Marxpar Water, beads were mixed and acquired as a single tube at a maximum rate of 500 events/sec. The FCS file from mixed single stained beads was imputed in CATALYST R[26] using the NNLS method. The output compensation matrix (Figure 6 C) was applied to all the files and compensated FCS files were edited and processed to data analysis.

Data analysis

Debarcoded - batch normalized – concatenated - compensated FCS files were gated on nucleated - single - biological - non beads - $CD45^+$ live events (Supplemental figure 1) and processed for phenotype analysis using R 4.0.0 and a modified version of the Cytofkit package[27] (http://github.com/i-cyto/cytofkitlab), including UMAP computation using the uwot package. Dimension reduction was performed using t-SNE or UMAP algorithms. Multi-

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

dimensional scaling was performed using the CytoMDS R package[28]. Illustrations were edited using the Cytofkit ShinyAPP browser[27] and Cytobank (Beckman Coulter, Indianapolis, IN, USA).

Software

The cytoBatchNorm package as well as recommendations and commands are described on the github repository: https://github.com/i-cyto/cytoBatchNorm. The package is simply installed and launched using the commands below in R or RStudio.

Download and installation:

devtools::install_github("i-cyto/cytoBatchNorm")

Launching:

library(cytoBatchNorm)

cytoBatchNormGUI()

Results and Discussion

Computer-assisted treatment of data, whatever the kind of treatment, has to be finely tuned and controlled carefully to avoid computational artefacts as well as data distortions that could deteriorate data consistency and induce uncontrolled bias in the downstream analysis pipeline. Considering batch effects correction, this burden should be taken in consideration cautiously especially considering the complexity of the dataset (number of batches, samples, markers), as both the risk of such data distortion and the possibility that the experimenter misses it during visual control of corrected data increase with the dimensionality of the dataset.

As a rational and accessible solution, we developed an intuitive, user-friendly tool based on R Shiny package that does not require any practice in R language or programming skills. This tool can easily be installed and launched using simple commands referenced in the "software" section. The Web interface allows intuitive, point-and-click navigation through the different steps of the normalization pipeline: selection of the dataset, identification of the different batches/control tubes and of the referent control tube, selection of the channels and tuning of the batch correction process as well as launching the correction.

A detailed workflow (DWF) is provided as supplemental data and describes the procedure step by step.

"Create Bunch" menu

The "Create Bunch" menu (Figure 1A, DWF steps 1-5) asks for the name of the current experiment to be created, storage directory, cytometry technology as well as the directory containing the FCS files to adjust. To finish, click on the "Create" button.

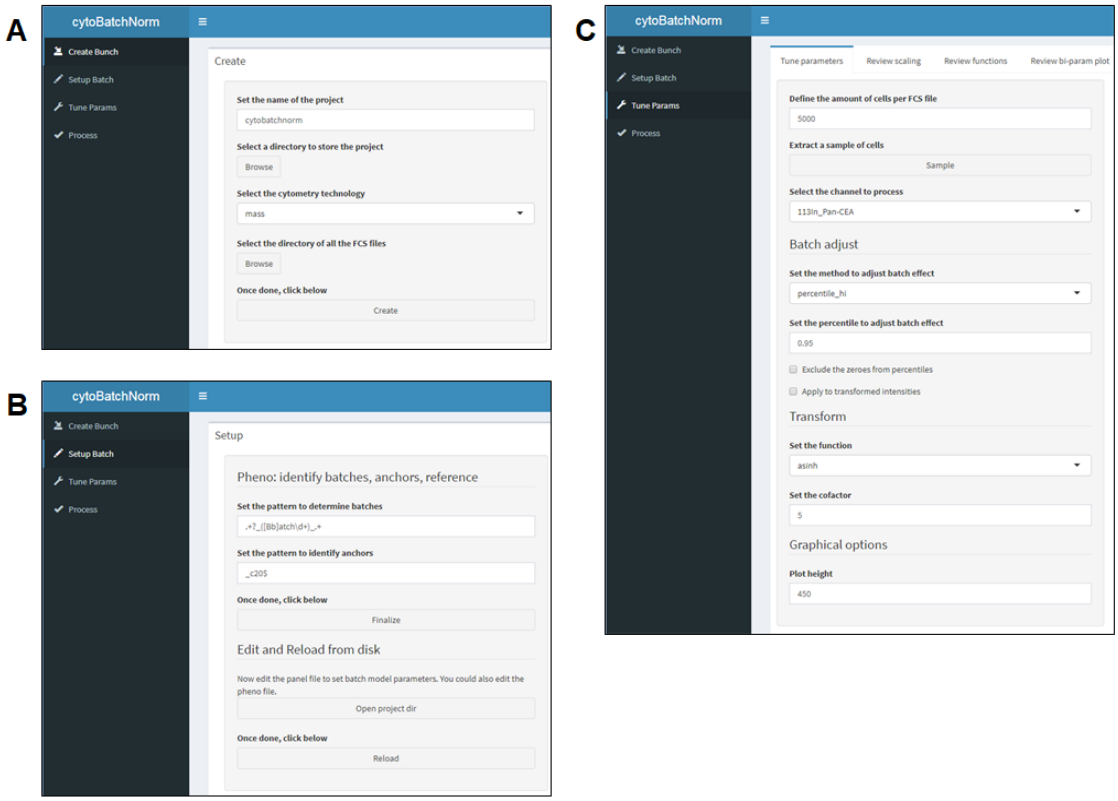


Figure 1: Sequential operation of the batch effect normalization interface. A: Bunch creation menu. B: Batch setup menu (detailed in figure 2). C: Tuning parameters menu (detailed in figure 3 and 4).

