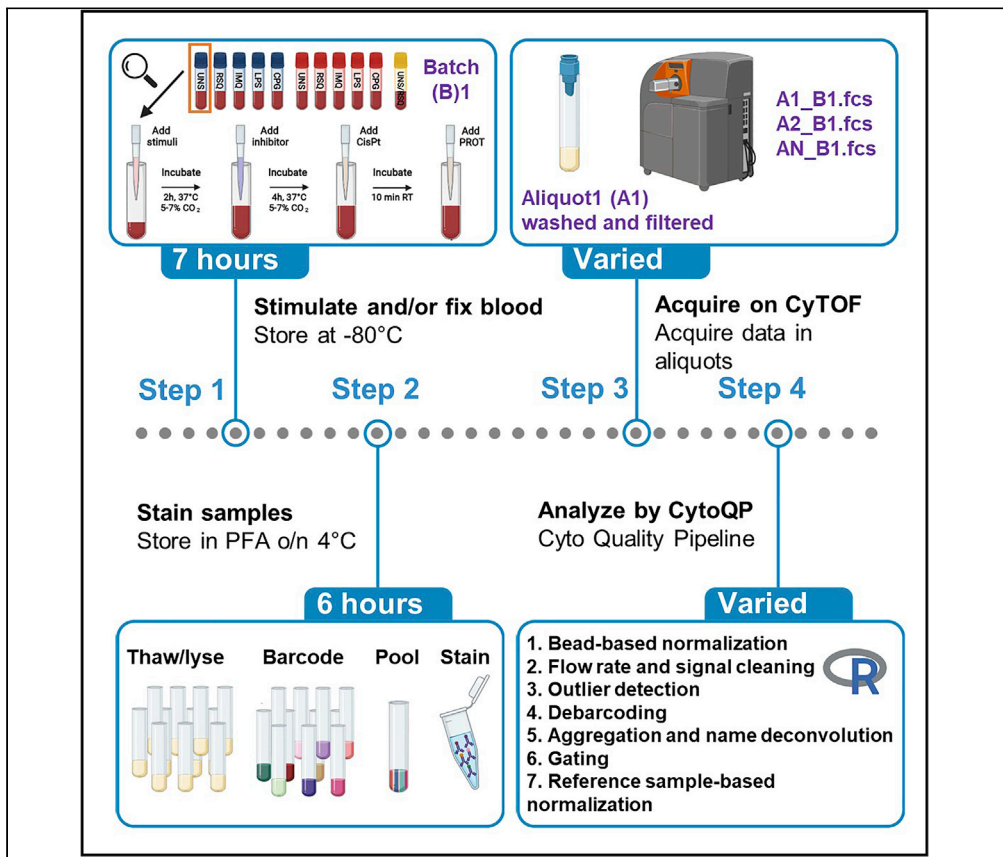# STAR Protocols

Protocol

Protocol for large scale whole blood immune monitoring by mass cytometry and Cyto Quality Pipeline

Paulina Rybakowska, Sofie Van Gassen, Jordi Martorell Marugán, Katrien Quintelier, Yvan Saeys, Marta E. Alarcón-Riquelme, Concepción Marañón

paulina.rybakowska@genyo.es (P.R.)
concepcion.maranon@genyo.es (C.M.)

Highlights

Steps to prepare and preserve samples for retrospective functional mass cytometry studies

Combine experimental design with data preprocessing and quality control pipeline CytoQP

Use sample barcoding to limit intra-batch variations

Use a reference sample to achieve comparability for multibatch and multicenter studies

Mass cytometry (MC) is a powerful large-scale immune monitoring technology. To maximize MC data quality, we present a protocol for whole blood analysis together with an R package, Cyto Quality Pipeline (CytoQP), which minimizes the experimental artifacts and batch effects to ensure data reproducibility. We describe the steps to stimulate, fix, and freeze blood samples before acquisition to make them suitable for retrospective studies. We then detail the use of barcoding and reference samples to facilitate multicenter and multi-batch experiments.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

STAR Protocols

Protocol

Protocol for large scale whole blood immune monitoring by mass cytometry and Cyto Quality Pipeline

Paulina Rybakowska,[1,8,*] Sofie Van Gassen,[2,3] Jordi Martorell Marugán,[1,4,5] Katrien Quintelier,[2,3,6] Yvan Saeys,[2,3] Marta E. Alarcón-Riquelme,[1,7] and Concepción Marañón[1,9,*]

[1]GENYO, Centre for Genomics and Oncological Research Pfizer/University of Granada/ Andalusian Regional Government, PTS, 18016 Granada, Spain

[2]Department of Applied Mathematics, Computer Sciences and Statistics, Ghent University, 9000 Gent, Belgium

[3]Data Mining and Modeling for Biomedicine, VIB Center for Inflammation Research, 9000 Gent, Belgium

[4]Department of Statistics and Operational Research, University of Granada, 18071 Granada, Spain

[5]Data Science for Health Research Unit, Fondazione Bruno Kessler, 38123 Trento, Italy

[6]Department of Pulmonary Diseases, Erasmus MC, 3015 Rotterdam, the Netherlands

[7]Institute for Environmental Medicine, Karolinska Institute, 17177 Stockholm, Sweden

[8]Technical contact

[9]Lead contact

*Correspondence: paulina.rybakowska@genyo.es (P.R.), concepcion.maranon@genyo.es (C.M.)
https://doi.org/10.1016/j.xpro.2022.101697

SUMMARY

Mass cytometry (MC) is a powerful large-scale immune monitoring technology. To maximize MC data quality, we present a protocol for whole blood analysis together with an R package, Cyto Quality Pipeline (CytoQP), which minimizes the experimental artifacts and batch effects to ensure data reproducibility. We describe the steps to stimulate, fix, and freeze blood samples before acquisition to make them suitable for retrospective studies. We then detail the use of barcoding and reference samples to facilitate multicenter and multi-batch experiments.

For complete details on the use and execution of this protocol, please refer to Rybakowska et al. (2021a) and (2021b).

BEFORE YOU BEGIN

This protocol is focused on blood functional studies where cell cytokine responses to stimulation are measured. You will follow an experimental procedure where we stimulated the blood samples with toll-like receptor agonists, fixed them using Proteomic stabilizer (PROT), and stored frozen. As an example, we performed 1 experiment in which blood from 2 donors was stimulated using 5 different conditions, aliquoted into 3 tubes each (simulating a multi-batch experiment), and stored at −80°C. Additionally, blood was drawn from donor 2 to create a reference sample, which contains the information of the technical variability introduced during sample preparation, staining and acquisition. This sample was aliquoted and stored. Refer to Figures 1 and 2 for experimental details. Subsequently, we will use the generated data to illustrate the preprocessing and quality control with an R package called Cyto Quality Pipeline (CytoQP). Figure 3 presents the computational workflow overview. These protocols are easily scalable to large scale multicenter, multi-batch studies, as the samples can be collected in multiple laboratories, stained and acquired in various acquisition batches.

Institutional permissions

Two human healthy donors were enrolled under a protocol approved by the Ethical Committee of Centro Granada (CEI-Granada) according to the Helsinki declaration of 1975, as revised in 2013.
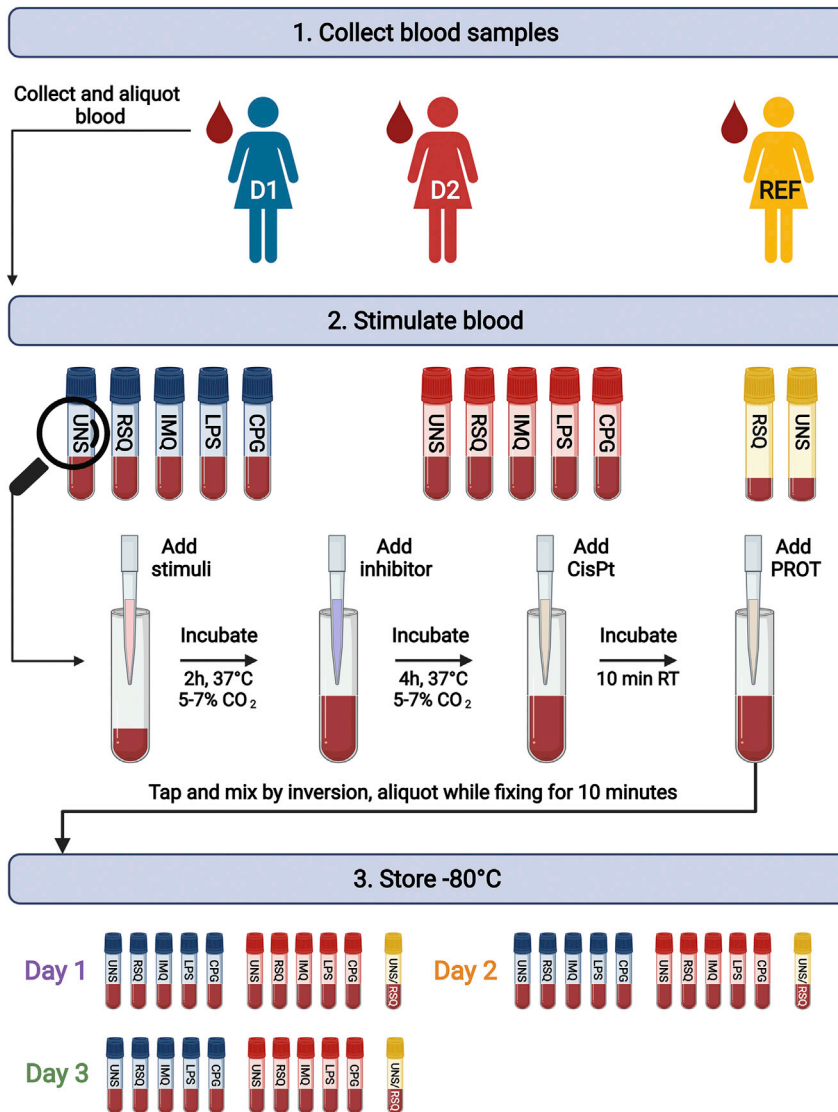
Figure 1. Sample collection and stimulation workflow

(1) Blood from two donors (D1 – blue and D2 – red) was drawn and aliquoted. Additional volume of blood was collected from D2 to create a reference sample (REF - yellow). (2) Donors' blood was stimulated with 4 different stimulation agents, or left unstimulated. Half volume of REF sample was stimulated with resiquimod (RSQ) and the other half was left unstimulated (UNS). The magnifier shows a representative example of the stimulation protocol for one sample, although all the samples were processed in parallel. Stimuli was added to the blood in the proportion 1:1, and left for 2 h at 37°C with 5%–7% CO_2. Next, protein transport inhibitor was added and samples were incubated for another 4 h. After the incubation time was completed, the viability staining was performed using CisPt for 10 min (at this point, 2 conditions for REF sample were pooled and stained together) and PROT was added for sample fixation. Samples were mixed well and aliquoted into three batches (Day1, Day2, Day3). (3) After completing 10 min fixation time samples were stored at −80°C until staining.

All donors signed an informed consent according to the ethical protocol of the Andalusian Biobank and the PRECISESADS project. The Granada node of the Andalusian Health System Biobank collected whole blood samples.

We would like to remind that adequate permissions should be obtained when working with human biological samples.
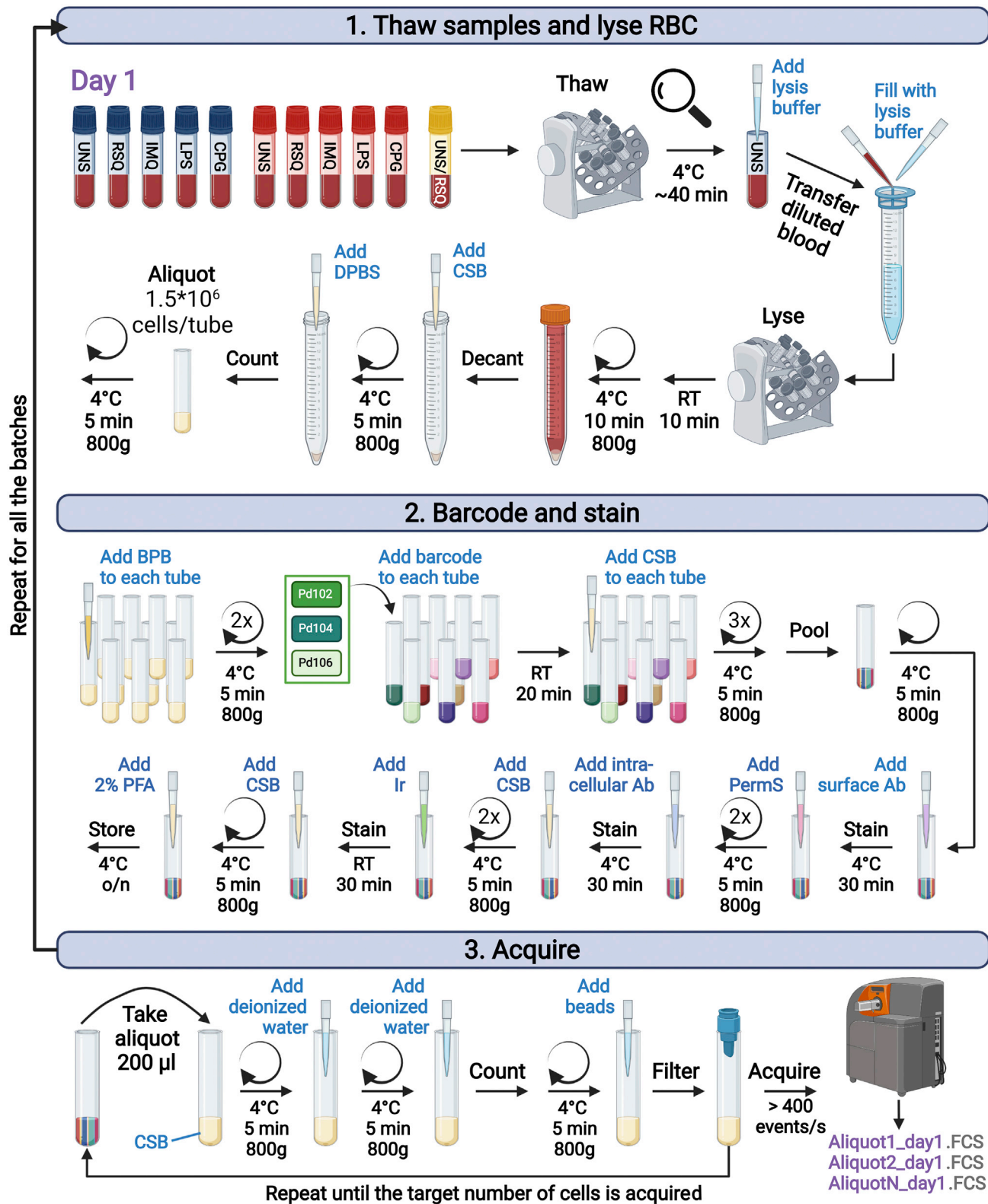
1. Thaw samples and lyse RBC