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

"Setup Batch" menu

The "Setup Batch" menu (Figure 1B, DWF steps 6-10) allows to define keywords for automatic identification of batches and control tubes (termed "anchor", "C20" in our example) for each batch inside the FCS files directory. The "finalize" button launches the identification process and creates two tables (.xlsx) "pheno" and "panel" which can be accessed using the "open project dir" button (Figures 1B and 2A-2, DWF step 8). To indicate the markers to be normalized and the control files, these tables must be modified and saved using appropriate spreadsheet software. Once all the information has been entered, the interface is updated by clicking on the "reload" button (Figure 2A-2 and 2B-2, DWF step 10). The "pheno" table (Figure 2A-3) lists the files identified as control tubes (column "sample_ID_is_ref" = "Y") and the reference batch control tube (column "sample_ID_is_ref" = "Y" and column "batch_is_ref" = "Y"). The choice of the reference batch is decided by the user, setting only one "Y" in the "batch_is_ref" column onto the proper row/file in the .xlsx "pheno" table, saving it, and then clicking on the "reload" button. The "panel" table (Figure 2B-3) summarizes the attributes of the FCS files loaded as well as the percentile value that will be used for each channel's adjustment (by default 0.95), and can be edited the same way as the "pheno" table, if necessary, saved and reloaded similarly.

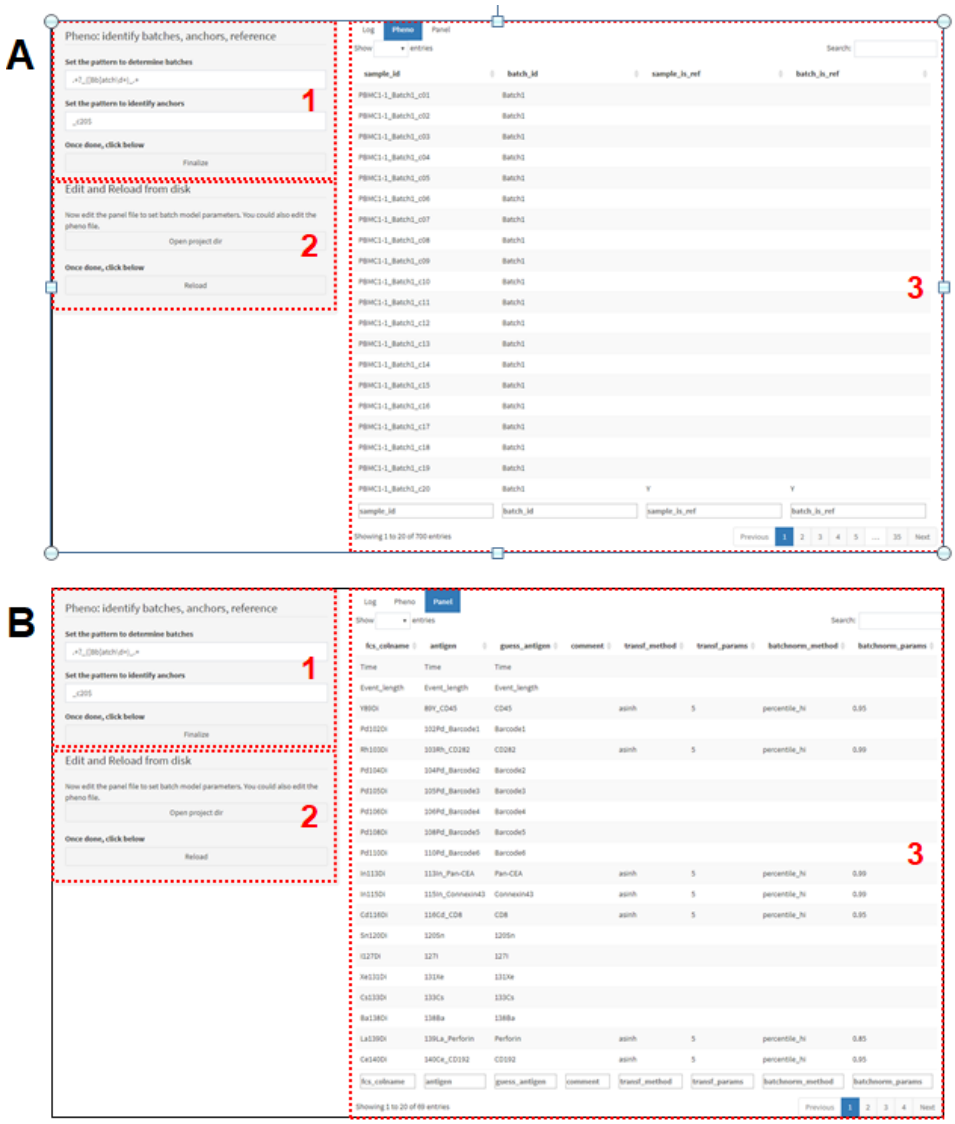


Figure 2: Batch setup menu. A: Tab for setup of the phenotype of the data. B: Tab for setup of the panel of the data. First step (A1, B1) is the identification of the batches and the referent control tubes of each batch using keywords. We strongly recommend to clearly standardize keywords in files nomenclature, i.e. « Batch » and « C20 » respectively in the example displayed. Identification is launched by clicking on the « finalize » button. Users can then check if this identification of batches and referent control tubes is accurate in the pheno tab (A3) and if the indexing of the panel is adequate (B3). Those tables can be modified with spreadsheet software using the .xlsx files available in the project directory, and then reloaded using the « reload » button (A2, B2).

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

Tune Params" menu

The CytoBatchNorm interface makes two major additions to the normalization settings. Firstly, the percentile is defined independently for each marker. Secondly, a display allows you to determine which percentile will be the most appropriate and most accurate for normalizing the whole experiment. The "Tune parameters" step (Figure 1C and Figure 3, DWF steps 11-18) offers a first tab which presents the histograms of all control tubes from all batches for the selected channel in a ridgeline plot and superimposes a line through histograms for each percentile of the default percentile set as illustrated in Figure 3A (0.20, 0.40, 0.60, 0.70, 0.80, 0.90, 0.95, 0.97, 0.99). Beforehand, a sampling of files has to be achieved using the "sample" button. Then, users can preview the effect of normalization with different percentiles and different transformations in the adjusted and raw intensity graphs. To refine and validate the previewed batch effect correction, we also developed a set of control plots presenting and comparing multiple batches at once. The "Review scaling" tab (Figure 3B) displays the scaling factors that would be applied with the selected parameters. The "Review functions" tab (Figure 3C) shows the raw vs pre-normalized events for the active channel which allows a direct control of the linearity of the normalization. The "Review bi-parameters plot" tab (Figure 3D) allows to build a dot-plots of the active channel vs another channel and to check whether an artefact would be induced by the normalization. These four tabs allow an enlightened determination of which percentile value is the most suitable to scale a given channel, considering the distribution of events along the range, some possible variations in this distribution across batches and notably questioning the relevance of choosing a percentile value in the positive or negative peak or in the positive "queue" if so.