Day 1

UNS | RSQ | IMQ | LPS | CPG   UNS | RSQ | IMQ | LPS | CPG   UNS/RSQ

Thaw

4°C ~40 min

Add lysis buffer

UNS

Transfer diluted blood

Fill with lysis buffer

Lyse

RT 10 min

4°C 10 min 800g

Decant

Add CSB

4°C 5 min 800g

Add DPBS

Count

Aliquot 1.5*10^6 cells/tube

4°C 5 min 800g

Repeat for all the batches

2. Barcode and stain

Add BPB to each tube

2x

4°C 5 min 800g

Pd102
Pd104
Pd106

Add barcode to each tube

RT 20 min

Add CSB to each tube

3x

4°C 5 min 800g

Pool

4°C 5 min 800g

4°C 5 min 800g

Add surface Ab

Stain

4°C 30 min

Add PermS

2x

4°C 5 min 800g

Add intra-cellular Ab

Stain

4°C 30 min

Add CSB

2x

4°C 5 min 800g

Add Ir

Stain

RT 30 min

Add CSB

4°C 5 min 800g

Add 2% PFA

Store

4°C o/n

3. Acquire

Take aliquot 200 μl

CSB

Add deionized water

4°C 5 min 800g

Add deionized water

4°C 5 min 800g

Count

Add beads

4°C 5 min 800g

Filter

Acquire > 400 events/s

Aliquot1_day1.FCS
Aliquot2_day1.FCS
AliquotN_day1.FCS

Repeat until the target number of cells is acquired

Figure 2. Experimental protocol workflow

(1) On three acquisition days, samples for each batch were thawed (here represented by Day1) in the cold room on hula mixer, for approximately 40 min. After thawing, lysis buffer was added to each tube (for the purpose of figure clarity magnifier shows one tube, though all samples were processed in parallel). Next, samples were filtered through a 100 μm strainer and transferred to a 15 mL tube containing lysis buffer. Tubes were washed with additional 1 mL of lysis buffer and 15 mL tubes were filled with lysis buffer. Samples were placed on hula mix and lysed for 10 min at RT. Next, they were centrifuged and the supernatant was decanted by inversion. Cells were washed with CSB followed by one wash with DPBS. Then, they were counted, aliquoted at 1.5*10^6 cells/mL and centrifuged. (2) Cells were barcoded using palladium-based barcoding (colored tubes represent different barcodes) and washed 3 times with CSB. After last wash, cells were pooled together into a single tube. Next, surface staining was performed, followed by 2 washing steps with Perm-S and intracellular staining. Afterward, cells were washed and stained with Ir, washed again and kept in PFA o/n at 4°C. (3) Samples were acquired on CyTOF, generating *Aliquot.FCS* files. Shortly, every aliquot of ∼200 μL was washed once with CSB and two times with MilliQ water. Before the last wash, 10 μL of cell suspension was used for cell counting. Cells were resuspended in bead solution at 0.8*10^6 cells/mL and acquired at an event rate below 400 events/s. This step was repeated until desired number of cells was acquired and the protocol was repeated on the following days for *Day2* and *Day3*.

Determine the total amount of samples and acquisition batches

🕐 Timing: 2 h

1. Estimate the number of samples that will be analyzed.
2. Calculate the total number of samples that will be included in the acquisition batch.

   Note: Acquisition using CyTOF is long, approximately 10^6 cells/hour, thus calculate the amount of cells according to the CyTOF availability and experimental setting. Remember, that one barcoded sample can be split up to 3 acquisition days using protocol presented below. It is worth to keep in mind that the less batches are run the less variability is introduced to the data.

3. Calculate the total amount of batches to be run, we advise to prepare at least one extra batch so the experiment can be repeated if needed.
4. Calculate the volume of reagents that are required to prepare a single batch of staining cocktails and reagents. Always prepare at least 10% excess to account for pipetting errors, in case of antibody cocktail preparation 20% excess is recommended for mastermix preparation and cocktail aliquoting.
5. Assign samples to the acquisition batches.

   Note: Proper batch design is critical to ensure even distribution of the biological groups. If multiple patient groups are studied, make sure that each batch contains a representation of each group, including healthy controls and recruitment centers. We recommend to use a software like Optimal Sample Assignment Tool (OSAT) to design batch distribution (Yan et al., 2012).

Prepare stimulation agents

🕐 Timing: 30 min per reagent

This step will depend on the stimuli used and the number of total experiments to be performed. In this example we have used 5 different stimulation conditions for 2 healthy donors (D1 and D2) in 3 consecutive experiments +1 backup experiment. Therefore, the stimulation agents were prepared for 8 samples each (see Figure 1 for more details). The stimulation agents were the toll-like receptor agonists LPS, ODN2006 (CpG), R848 (RSQ), R837 (IMQ), and medium alone (UNS). These were prepared at 2× concentration in RPMI, and the final blood stimulation was performed at 1× concentration in a final volume of 1 mL (500 μL of stimuli + 500 μL of blood) in sterile conditions as in (Rybakowska et al., 2021a).

6. Prepare stimuli.
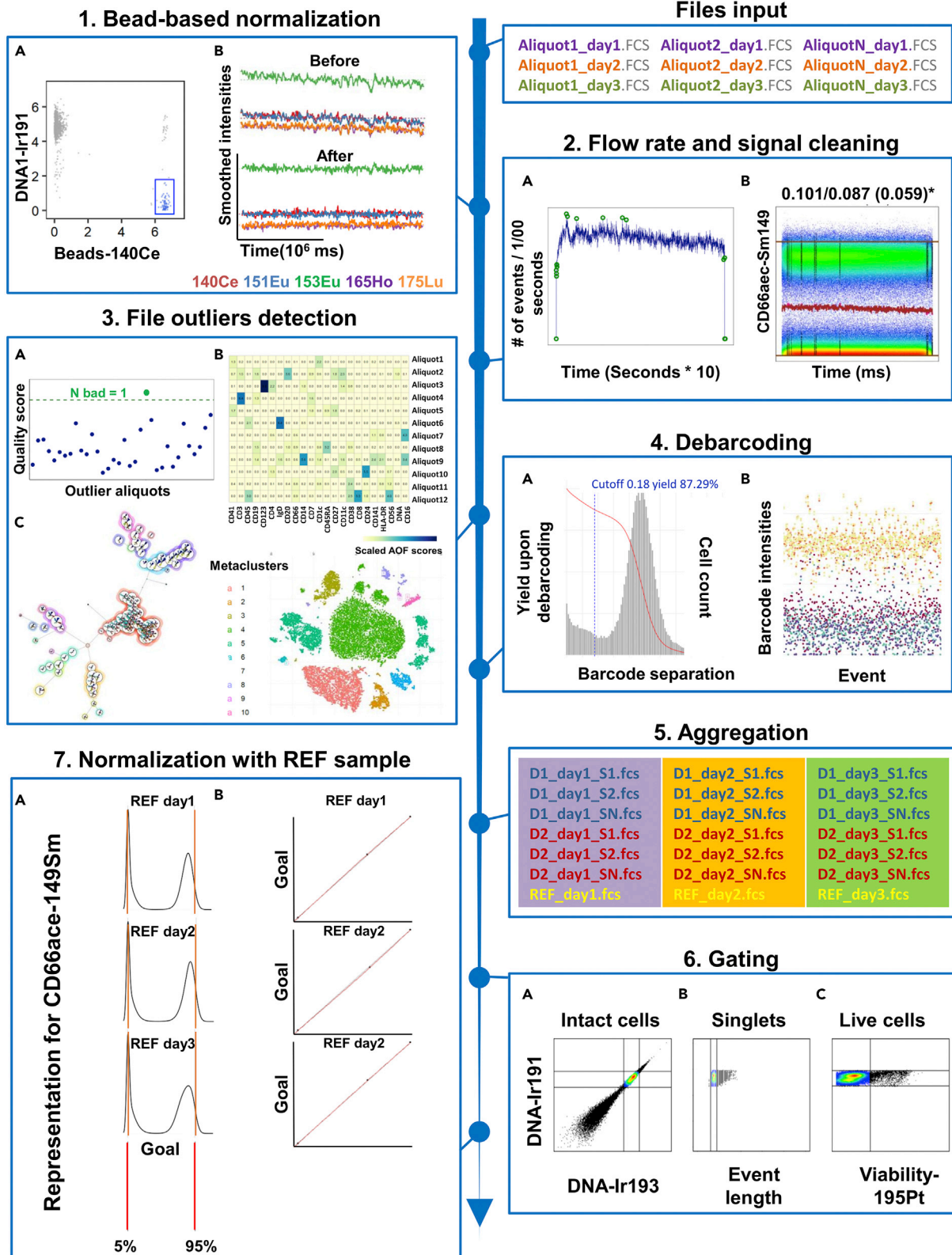   a. If the stimulation agents were frozen thaw them on ice.

1. Bead-based normalization

A

DNA1-Ir191

Beads-140Ce

B

Before

After

Smoothed intensities

Time(10^6 ms)

140Ce   151Eu   153Eu   165Ho   175Lu

Files input

Aliquot1_day1.FCS   Aliquot2_day1.FCS   AliquotN_day1.FCS
Aliquot1_day2.FCS   Aliquot2_day2.FCS   AliquotN_day2.FCS
Aliquot1_day3.FCS   Aliquot2_day3.FCS   AliquotN_day3.FCS

2. Flow rate and signal cleaning

A

# of events / 1/00 seconds

Time (Seconds * 10)

B

0.101/0.087 (0.059)*

CD66aec-Sm149

Time (ms)

3. File outliers detection

A

Quality score

N bad = 1

Outlier aliquots

B

Aliquot1
Aliquot2
Aliquot3
Aliquot4
Aliquot5
Aliquot6
Aliquot7
Aliquot8
Aliquot9
Aliquot10
Aliquot11
Aliquot12

CD41 CD3 CD5 CD19 CD123 CD4 IgD CD20 CD66 CD14 CD7 CD45RA CD27 CD11c CD1c CD38 CD8 CD24 CD141 HLA-DR CD56 DNA CD16

Scaled AOF scores

C

Metaclusters

1 2 3 4 5 6 7 8 9 10

4. Debarcoding

A

Yield upon debarcoding

Cutoff 0.18 yield 87.29%

Cell count

Barcode separation

B

Barcode intensities

Event

5. Aggregation

D1_day1_S1.fcs   D1_day2_S1.fcs   D1_day3_S1.fcs
D1_day1_S2.fcs   D1_day2_S2.fcs   D1_day3_S2.fcs
D1_day1_SN.fcs   D1_day2_SN.fcs   D1_day3_SN.fcs
D2_day1_S1.fcs   D2_day2_S1.fcs   D2_day3_S1.fcs
D2_day1_S2.fcs   D2_day2_S2.fcs   D2_day3_S2.fcs
D2_day1_SN.fcs   D2_day2_SN.fcs   D2_day3_SN.fcs
REF_day1.fcs   REF_day2.fcs   REF_day3.fcs

7. Normalization with REF sample

A

Representation for CD66ace-149Sm

REF day1
REF day2
REF day3

Goal

5%   95%

B

REF day1
Goal

REF day2
Goal

REF day2
Goal

6. Gating

A

Intact cells

DNA-Ir191

DNA-Ir193

B

Singlets

Event length

C

Live cells

Viability-195Pt

Figure 3. Preprocessing and data cleaning workflow

fcs files generated on 3 acquisition days are used as input. Files from 1 to N are shown to represent multiple aliquots. Each step generates the figures described below. (1) Bead-based normalization. (A) Dot plots generated for one representative aliquot with arcsine transformed DNA channel (191Ir) versus all bead channels (here represented by Ce140). The beads in the blue region are used for normalization. (B) Smoothed bead intensities in their positive channels (different beads represented by colors) along the time of acquisition for one representative aliquot. "Before" (top) and "After" (bottom) normalization. (2) Flow rate and signal cleaning. (A) Flow rate of one representative aliquot, is shown. The green circles indicate the outliers that were detected and removed from further analysis. (B) Signal cleaning for the same aliquot. The dot plot shows the arcsine transformed intensity for cleaned CD66ace-149Sm. The black lines indicate the indices removed from the analysis. The numbers above the plots show the mean change before and after cleaning and the maximum mean change for this aliquot, the asterisk marks this channel as the worst channel among the rest. (3) File outlier detection. (A) Outlier aliquot detection. On the x-axis the file name marked as outlier (here shown as *Outlier aliquots*) and on the y-axis Quality score. The green dotted line represents the threshold for outlier definition. Dots represent scores for each aliquot, the green dot shows a file score detected above the quality threshold. (B) Heatmap representation of Scaled AOF scores for each surface marker across aliquots collected on Day 1. (C) FlowSOM and t-SNE representation of clustering results colored by 10 metaclusters. (4) Debarcoding. (A) Distribution of the events by separation distance (gray histogram) and the cell yield (red line) are shown as a function of the separation threshold. The vertical, blue dotted line represents the cutoff for the separation threshold, the cell yield upon debarcoding is also shown. (B) All cell events assigned to one barcode, with each cell event represented as a dot. The intensities for each palladium isotype are presented with different colors. The data for one representative aliquot is shown. (5) Aggregation. Files are aggregated per batch (represented by different colors) and the donor-specific names are deconvoluted. (6) Gating. In each dot plot the selected cells are shown in color and discarded cells are shown in black. (A) intact, nucleated cells versus debris, (B) singlets versus doublets, and (C) viable cells versus dead cells. (7) Normalization with reference sample. (A) Distribution of the CD66-a-c-e marker in the reference sample on each staining day. The 5% and 95% percentiles are represented by the orange vertical lines and the computed mean for 3 experimental days is shown below in red lines (goal distribution). (B) The linear function that transforms the result is represented by a red line and the identity function represents the gray line (here behind the red line, as almost no batch effect is observed). Black dots shows 5% and 95% percentiles and limits.

b. Short spin to collect all the stimulation agent.
c. Dilute each stimulation agent to its 2× concentration in RPMI in sterile conditions, using 15 mL falcon tubes.

Note: The size of tube depends on the total amount of stimulation medium to be prepared.

d. Vortex for 10 s.
e. Split in aliquots of 550 μL into sterile, labeled 1.5 mL microcentrifuge tubes.
f. Seal the tubes with parafilm to avoid evaporation.
g. Place aliquots at −20°C (short-term) or −80°C (long-term storage).

See Table 1 for calculation details.

△ CRITICAL: It is important to determine the appropriate stimulus concentration before performing the final experiment.

△ CRITICAL: The protocol needs to be performed in sterile conditions under the cell culture hood.

The stimuli should be prepared using cold RPMI to ensure their stability.

⏸ **Pause point:** These stimulation agents can be stored at −20°C for short-term (1 month) or −80°C for long-term (at least 1 year).

Note: The number of the total stimulation tubes depends on the total number of individuals expected in the study. For multicenter studies, the batch of stimulation agents should be prepared in a central laboratory and then shipped on dry ice to the rest of collaborators. This will prevent introducing unwanted technical variation.

Note: If other than presented here stimuli are used it is advised to check their stability upon freezing, by testing in parallel fresh and frozen reagent. Note that LPS activity can vary from LOT to LOT. Therefore, it is important to ensure its similar behavior upon LOT change.

Table 1. An example of stimulus calculation and preparation for 4 experiments

Stimulation agent	Total number of aliquots	Final concentration	Working concentration (2×)	Stock concentration	Stimulation medium (2×)/ tube	[a]Total stimulation medium (2×) volume + 10%	Stock volume needed	RPMI needed
RSQ	8	1.25 µg/mL	2.5 µg/mL	1 mg/mL	500 µL	4400 µL	11 µL	4389 µL
IMQ	8	10 µg/mL	20 µg/mL	1 mg/mL	500 µL	4400 µL	88 µL	4312 µL
CpG	8	2 µM	4 µM	500 mM	500 µL	4400 µL	25.2 µL	4364.8.6 µL
LPS	8	0.05 µg/mL	1 µg/mL	1 mg/mL	500 µL	4400 µL	4.4 µL	4395.6 µL
UNS (RPMI)	8	N/A	N/A	N/A	500 µL	4400 µL	N/A	4400 µL

[a]Stimulation medium refers to RPMI with the stimulation agent. It is prepared and frozen at 2× concentration because the blood is stimulated as following: 500 µL of blood + 500 µL of stimulation medium. To obtain 8 aliquots of each stimulation medium we prepare 4400 µL ((500 * 8) + 10%) in total.

Conjugate and titrate antibody and confirm clone fixation resistance

© Timing: 4–6 h for antibody conjugation, 4–5 h for antibody titration. Conjugation time depends on the protocol used: X8 for lanthanide isotopes (shorter) vs MCP-9 for cadmium isotopes (longer). Staining time is shorter for surface and longer with intracellular marker titration.

Although many markers are currently available in the conjugated form, self-conjugation provides flexibility for panel optimization and can be necessary when antibody clones compatible with fixation method are needed or probes are not available in the catalog. The antibody conjugation protocol was recently described (Thrash et al., 2020) and it will not be discussed here, although some useful tips are added. Before and after performing the conjugation, measure the antibody concentration using nanodrop or perform a BCA protein assay. The conjugation protocol can be performed in a large scale, which means that more than 100 µg can be conjugated at once. For more details see (Hartmann et al., 2019).

Proper antibody titration and determination of antibody compatibility with fixation is a critical step when performing mass cytometry (MC) experiments with stabilized blood samples. An example of good and bad antigen preservation is shown in Figure 4.

An example of how to titrate antibodies is shown in (Thrash et al., 2020), and won't be developed in this paper. Titration should be carried out under the same conditions as for the experimental samples, which means that samples should be fixed in the same way after the same treatment.

⚠ CRITICAL: Titrate the antibody using the rule $5*10^7$ cells/mL (in this example $1.5*10^6$ cells/ 30 µL), in this way the experiments can be easily scalable when the barcoding protocol is used (see below).

Confirm that the targeted antigens are resistant to the fixation: the pattern of the expression is preserved, as shown in Figure 4 and in (Rybakowska et al., 2021a).