In mass cytometry, zero-values can represent a large part of events depending on the couple marker/metal-tag, which raises questions. Excluding them may greatly change the distribution of percentiles along some channels. Beyond the "zero values" lays the consideration of the negative peak for a given channel, including in the case of unimodal distribution of the events. Indeed, correcting batch effect on the positive peak of a given channel (which is intuitive, as for CD8 in our illustration, Figure 3A) will rescale the entire channel in order to align the positive peak of the batch control tubes to the referent control, shifting values up or down along the whole scale. Depending on the amplitude of the scaling required, it is possible for almost negative events to be moved away from the commonly considered "negative zone" of a channel. To enable users to control this potential adverse effect of a single-percentile adjustment, we implemented a bi-percentile adjustment function, with the idea of defining one percentile for the alignment of the positive peak and another percentile for the alignment of the negative peak (DWF steps 14 and 15). As illustrated in Figure 4, this is performed by choosing the "percentile_lohi" option in the "method" menu and entering two percentile values separated with a comma in the "percentile" window (Figure 4A-3). As for the "percentile_hi" option, pre-diagnostic graphs are automatically displayed (Figure 4B). When compared to a single 0.95 percentile adjustment on the positive peak for the CD8 channel of our dataset (Figure 4C), bi-percentile adjustment on 0.95 (positive peak) and 0.70 (light green line in the queue of the negative events) clearly harmonizes the distribution of events between batches, but with a risk of distortion. Indeed, although linear, the bi-percentile method can cause "weak" events to shift to zero/negative values differently from batch to batch. To avoid this effect and maintain the positive or zero intensities inherent in mass cytometry, we recommend using the "percentile_lohi_pos" method.

The individual percentiles chosen for each channel in the "Tune Params" menu are retained in the interface and will be applied when clicking on the "preview" button from the "Process menu" (see below). This represents a major improvement, as compared to the original package in which users had to realize multiple successive runs, one for each given percentile value, with specification of which channels had to be adjusted for each specific run. Those channel-specific percentiles can also be specified manually in the "Panel" table stored in the project directory. Doing so, the "Panel" table has to be saved, closed and reloaded manually using the "reload" button (DWF steps 9 and 10) before going to the "Process" menu (Figure 5).

"Process" menu

In the "Process" function (Figure 5, DWF steps 19-23), clicking on the "preview" button generates PDF files summarizing both raw and a preview of the adjustments to be realized on all control tubes. A "prefix" and a "suffix" can be conveniently added to the output file names (DWF steps 20 and 21). After having reviewed the normalization of the control samples, the normalization of all the FCS files is launched by clicking the "process" button (DWF step 22). After calculation, the "review" button (DWF step 23) creates a PDF report showing the channel histograms before and after normalization (as a mirror of the "Tune parameters" menu) for each acquired FCS file (control and experimental samples) of each batch, allowing a rapid visual control of the adjusted FCS files. Although this visual check may be sufficient for some markers (such as those with a bimodal distribution), the best way to assess the relevance of the adjustment is to check the adjusted FCS files (all or a representative sample) with adequate and biologically relevant bi-parametric plots to understand the changes in intensity distribution and identify any aberrations. This control will be carried out using standard software, which can also be used for rapid manual gating, which we recommend, particularly for cytokine channels that do not have easily identifiable positive peaks. A summary of the final adjustments is exposed in Figure 6 (A and B), as well as downstream compensations (C) for Dataset 1.

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

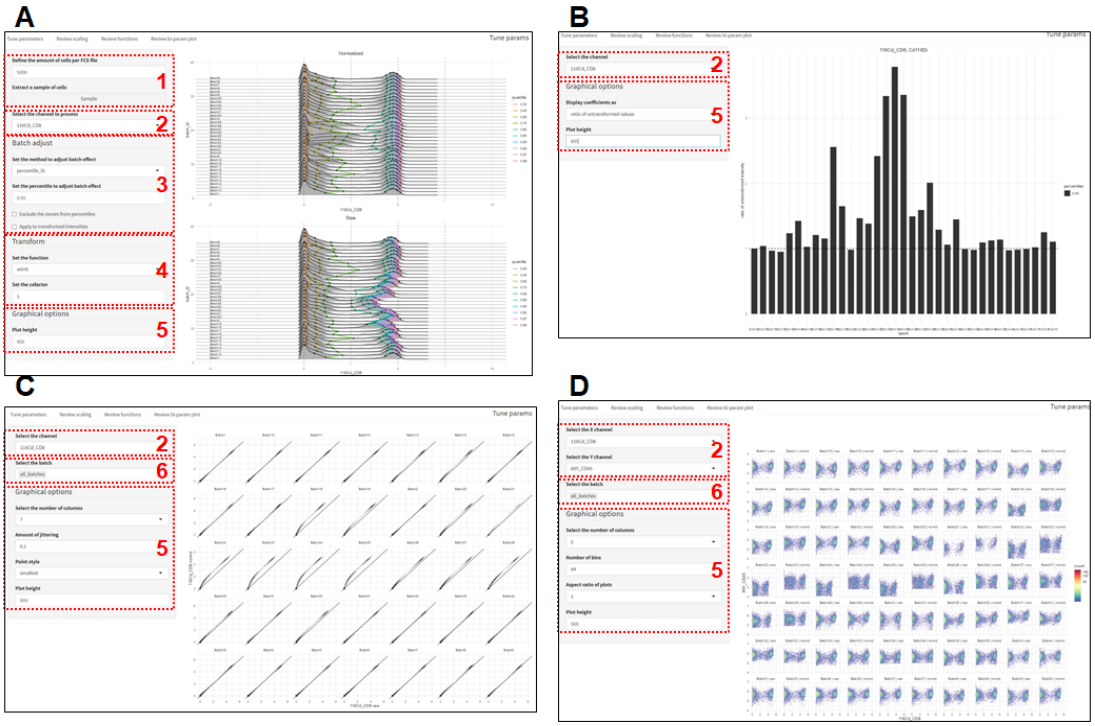


Figure 3: Detailed "Tune parameters" tabs. A: "Tune parameters" main tab. B: "Review scaling" tab. C: "Review functions" tab. D: "Review bi-parameters plot" tab. Tools for parameters tuning include a sampling function (A1), a channels selection function (A2, B2, C2, D2), batch adjustment functions (A3), transform functions (A4), graphical options (A5, B5, C5, D5) and a batch selection function (C6, D6).