Prepare batches of frozen antibody cocktails

© Timing: 1 h

These steps will depend on the number of total barcoded tubes that will be acquired and the staining volume that will be used during antibody incubation. In this example we used 10 samples from 2 donors in 3 consecutive experiments. In each experiment one reference sample was added, giving a total of 11 samples per batch. For each individual we stained $1.5*10^6$ cells/sample (using the rule $5*10^7$/mL). Taken together, 3 barcoded tubes were acquired on 3 different days, (Figures 1 and 2).
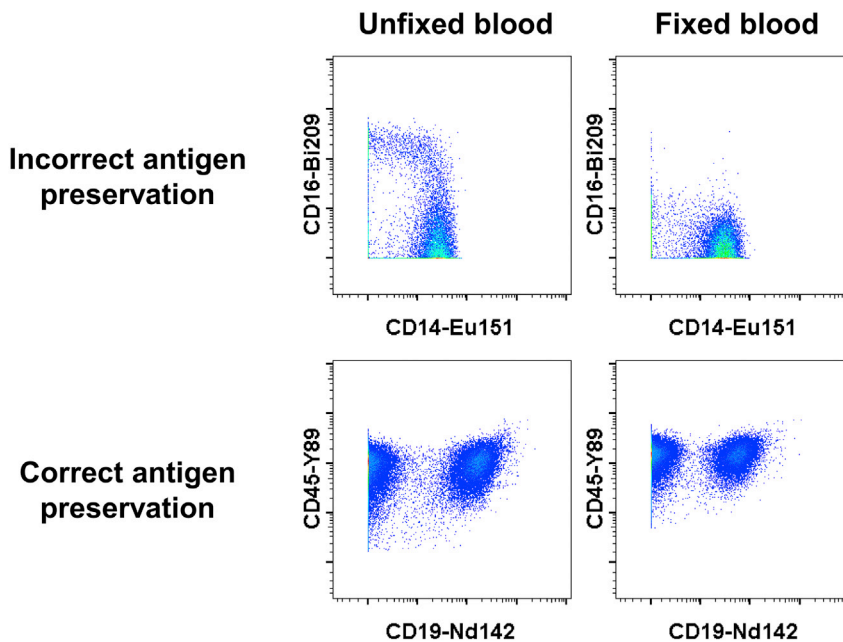
Figure 4. Antigen preservation after fixation

An example of blood samples stained with CD16, CD14, CD19 and CD45 markers before (fresh, unfixed, left) and after fixation with PROT (right). Top panel shows incorrect antigen preservation after fixation in monocytes (parental gate: CD45[+]CD66ace[−]CD3[−]CD19[−]CD7[−]CD11c[high] HLA-DR[+]). The loss of CD16 staining is observed for fixed blood compared to unfixed blood. Bottom panel shows correct antigen preservation, the CD19 staining is preserved after fixation and shows a similar pattern to the fresh sample (PBMC gate: CD45[+]CD66ace[−]).

We recommend to prepare at least one extra aliquot of antibody cocktail in case needed, thus in this example antibody cocktail is calculated for a total of 4 experiments. Two separate cocktails for cell surface and intracellular staining were prepared, as in (Rybakowska et al., 2021a).

7. Calculate the amount of antibody needed.

   Note: See Table 2 for the antibody calculation and Tables 3 and 4 for the surface and cytokine cocktail preparation, respectively.

   Note: In this case we prepare ~20% more of the mastermix than needed. This is to account for pipetting error and also to avoid touching the bottom of the tube while aliquoting. The same rule is applied to the antibody aliquot volumes.

8. Prepare the antibody mastermix.

Table 2. Calculation of total volume of antibody needed for staining

	1 sample	11 samples (1 batch)	48 samples (4 batches)	Final volume to stain (+ ~20%)
Number of cells / sample	1.5×10^6 cells	16.5×10^6 cells	66×10^6 cells	
Staining volume	30 μL	330 μL	1320 μL	1600 μL

The number of cells and staining volume for one single sample is shown and multiplied by 11 samples included in one staining batch. Next, a factor 4 is used to prepare antibody cocktail for 4 batches (48 samples). To account for pipetting error and avoid touching the bottom of the tube ~20% is added to the final staining volume. This calculation allows to easily scale up experiments to a multibatch (4 in this case) setting, preserving cell concentration used for antibody titration. It shows the staining volume needed to stain one batch but also allows to calculate total staining volume required to prepare antibody cocktails aliquots.

Table 3. Mastermix preparation for surface staining

Antigen	Metal	Clone	ᵃDilution factor	Total [ul] amount for 330 μL / 11 sample	Total [ul] amount for 1,600 μL / 48 samples
CD3	115In	UCHT1	400	0.825	4
CD45	141Pr	HI30	400	0.825	4
CD19	142Nd	HIB19	400	0.825	4
CD123	143Nd	6H6	400	0.825	4
CD4	145Nd	RPA-T4	400	0.825	4
IgD	146Nd	IA6-2	200	1.65	8
CD20	147Sm	2H7	400	0.825	4
CD66a/c/e	149Sm	ASL-32	400	0.825	4
CD14	151Eu	M5E2	200	1.65	8
CD7	153Eu	CD7-6B7	100	3.3	16
CD1c	154Sm	L161	800	0.4125	2
CD45RA	155Gd	HI100	400	0.825	4
CD27	158Gd	L128	400	0.825	4
CD11c	159Tb	Bu15	400	0.825	4
CD38	167Er	HIT2	400	0.825	4
CD8	168Er	SK1	800	0.4125	2
CD24	169Tm	ML5	100	3.3	16
CD141	173Yb	1A4	400	0.825	4
HLA-DR	174Yb	L243	100	3.3	16
CD56	176Yb	N901	100	3.3	16
CD16	209Bi	3G8	100	3.3	16
CD41	89Y	HIP8	400	0.825	4
			Total volume	31.35	152
			Volume to aliquot + ~20%	38	n/a

Antibody volumes needed to prepare surface stating cocktail for one batch (11 samples) and multibatch experiments, 4 batches (48 samples), taking into consideration a dilution factor for each antibody and final staining volume obtained in Table 2. In this case 35 μL of antibody mixture (with 20% excess) is aliquoted in one single tube. In total 4 tubes are prepared.
ᵃFinal antibody concentration depends on the cell type, antibody lot and conditions. Therefore, users should titrate the antibody according to their own setting.

    a. Centrifuge the antibody stock at 10,000 g for 10 min, 4°C, to sediment aggregates.
    b. Combine the antibodies into the mastermix according to the amount needed (see Table 3), using 1.5 mL loBind tube.
9. Aliquot and freeze mastermix.
    a. Centrifuge the mastermix at 15,000 g for 15 min, 4°C.
    b. Prepare and label Safe-lock 0.5 mL tubes.
    c. Split the mastermix into aliquots in individual labeled tubes.

Note: To account for pipetting error, aliquot at least 20% more than needed.

    d. Seal the tubes with parafilm to avoid evaporation.
Place aliquots at −80°C.

Note: the antibody cocktail can be stored for at least 1 year.

⚠ CRITICAL: To aliquot the mastermix use only the supernatant, without touching the bottom of the tube, where antibody aggregates are deposited after centrifugation. To do this, prepare ~20% more than needed.

Table 4. Mastermix preparation for cytokine staining

Antigen	Metal	Clone	Dilution factor	Total [ul] amount for 330 μL / 11 samples	Total [ul] amount for 1,600 μL / 48 samples
IL-2	144Nd	MQ1-17H1	400	0.825	4
IL-17	148Nd	BL168	100	3.3	16
MIP1B	150Nd	D21-1351	400	0.825	4
TNFa	152Sm	Mab11	400	0.825	4
IL-6	156Gd	MQ2-13A5	200	1.65	8
IFN-a	160Gd	LT27:295	2,000	0.165	0.8
IL-23p19	161Dy	23dcdp	400	0.825	4
TGF-b	163Dy	TW4-6H10	100	3.3	16
IL-10	166Er	JES3-9D7	100	3.3	16
IL-12p70	170Er	REA123	800	0.4125	2
Granzyme B	171Yb	GB11	100	3.3	16
MCP-1	175Lu	5D3-F7	200	1.65	8
			Total volume	20.38	98.8
			Volume to aliquot + ~20%	24	

Antibody volumes needed to prepare intracellular stating cocktail for one batch (11 samples) and multibatch experiments, 4 batches (48 samples), considering a dilution factor for each antibody and final staining volume (1,600 μL) obtained in Table 2. In this case 23 μL of antibody mixture (with 20% excess) is aliquoted in one single tube. In total 4 tubes are prepared.

Note: In our experiments we stain 1.5×10^6 cells/sample, but this number could be scaled considering the frequency of target cells. Perform staining using the rule 5×10^7 cells/mL.

This protocol is adapted from (Schulz et al., 2019).

Stimulate and stabilize blood samples

⏱ Timing: 7 h

10. Collect blood in EDTA-K2 tubes.

    Alternatives: Other anticoagulants like heparin could be used, however cytokine expression and blood stability for longer time storage should be verified.

11. Prepare stimulation agents.
    a. Thaw the stimulation agents and let them warm-up to RT.
    b. Short-spin the tubes to make sure that all the liquid is in the tube.
    c. Transfer 500 μL of stimulation agent to the sterile 5 mL FACS tubes with cap. For the control tube add 500 μL of RPMI.
12. Start stimulation.
    a. Mix blood by inversion every time before transferring the blood.
    b. Aliquot 500 μL of blood to the tubes containing the stimuli, gently pipet once up and down to ensure that all cells from the tip are in the stimulation tube.
    c. Vortex gently.
    d. Incubate for 2 h at 37°C, 5%–7% CO_2.

    Note: Make sure that the tube is loosely closed, so the air exchange can take place.

13. Block cytokine secretion.
    a. Add 2 μL of 500× eBioscience™ Protein Transport Inhibitor Cocktail to each tube to obtain final concentration of 1×, vortex gently.

Table 5. Reagent calculation for whole blood stimulation and reference sample preparation

	Donor		Reference sample			
	1 tube	4 experiments +10%	1 tube		4 experiments + 10%	
Stimulation agent 2×/plain medium	500 µL	2,200 µL	250 µL (RSQ)	250 µL (RPMI)	1100 µL (RSQ)	1100 µL (RPMI)
Blood	500 µL	2,200 µL	250 µL	250 µL	1100 µL	1100 µL
Protein transport inhibitor	2 µL	8.8 µL	1 µL	1 µL	4.4 µL	4.4 µL
CisPt (5 mM)	1 µL	4.4 µL	1 µL		4.4 µL	
Proteomic Stabilizer	1.4 mL	6.2 mL	1.4 µL		6.2 mL	

Reagent volumes needed to stimulate, fix and freeze 1 donor sample in 4 experiments (columns for Donor) and 1 reference sample in 4 experiments (columns for reference sample). Note that half of the reference sample is stimulated with RSQ and the other half is left unstimulated. These samples are pooled after stimulation; thus the calculation is done separately up to the protein transport inhibitor.

Note: eBioscience™ Protein Transport Inhibitor Cocktail is a mixture of the monensin and brefeldin A however it can be personalized depending on the experimental setting.

Alternatives: eBioscience™ Monensin Solution (1000×) and eBioscience™ Brefeldin A Solution (1000×) could be purchased and mixed in-house.

    b. Incubate the tubes for another 4 h at 37°C, 5%–7% CO_2.

14. Stabilize the blood.

    a. Perform live/dead staining adding 1 µL of cisplatin (CisPt, 5 mM) to a final concentration of 5 µM.

    b. Incubate for 10 min at RT.

    c. Add 1.4 volume of Proteomic Stabilizer (1.4 mL). Tightly close the tubes and mix by inversion.

    d. Incubate for 10 min at RT.

    e. Freeze 5 mL FACS tubes directly at −80°C.

See Table 5 column 1 and 2 for calculation details.

△ CRITICAL: If stimulation is performed, blood should be processed within 3 h from collection. Only for surface staining, blood can be processed within 24 h. Avoid vigorous pipetting of the blood to prevent granulocytes' death.

It is important to place the stabilized blood tubes at −80°C as soon as the 10 min incubation time is over. Make sure that samples are tightly closed.

▐▐ Pause point: Samples can be left at −80°C for at least 1 year as shown in (Rybakowska et al., 2021a).

Note: If a lot of samples will be stimulated at the same time we recommend to prepare stimulation tubes before the experiment is started. It is advisable to label the tubes using printer and cryolabels.

Additionally, each label should be protected with transparent tape to avoid label loss during freezing.

For multicenter studies, samples should be frozen at −80°C and shipped to the central laboratory on dry ice if necessary.

Alternatives: Blood could be also fixed using Phosphoflow Fix and Lyse (BD Biosciences), for more detail please refer to (Rybakowska et al., 2021a).

Prepare reference sample, stimulate and preserve

🕐 Timing: 6 h

Multibatch experiments suffer from day-to-day variation. Hence, measures allowing estimation and correction of batch effects are needed. Inclusion of reference sample is one way to deal with this, as it allows to measure the variability and correct it (Rybakowska et al., 2020). In this chapter we show how to prepare a reference sample and further in the analysis part how to use it to correct for batch effect.

Prepare one aliquot of the reference sample per acquisition batch plus some backup aliquots. In this example 3 barcoded batches were acquired on 3 different days, and thus 4 reference sample aliquots were prepared. Start from a 10% excess of blood volume. The reference sample should be preserved with the same protocol as the experimental samples. In the case of intracellular cytokine staining the reference sample should contain positive and negative cells for all studied cytokines. In our setting we used 1:1 mixture of unstimulated and RSQ-stimulated blood as reference sample. The RSQ reagent was prepared with the same concentration as shown in Table 1 with a final volume 1,100 μL (Table 5).

> ⚠ CRITICAL: If the mixture of unstimulated and stimulated samples is used as a reference, pool them after the stimulation and proceed with the protocol, see Table 5. This will ensure that all the cells are evenly distributed across the stored aliquots.

15. Prepare a stimulation tube. Only in the case of panels including markers induced by stimulation (e.g., cytokines, phosphoproteins or activation markers).
    a. Prepare two 5 mL FACS tubes containing 1.1 mL of either 2✕ RSQ or RPMI.

    Note: The size of the tube will depend on the final culture volume. After the incubation time is over, blood from stimulated and unstimulated tubes will be joined together, stained with CisPt and immediately fixed in one tube, so make sure that 1.4 volume of Proteomic Stabilizer can be added.

    b. Mix blood by inverting the blood tube up and down for a few times, to homogenize the cells to avoid sedimentation.
    c. Add 1.1 mL of blood to the stimulus, pipette up and down for few times.
    d. Vortex gently.
    e. Incubate for 2 h at 37°C and 5%–7% CO_2.

    Note: Make sure that the tube is loosely closed, so air exchange can occur.

16. Block cytokine secretion. Only in the case of panels including intracellular cytokines.
    a. Add 4.4 μL of 500✕ eBioscience™ Protein Transport Inhibitor Cocktail to each tube to reach a final concentration of 1✕, vortex gently.
    b. Incubate tubes for 4 h at 37°C and 5%–7% CO_2.
17. Stabilize the blood.
    a. Pool stimulated and unstimulated blood together in one single 15 mL tube, vortex gently.
    b. Perform live/dead staining adding 4.4 μL of CisPt to a final concentration of 5 μM.
    c. Incubate for 10 min at RT.
    d. Add 1.4 volumes (6.2 mL) of Proteomic Stabilizer.
    e. Close the tubes tightly and mix by inversion.
    f. Incubate for 10 min RT.

g. During the incubation time prepare aliquots of 2.4 mL of stabilized blood in 5 mL FACS tubes with cap and close them tightly.

Note: While aliquoting, mix well the blood/PROT, so the cell distribution is equal in each aliquot. Use the automated single or multichannel pipet to prepare aliquots.

h. Place 5 mL FACS tubes at −80°C.

See Table 5, column 4 and 5 for calculation details.

⚠ CRITICAL: It is important to place the stabilized blood tubes at −80°C as soon as the 10 min incubation time is over.

Note: If a big volume of blood needs to be stimulated and stabilized, we recommend using larger tubes or vented capped T flasks. Gentle mixing will prevent cell sedimentation and ensure even cell distribution. If it is necessary to split the sample in two or more tubes, pool all the simulated blood before aliquoting and mix them well, so all the reference tubes have the same cell composition.

For multicenter experiments it is critical to prepare reference samples in the central laboratory.

⏸ Pause point: Samples can be left at −80°C for at least 1 year as shown in (Rybakowska et al., 2021a).

Perform pilot experiment

⏱ Timing: may vary depending on the scale of pilot study

18. Using the final protocol and having reagents prepared, perform a pilot study to verify that all the steps work fine together before starting with precious patients' samples.