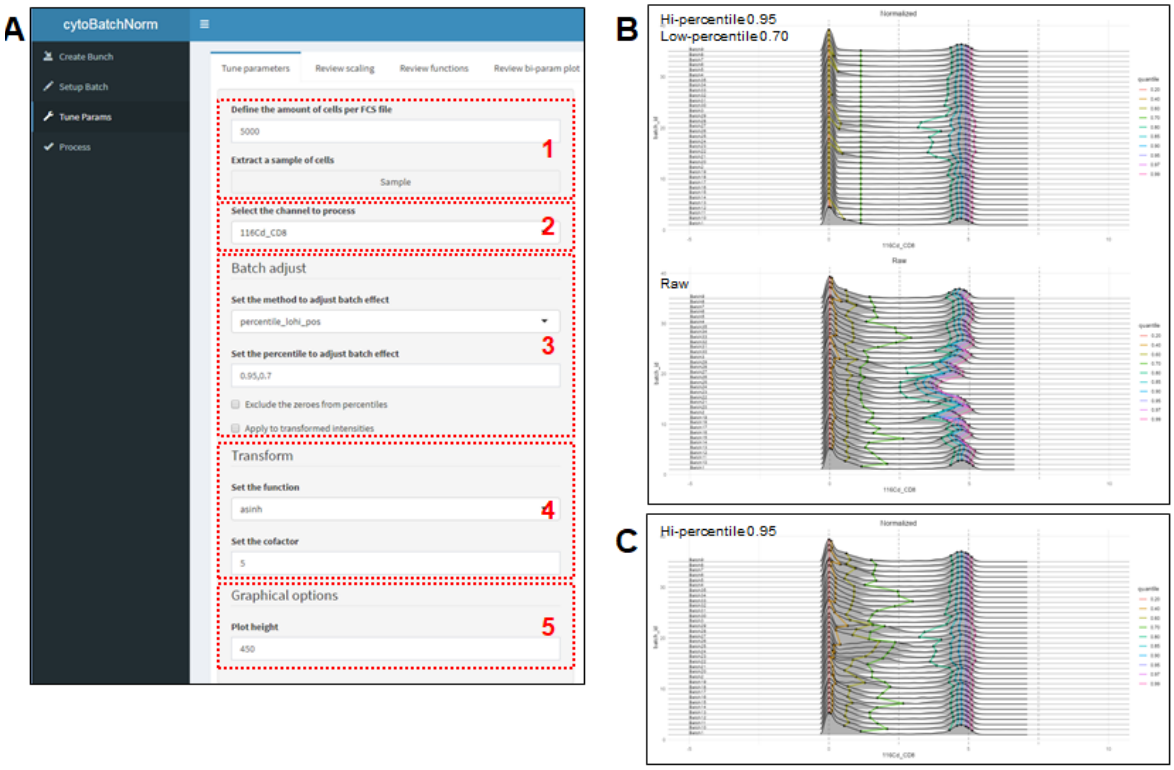


Figure 4: Bi-percentile adjustment. The Batch adjust section in the Tune parameters menu (A-3) allows to choose a "percentile_lohi" option to define two values of percentiles, separated by a comma, which will serve for adjustment of the selected channel (A-2). With the bi-percentile scaling, the distribution of positive and negative peaks is clearly homogenized (B) when compared with the single "percentile_hi" method (C). Tools for parameters tuning include sampling functions (1), a channels selection function (2), batch adjustment functions (3), transform functions (4), graphical options (5).

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

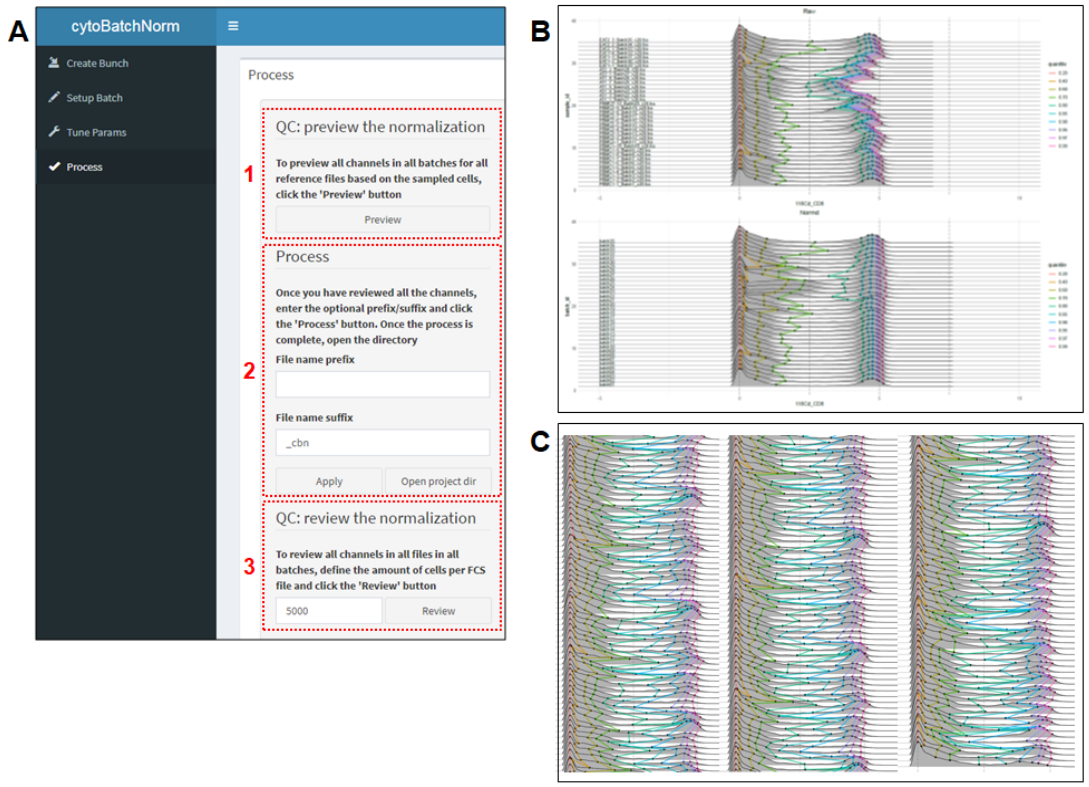
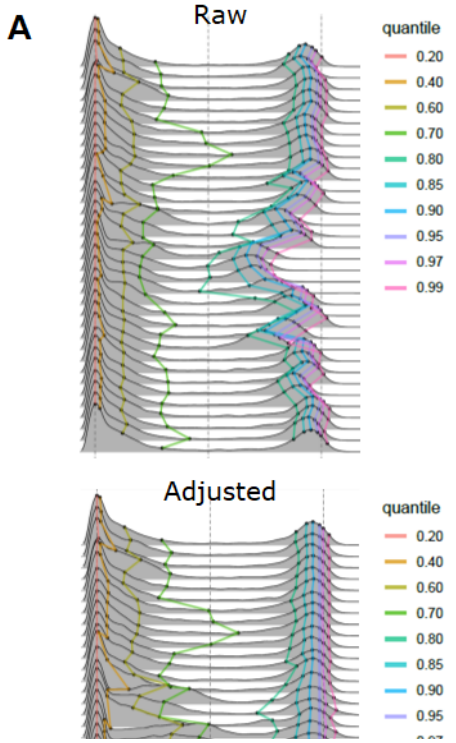


Figure 5: Detailed "process" tab. The Process menu allows to preview (A-1) the adjustment on all channels and all batches for reference files on PDF (B) consequently to parameters tuned previously. When satisfying, processing of FCS files for batch effects adjustments is launched by clicking on the "apply" button (A-2). The 'Review' button (A-3) can be used to view the result of the adjustment of all the channels on all the files and all the batches in a PDF file(C, samples of the output PDF file for CD8 channel).



B

Target	Metal	CBN percentile	Target	Metal	CBN percentile
CD45	Y89Di	0.95	FoxP3	Gd157Di	0.95
CD282	Rh103Di	0.99	CD33	Gd158Di	0.95
Pan-CEA	In113Di	0.99	CD11c	Tb159Di	0.99
Connexin43	In115Di	0.99	CD28	Gd160Di	0.95
CD8	Cd116Di	0.95	CD26	Dy161Di	0.95
Perforin	La139Di	0.85	CD66b	Dy162Di	0.99
CD192	Ce140Di	0.95	CD172a	Dy163Di	0.99
CD196	Pr141Di	0.9	CD161	Dy164Di	0.95
CD19	Nd142Di	0.95	CD45-RO	Ho165Di	0.95
CD127	Nd143Di	0.95	CD24	Er166Di	0.95
CD38	Nd144Di	0.95	CD197	Er167Di	0.95
CD163	Nd145Di	0.99	CD206	Er168Di	0.99
IgD	Nd146Di	0.95	CD25	Tm169Di	0.95
CD20	Sm147Di	0.95	CD45-RA	Er170Di	0.8
CD34	Nd148Di	0.95	GranzymeB	Yb171Di	0.95
CD194	Sm149Di	0.95	CX3CR1	Yb172Di	0.95
CD64	Nd150Di	0.99	HLA-DR	Yb173Di	0.95
CD123	Eu151Di	0.95	CD4	Yb174Di	0.95
TCRgd	Sm152Di	0.95	CD14	Lu175Di	0.8
CD185	Eu153Di	0.95	CD56	Yb176Di	0.85
CD3	Sm154Di	0.95	CD1c	Pt195Di	0.85
CD27	Gd155Di	0.95	CADM1	Pt196Di	0.95
CD183	Gd156Di	0.95	CD16	Bi209Di	0.95

C

Figure 6: Mass cytometry data pre-processing. A: Illustration of batch adjustment on CD8 marker plots. B: Summary of percentiles used for batch adjustment of each channel of the panel. C: Compensation matrix calculated using CATALYST.

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

Batch effects correction

Three levels of batch effects were identified and corrected in Dataset 1 as illustrated in Figure 7, termed as "single", "barcode" and "time" batch effects. "Single" effects reflect specific normalization to a single batch for a given channel without any link to the batch or tissue-series it belongs to, as illustrated in Figure 7A (pink arrows). "Barcode" effects refer to batch homogenous specificity in correction levels for a given channel (Figure 7A, red brackets). The "time" effect was seen specifically in one experiment on PBMC (20 samples in one barcode), which displayed high variation of certain channels intensity during the 10 batch/runs of this barcode acquisition, as illustrated for CD45-$_{89}$Y (Figure 7B). This probably reflects tolerable instability of the plasma torch, argon pressure, or TOF detector that were not corrected by normalization beads. These variations that were not corrected by the normalization on Four-Elements EQbeads (Standard BioTools Inc, San Francisco, CA, US) illustrates a very powerful implementation of batch correction for minimizing an instrument-related batch effect over time. Lastly, comparison of raw and batch-corrected control files on dimension reduction maps recapitulates the improvement of data homogeneity after processing to batch correction. Figure 7C illustrates accurate correction of batch effects-related CD8+ cells aggregation and expression level seen on a t-SNE map in the control sample from batch 23 (AT tissue, "Raw" vs "Normalized") when compared to the reference batch 1. Multi-Dimensional Scaling (Figure 7D) using the CytoMDS R package[28] also illustrates reduced dispersion of AT samples (red circle) after correction (pink dots) compared to raw data (blue dots).