Note: If multiple centers are involved, include them in the pilot experiment to verify that the obtained results are comparable, as done in (Jamin et al., 2016).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies (dilutions in Table 3)		
Anti-Human CD45 (HI30)-141Pr—100 Tests	Standard BioTools	Cat#3141009B
Anti-Human CD19 (HIB19)-142Nd—100 Tests	Standard BioTools	Cat#3142001B
Anti-Human CD123/IL-3R (6H6)-143Nd—100 Tests	Standard BioTools	Cat#3143014B
Anti-Human CD4 (RPA-T4)-145Nd—100 Tests	Standard BioTools	Cat#3145001B
Anti-Human IgD (IA6-2)-146Nd—100 Tests	Standard BioTools	Cat#3146005B
Anti-Human CD20 (2H7)-147Sm—100 Tests	Standard BioTools	Cat#3147001B
Anti-Human CD66a/c/e (ASL-32)-149Sm—100 Tests	Standard BioTools	Cat#3149018B
Anti-Human CD14 (HCD14)-156Gd—100 Tests	Standard BioTools	Cat#3156019B
Anti-Human CD7 (CD7-6B7)-153Eu—100 Tests	Standard BioTools	Cat#3153014B
Anti-Human CD45RA (HI100)-155Gd—100 Tests	Standard BioTools	Cat#3155011B
Anti-Human CD27 (L128)-158Gd—100 Tests	Standard BioTools	Cat#3158010B
Anti-Human CD11c (Bu15)-159Tb—100 Tests	Standard BioTools	Cat#3159001B
Anti-Human CD38 (HIT2)-167Er—100 Tests	Standard BioTools	Cat#3167001B
Anti-Human CD8 (SK1)-168Er—100 Tests	Standard BioTools	Cat#3168002B
Anti-Human CD24 (ML5)-169Tm—100 Tests	Standard BioTools	Cat#3169004B

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Anti-Human CD141 (1A4)-173Yb—100 Tests	Standard BioTools	Cat#3173002B
Anti-Human HLA-DR (L243)-174Yb—100 Tests	Standard BioTools	Cat#3174001B
Anti-Human CD56 (N901)-176Yb—100 Tests	Standard BioTools	Cat#3176009B
Anti-Human CD16 (3G8)-209Bi—100 Tests	Standard BioTools	Cat#3209002B
Anti-Human CD41 (HIP8)-89Y—100 Tests	Standard BioTools	Cat#3089004B
Anti-Human IL-2 (MQ1-17H12)-144Nd—100 Tests	Standard BioTools	Cat#3144021B
Anti-Human IL-17A (BL168)-148Nd—100 Tests	Standard BioTools	Cat#3148008B
Anti-Human MIP1beta (D21-1351)-150Nd—100 Tests	Standard BioTools	Cat#3150004B
Anti-Human TNFa (Mab11)-152Sm—100 Tests	Standard BioTools	Cat#3152002B
Anti-Human IL-6 (MQ2-13AS)-156Gd—100 Tests	Standard BioTools	Cat#3156011B
Anti-Human IL-23 p19 (23dcdp)-161Dy—100 Tests	Standard BioTools	Cat#3161010B
Anti-Human TGFbeta (TW4-6H10)-163Dy—100 Tests	Standard BioTools	Cat#3163010B
Anti-Human IL-10 (JES3-9D7)-166Er—100 Tests	Standard BioTools	Cat#3166008B
Anti-Human Granzyme B (GB11)-171Yb—100 Tests	Standard BioTools	Cat#3171002B
IL-12 (p70) Antibody, anti-human, pure, REAfinity™	Miltenyi Biotec	Cat#130-108-051
IFN-α Antibody, anti-human, pure	Miltenyi Biotec	Cat#130-108-050
CD3 Monoclonal Antibody (UCHT1), eBioscience™	Thermo Fisher Scientific	Cat#14-0038-82
Purified Mouse Anti-Human MCP-1 Clone 5D3-F7 (RUO)	BD Biosciences	Cat#550416
Purified anti-human CD1c Antibody	BioLegend	Cat#331502
Biological samples		
Human peripheral blood (females, 20–30 years old)	Healthy controls	N/A
Chemicals, peptides, and recombinant proteins		
Cisplatin	Sigma Aldrich	Cat#479306-1G
Indium	Sigma Aldrich	Cat#203440-1G
Maxpar® Perm-S Buffer—250 mL	Standard BioTools	Cat#201066
Maxpar® Fix and Perm Buffer—100 mL	Standard BioTools	Cat#201067
Maxpar® Cell Staining Buffer—500 mL	Standard BioTools	Cat#201068
EQ Four Element Calibration Beads—100 mL	Standard BioTools	Cat#201078
Cell-ID™ 20-Plex Pd Barcoding Kit	Standard BioTools	Cat#201060
Cell-ID™ Intercalator-Rh—500 µM	Standard BioTools	Cat#201103A
R848 (Resiquimod)	InvivoGen	Cat#tlrl-r848-5
Imiquimod (R837)	InvivoGen	Cat#tlrl-imq
LPS-EK (LPS from E. coli K12)	InvivoGen	Cat# tlrl-peklps
ODN 2006 (ODN 7909)	InvivoGen	Cat#tlrl-2006-5
RPMI 1640 Medium, GlutaMAX™ Supplement. 500 mL	Thermo Fisher Scientific	Cat#61870-010
eBioscience™ Protein Transport Inhibitor Cocktail (500×)	Thermo Fisher Scientific	Cat#00-4980-93
Proteomic Stabilizer PROT1	Smart Tube Inc.	Cat#PROT1-1L
Antibody Stabilizer PBS	CANDOR Bioscience	Cat#131 125
Sodium azide, ReagentPlus®, ≥99.5%	Sigma-Aldrich	Cat#S2002-5G
Bond-Breaker™ TCEP Solution, Neutral pH	Thermo Fisher Scientific	Cat#77720
Pierce™ 16% Formaldehyde (w/v), Methanol-free	Thermo Scientific	Cat#28908
Thaw-Lyse buffer	Smart Tube Inc.	Cat#PROT1-1L
Heparin	Sigma-Aldrich	Cat#H3393-10KU
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	Cat# D8418-100ml
Dulbecco's Phosphate Buffered Saline (DPBS)	Fisher Scientific S.L	Cat#14190-094
Critical commercial assays		
Maxpar® X8 Antibody Labeling Kit, 154Sm—4 Rxn	Standard BioTools	Cat#201154A
Maxpar® X8 Antibody Labeling Kit, 160Gd—4 Rxn	Standard BioTools	Cat#201160A
Maxpar® X8 Antibody Labeling Kit, 170Er—4 Rxn	Standard BioTools	Cat#201170A
Maxpar® X8 Antibody Labeling Kit, 175Lu—4 Rxn	Standard BioTools	Cat#201175A
Deposited data		
FlowRepository: FR-FCM-Z3YR	(Rybakowska et al., 2021a, 2021b)	www.flowrepository.org