Finally, cytoBatchNorm can be used on all operating systems, including Windows.

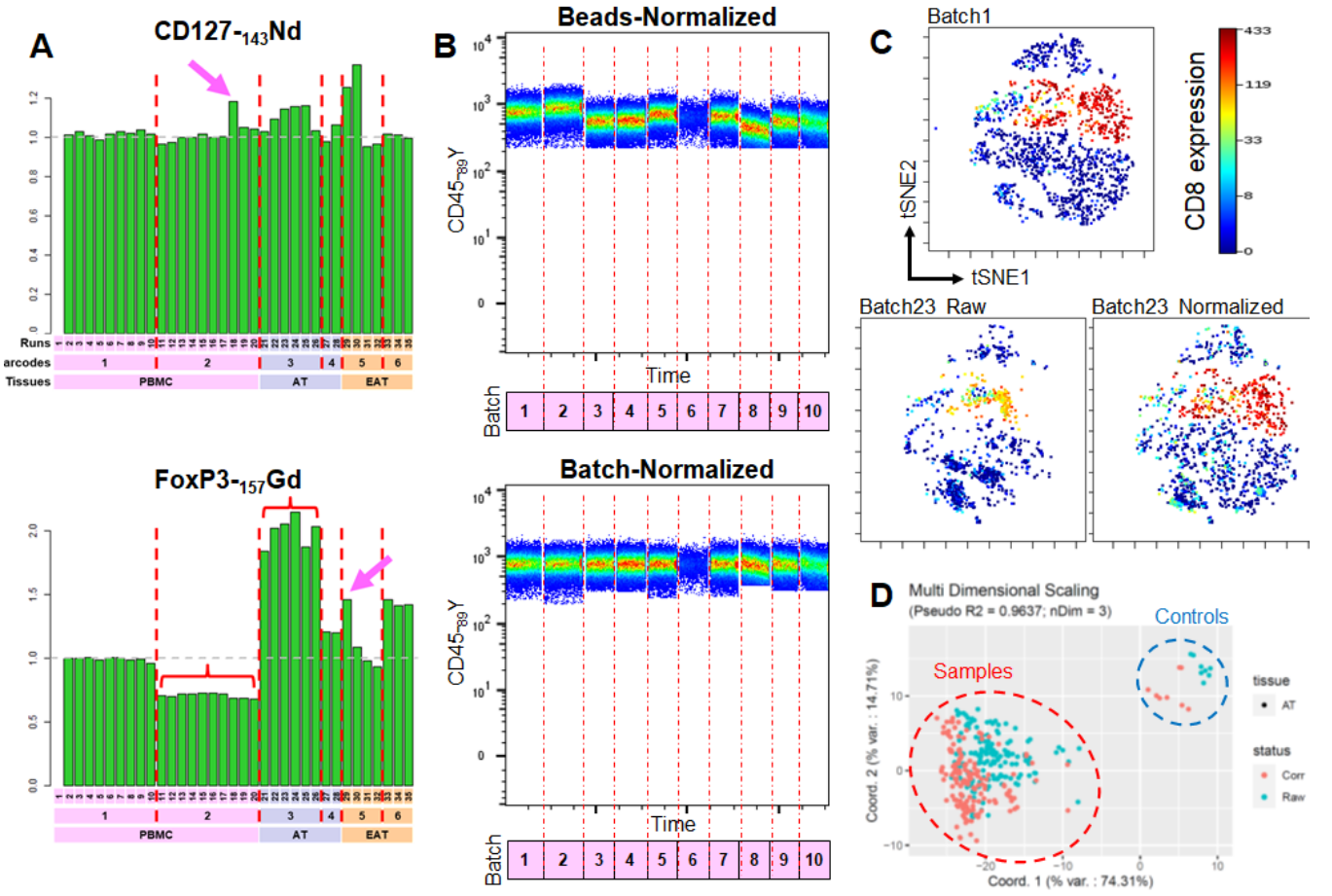


Figure 7: Different levels of corrected batch effects. A: scaling factors for CD127 and FoxP3 channels. Pink arrows point to "tube" specific batch effects; red brackets point to "barcode" specific batch effects. B: dot plots of PBMC batches 11 to 20 showing basal CD45 levels as a function of time; after bead normalization, variations were not corrected (top plot), illustrating 'time' specific batch effects; CytoBatchNorm corrects them (bottom plot). C: CD8 expression on t-SNE dimension reduction map of lineage markers from Batch1 (PMBC) and Batch 23 (AT tissue) control samples illustrates the improvement in data consistency following batch effect correction with CytoBatchNorm. D: Multi-Dimensional Scaling using the CytoMDS R package shows reduction of AT samples dispersion after batch effect correction (pink dots) compared to raw data (blue dots).

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

Benchmarking

Benchmarking of CytoBatchNorm was performed on Datasets 1 and 2 by two different scientists. Dataset 1 served for comparison between CytoBatchNorm and CytofBatchAdjust[19], while Dataset 2 served for comparison between CytoBatchNorm and CytoNorm[18].

As shown in Figure 8A, on first use, the set-up time with CytoBatchNorm is approximately one hour for both datasets, reflecting the time spent examining all channels and configuring all percentiles in the dataset. On the second use (to re-adjust the parameters), the handling time falls sharply from an hour to 10-20 minutes, as experimenters get used to both the interface and their dataset. In comparison, on dataset 2, this time remains twice as long as that of CytoNorm, whose methodology does not allow for adjustment of channel corrections. For Dataset 1, correction by CytofBatchAdjust required numerous successive runs to determine the best percentile for each channel (as well as a systemic visual examination of the 700 FCS files making up Dataset 1), which took 16 hours. In the end, using CytoBatchNorm was faster or equivalent, in particular because it allows a pre-visualization of the normalization. In addition, when the same panel is used in a new experiment, it is possible to re-use the previous settings, which reduces the time spent looking for the right parameters.

Key populations variance comparison between Dataset 2 control files batch-corrected with either CytoBatchNorm or CytoNorm demonstrates CytoBatchNorm is at least as accurate as (CD8, PD1) or more accurate (CD4, HLA-DR) than CytoNorm (Figure 8B). On the opposite, CytoBatchNorm was slightly less accurate in reducing median variance of key markers (Figure 8B), except for PD1.

Strikingly, CytoNorm introduced artifactual deformation of some populations in some experimental samples, as illustrated in Figure 8C for CD8 and HLA-DR in sample D (red boxes) from the third batch of Dataset 2, but not in the batch specific control sample (Control 3). This illustrates 1) the absolute need for cautious control of potential bias introduced by specific algorithms miscomputation during data processing, 2) the inaccuracy of quantile-based normalization for cytometry data.

Conclusion

Computer assistance in the treatment and analysis of omics data is ineluctable and has to be performed properly and wisely. Refining algorithms and free-access R packages to this aim will greatly enhance the still recent implementation of computational cytometry as well as the downstream results accuracy. One first, basic but essential step in cytometry data analysis is their standardization, including batch effects correction. We present the CytoBatchNorm R package which

A

Datasets	Package	Handling time		Processing time
		First run	Second run	
Dataset 1	CytoBatchNorm	1h	10 min	27 min
	CytofBatchAdjust	16h	16h	120 min
Dataset 2	CytoBatchNorm	1h	20 min	15 min
	CytoNorm	10 min	10 min	20min

B

Key populations frequency variance (Dataset 2)				
	CD8	CD4	PD1	HLA-DR
CytoBatchNorm	11.94	5.17	28.64	12.60
CytoNorm	11.95	5.43	28.64	17.84
Key markers median variance (Dataset 2)				
	CD8	CD4	PD1	HLA-DR
CytoBatchNorm	16.88	7.71	79.04	15.23
CytoNorm	12.06	3.90	80.10	5.88

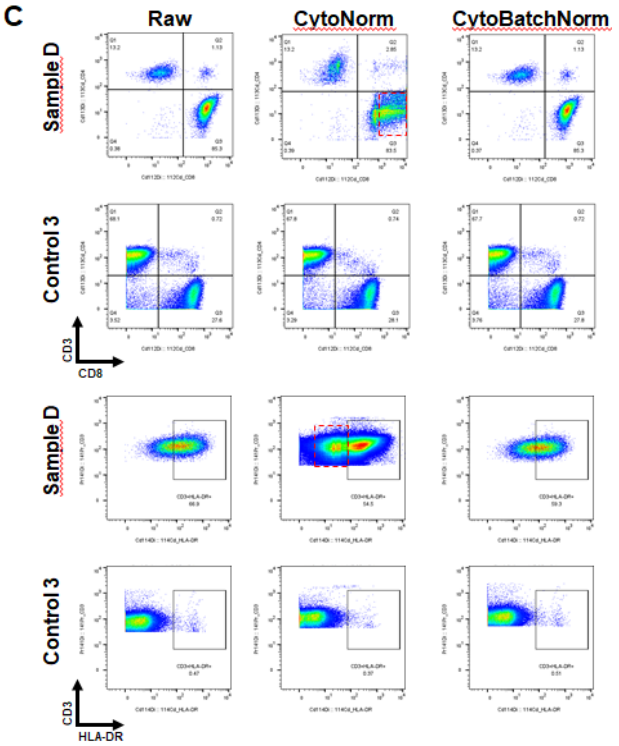


Figure 8: Benchmarking of CytoBatchNorm versus CytoNorm and CytofBatchAdjust. A: comparison of handling and processing time. B: Variance of key populations frequency and key markers median. C: Illustration of artifacts introduced in experimental samples by CytoNorm (red boxes).

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

is the most user-friendly package available for batch effects correction, with a live assessment of correction accuracy, and which out-performs existing packages in terms of both tuning possibilities and efficiency. CytoBatchNorm will help the cytometry community to adequately scale their data amongst batches, allowing reliable reduction of variability and improvement of subsequent dimension reduction and clustering in user's analysis pipeline.

Acknowledgements

We thank Philippe Hauchamps (PhD student) from Computational Biology and Bioinformatics department, Duve Institute, Catholic University of Louvain, Belgium, for his active improvement of the CytoMDS R package used for figure 7D edition.

Data availability statement: as data is for illustrative purposes only and has not yet been published for scientific study, available or submitted to a repository, it has not yet been made available nor submitted to FlowRepository.

Funding statement: This study was supported by grants from the Fédération Française de Cardiologie, the Fondation Leducq convention 16CVD01 "Defining and targeting epigenetic pathways in monocytes and macrophages that contribute to cardiovascular disease", the European Genomic Institute for Diabetes (EGID, ANR-10-LABX-0046) and the Agence Nationale de la Recherche (TOMIS leukocytes: ANR-CE14-0003-01).

Conflict of interest disclosure: None

Authors Contribution:

	S.G.	N.A.	A-S.C.	E.W.	L.P.	S.N.	A.H.	M.A.	D.M.	B.S.	D.D.	O.M-C.
Conceptualization	X		X									X
Data Curation	X	X										X
Formal Analysis	X	X										X
Funding Acquisition									X	X	X	
Investigation	X	X		X	X	X		X				X
Methodology	X		X									X
Project Administration	X											X
Resources	X	X					X	X	X	X	X	X
Software	X	X	X									X
Supervision	X											X
Validation	X	X	X									X
Visualization	X	X										X
Writing – Original Draft Preparation	X	X										X
Writing – Review & Editing	X	X	X				X	X				X

Ethics approval statement: The clinical cohorts were constituted as part of the POMI-AF study (NCT#03376165), approved by the institutional ethics committee (Comité de Protection des Personnes Ile de France V).

Patient consent statement: Written informed consent was obtained from all patients before inclusion.

Permission to reproduce material from other sources: None

Clinical trial registration: None

References

1. Rybakowska P, Alarcón-Riquelme ME, Marañón C. Key steps and methods in the experimental design and data analysis of highly multi-parametric flow and mass cytometry. Comput Struct Biotechnol J 18 (2020): 874–886.
2. Marsh-Wakefield FM, Mitchell AJ, Norton SE, et al. Making the most of high-dimensional cytometry data. Immunol Cell Biol 99 (2021): 680–696.
3. Rybakowska P, Van Gassen S, Quintelier K, et al. Data processing workflow for large-scale immune monitoring studies by mass cytometry. Comput Struct Biotechnol J 19 (2021): 3160–3175.
4. Wagar LE. Live Cell Barcoding for Efficient Analysis of Small Samples by Mass Cytometry. Mass Cytom Methods Protoc 10 (2019): 125–135.
5. Zunder ER, Finck R, Behbehani GK, et al. Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. Nat Protoc 10 (2015): 316–333.
6. Leipold MD, Obermoser G, Fenwick C, et al. Comparison of CyTOF assays across sites: Results of a six-center pilot study. J Immunol Methods 453 (2018): 37–43.
7. Lähnemann D, Köster J, Szczurek E, et al. Eleven grand challenges in single-cell data science. Genome Biol 21(2020): 1–35.
8. Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 11 (2010): 733–739.
9. Rybakowska P, Alarcón-Riquelme ME, Marañón C. Key steps and methods in the experimental design and data analysis of highly multi-parametric flow and mass cytometry. Comput Struct Biotechnol J 18 (2020): 874–886.
10. Čuklina J, Pedrioli PGA, Aebersold R. Review of Batch Effects Prevention, Diagnostics, and Correction Approaches. Mass Spectrom Data Anal Proteomics (2020): 373–387.
11. Goh WWB, Wang W, Wong L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. Trends Biotechnol 35 (2017): 498–507.
12. Kalina T, Flores-Montero J, van der Velden VHJ, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia 26 (2012):1986–2010.
13. Jamin C, Le Lann L, Alvarez-Errico D, et al. Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. Autoimmun Rev 15 (2016): 1038–1045.
14. Kleinsteuber K, Corleis B, Rashidi N, et al. Standardization and quality control for high-dimensional mass cytometry studies of human samples. Cytom Part J Int Soc Anal Cytol 89 (2016): 903–913.
15. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 16 (2019): 1289–1296.
16. Hahne F, Khodabakhshi AH, Bashashati A, et al. Per-channel basis normalization methods for flow cytometry data. Cytometry A 77A (2010): 121–131.
17. Finak G, Jiang W, Krouse K, et al. High-throughput flow cytometry data normalization for clinical trials. Cytometry A 85 (2014): 277–286.
18. Gassen SV, Gaudilliere B, Angst MS, et al. CytoNorm: A Normalization Algorithm for Cytometry Data. Cytometry A 97 (2020): 268–278.
19. Schuyler RP, Jackson C, Garcia-Perez JE, et al. Minimizing Batch Effects in Mass Cytometry Data. Front Immunol 10 (2019): 2367.
20. Shaham U, Stanton KP, Zhao J, et al. Removal of batch effects using distribution-matching residual networks. Bioinformatics 33 (2017): 2539–2546.
21. Lakkis J, Wang D, Zhang Y, et al. A joint deep learning model enables simultaneous batch effect correction, denoising and clustering in single-cell transcriptomics. Genome Res (2021): 271874.120.
22. Trussart M, Teh CE, Tan T, et al. Removing unwanted variation with CytofRUV to integrate multiple CyTOF datasets. eLife 9 (2020): e59630.
23. Ogishi M, Yang R, Gruber C, et al. Multibatch Cytometry Data Integration for Optimal Immunophenotyping. J Immunol Baltim Md 206 (2021): 206–213.
24. Ninni S, Dombrowicz D, Kuznetsova T, et al. Hematopoietic Somatic Mosaicism Is Associated With an Increased Risk of Postoperative Atrial Fibrillation. J Am Coll Cardiol 81 (2021): 1263–1278.
25. Arnaud M, Chiffelle J, Genolet R, et al. Sensitive identification of neoantigens and cognate TCRs in human solid tumors. Nat Biotechnol 40 (2022): 656–660.
26. Chevrier S, Crowell HL, Zanotelli VRT, et al. Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. Cell Syst 6 (2018): 612-620.
27. Chen H, Lau MC, Wong MT, et al. Cytofkit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline. PLoS Comput Biol 12 (2016): e1005112
28. Hauchamps P, Gatto L. CytoMDS: Low Dimensions projection of cytometry samples. R package version (2024).

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.

Granjeaud S, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107095

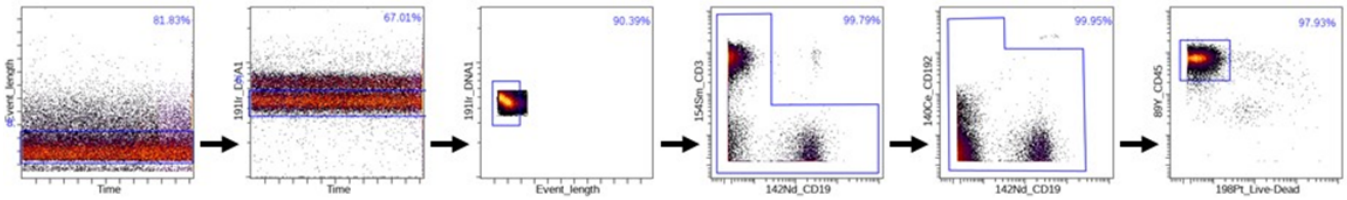


Figure S1: Pretreatment of reference FCS files. Manual gating of CD45+ live cells after exclusion of doublets, non-biological events and beads residues.

Citation: Samuel Granjeaud, Naoill Abdellaoui, Anne-Sophie Chrétien, Eloise Woitrain, Laurent Pineau, Sandro Ninni, Alexandre Harari, Marion Arnaud, David Montaigne, Bart Staels, David Dombrowicz, Olivier Molendi-Coste. CytoBatchNorm: An R package with Graphical Interface for Batch Effects Correction of Cytometry Data. Journal of Bioinformatics and Systems Biology. 7 (2024): 237-251.