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
CyTOF Software Version 6.7.1014	Standard BioTools	https://www.fluidigm.com/software
R software	(R Core Team, 2013)	http://www.R-project.org/
RStudio	RStudio Team	https://rstudio.com/
CytoQP	This paper, https://doi.org/10.5281/zenodo.6947414	https://github.com/prybakowska/CytoQP
BiocParallel	(Morgan et al., 2022b)	https://bioconductor.org/packages/release/bioc/html/BiocParallel.html
WinRAR	N/A	https://www.win-rar.com/start.html?&L=12
Other		
Amicon Ultra-0.5 Centrifugal Filter Unit, 3 kDa	Millipore	Cat#UFC500396
Amicon Ultra-0.5 Centrifugal Filter Unit, 50 kDa	Millipore	Cat#UFC505096
MACS SmartStrainers (100 μm)	Miltenyi Biotec	Cat#130-110-917
TC20 Automated Cell Counter	Bio-Rad	Cat#1450102
Cell Counting Slides for TC10™/TC20™ Cell Counter, Dual-Chamber, 30 slides, 60 counts	Bio-Rad	Cat#1450011
TC20™ Data Analyzer Software	Bio-Rad	Cat#SOFT-TC20-Data-Analyzer
Blood Collection tube, BD Vacutainer® K2EDTA 10mL	BD	Cat# BDAM367525
Axygen® 1.7 mL MaxyClear Snaplock Microcentrifuge Tube, Clear, Nonsterile, 500 Tubes/Pack, 10 Packs/Case	Axygen	Cat# MCT-175-C
Eppendorf Safe-Lock Tubes, 0,5 mL, Eppendorf Quality™, incoloro, 500 tubos.	Eppendorf	Cat#525-0129
Eppendorf® Protein LoBind Tube, 1.5 mL	Eppendorf	Cat# 0224311081
5 mL polystyrene Round-bottom Tube (FACS tube)	Fisher Scientific	Cat#352054
HulaMixer Sample Mixer	Thermo Fisher Scientific	Cat#15920D
Cytof2 upgraded to HELIOS	Fluidigm	N/A
NB injector	Fluidigm	Cat#107018
HIMAC CT15RE tabletop centrifuges, HITACHI, Rotor T 15A61-1912	Himac centrifuges	N/A
VWR Digital microcentrifuges VWR2416 digital microcentrifuge, with 24 place rotor	VWR	Cat#521-2839
Centrifuge Mini Spin Eppendorf, fixed –angle rotor F45-12-11	Eppendorf	Cat#521-0012
Multi-Purpose High Speed Centrifuges GYROZEN with Swing-Out Rotor GRS-G-r250-4	Gyrozen	Cat# GZ-1248R
Eppendorf® Centrifuge 5702R	Eppendorf	Cat#EP022628012
Thermo Scientific™ Forma™ 900 Series −86°C Upright Ultra-Low Temperature Freezers	Thermo Scientific	Model# 9470E-907
Liebherr™ NoFrost Comfort Freezer	Fisher Scientific	Cat#15160333
Water-Jacketed, IR Autoflow Automatic CO2 Incubator, UN-4750E	NuAire	N/A
Laminar Air Flow: Telstar bio II A	Telstar	N/A
Brady[R] LABEXPERT printer	Brady	N/A
Brady™ Polyester LABXPERT™ and IDXPERT™ Labels	Brady	Brand: Brady™ 806807
Fisherbrand™ S-Series Heated Ultrasonic Cleaning Bath (FB15046.)	Fisher Scientific	N/A
Falcon™ Round-Bottom Polystyrene Test Tubes with Cell Strainer Snap Cap, 5 mL	Fisher Scientific	Cat#352235
Centrifuge tubes, 15 mL	VWR	Cat#10-9152
Centrifuge tubes, SuperClear, 50 mL	VWR	Cat#525-0156
Vortex mixer, VWR®	VWR	Cat#444-1372
Bemis™ Parafilm™ M Laboratory Wrapping Film	Fisher Scientific	Cat#HS234526C

MATERIALS AND EQUIPMENT

- Centrifugation × g is calculated using the maximum radius.
- All centrifugations are set at 4°C.

- Acceleration and deceleration set to 9 for all centrifugations.
- All the centrifugation during sample thawing and staining are done using refrigerated benchtop centrifuge: Multi-Purpose High Speed Centrifuges Gyrozen 1248R with Swing-Out Rotor GRS-G-r250-4.
- Antibody cocktail are centrifuged using refrigerated microcentrifuge: Himac CT15RE tabletop centrifuges, Hitachi, Rotor T 15A61-1912.
- Barcodes are prepared using Tabletop mini-centrifuge: Centrifuge Mini Spin Eppendorf, fixed –angle rotor F45-12-11, at RT.
- Before CyTOF acquisition samples are prepared using Eppendorf (5702R) centrifuge.

In this protocol the following equipment refers to:

- Vortex: Vortex mixer, VWR®.
- Automated cell counter: TC20 Automated Cell Counter with disposable counting slides (Cell Counting Slides for TC10™/TC20™ Cell Counter).
- Program for cell counting: TC20 Data Analyzer.
- Lab Printer: BradyR LABEXPERT printer.
- Freezer −80°C: Thermo Scientific™ Forma™ 900 Series −86°C Upright Ultra-Low Temperature Freezers.
- Freezer −20°C: Liebherr™ NoFrost Comfort Freezer.
- Sonicator: Fisherbrand™ S-Series Heated Ultrasonic Cleaning Bath (FB15046).
- CyTOF (Helios).
- Rotator: Hula Mixer, Sample Mixer, Invitrogen.
- Strainer for thawing the samples: MACS SmartStrainers (100 μm).
- Tubes with strainers for sample filtration and CyTOF acquisition: Falcon™ Round-Bottom Polystyrene Test Tubes with Cell Strainer Snap Cap, 5 mL.
- FACS tubes: 5 mL polystyrene Round-bottom Tube.

Recipes

Cis-platin (CisPt)

Reagent	Final concentration	Amount/volume
CisPt	5 mM (1000×)	1.5 mg
DMSO	N/A	1 mL

Aliquot and freeze at −20°C for long-term storage, at least 1 year.

⚠ CRITICAL: Cisplatin is a cytotoxic anticancer agent. It should always be handled in a fume hood while wearing protective gloves, face mask, and lab coat.

STEP-BY-STEP METHOD DETAILS
Thaw and lyse samples including reference sample

⏱ Timing: around 2 h, depending on the number of samples

Note: In this example we stain 10 samples + 1 reference sample. More samples/batches can be stained if required. However, to keep the lysis time comparable across the tubes help of another technician is needed, if more than 20 samples are analyzed. Due to long acquisition time on CyTOF (~10^6 cells/hour) the number of samples should be adjusted to the core facility working hours.

1. Thaw frozen donor samples and 1 reference sample.
   a. Take FACS tubes containing preserved blood from −80°C.

b. Place them on a tube rotator at 4°C, until completely thawed (40 min approx.), process samples as soon as they are thawed.

Note: Alternatively, samples can be thawed in cold water bath (∼ 10°C) with agitation.

c. While thawing proceed to step 2.
2. Prepare Thaw-lyse buffer and tubes.
   a. Dilute Thaw-lyse buffer 1000× to 1× in MilliQ water, 13 mL per sample will be required.
   b. Prepare eleven (11) 15 mL tubes and eleven (11) 5 mL FACS tubes. Label them according to the barcoding ID that will be used. Place 100 μm strainers above each 15 mL tube.
   c. Add 5 mL of Thaw-Lyse buffer to each tube, make sure that the strainer area is wet.
3. Lyse red blood cells.
   a. Once the fixed blood is thawed, add 2 mL of Thaw-Lyse buffer to each tube.
   b. Close the tubes tightly and mix by inversion.
   c. Transfer the content of the FACS tubes to the corresponding centrifuge tube with the strainer.

   Note: If clumps are present after thawing, pipet gently to break them down.

   d. Wash each FACS tube with 1 mL of Thaw-lyse buffer in order to collect all the cells and filter through the strainer.
   e. Fill the centrifuge tubes up to 15 mL with Thaw-Lyse buffer, discard the strainers caps, and incubate on the roller for 10 min at RT.
   f. Centrifuge tubes for 10 min at 4°C, 800 g and decant the supernatant.

   Note: If the pellet is still red repeat the lysis using 10 mL of Thaw-Lyse buffer and additional 10 min of incubation.

   g. Wash with 3 mL of CSB (Cell staining buffer), 5 min, 800 g, 4°C.
   h. Decant the supernatant.
   i. Add 2 mL of DPBS and count the cells.

   Note: It is strongly recommended to use an automated cell counter. In this experiment we use a TC20 cell counter.

   Alternatives: A hemocytometer could be used if an automated cell counter is not accessible. Importantly, this will cause add a substantial amount of time to the experiment.

   j. Transfer 1.5×10^6 cells to a labeled FACS tube.

   Note: During the counting we do not perform any life/dead staining. The fixation used in this protocol induce mild permeabilization, hence all cells will be positive for dyes like trypan blue.

   Note: If for some reason cell counts are lower than expected for a given sample, add less cells for this sample and recalculate the volume of antibodies needed conserving the final cell concentration in the staining step. We have stained 1.5×10^6 cells/sample, but the number can be adjusted according to the expected cell frequency of the target populations. If a smaller number of cells is included for a particular sample, it can be difficult to acquire the desired number of cells.

   ⚠ CRITICAL: It is important to standardize thawing conditions across all samples in the study, either thawing them at 4°C using tube rotator or agitating in water (∼ 10°C).

Process the samples as soon as they are thawed. If grumps appear pipet the clumps and perform two lysis steps.

Barcode samples

⏱ Timing: 1 h

To limit intra-batch variation barcoding (labeling individual samples with unique combinatorial barcodes) is introduced before staining.

In this step individual samples are barcoded using palladium-based barcoding (Cell-ID™ 20-Plex Pd Barcoding Kit). Afterward the samples will be pooled in a single staining tube.

4. Dilute Maxpar® 10× Barcode Perm buffer (BPB) with Maxpar® PBS.

   Note: Prepare 1.5 mL/sample, keep the buffer on ice.

5. Wash cells with 500 μL of BPB and centrifuge for 5 min 800 g 4°C.
6. Decant the supernatant and vortex gently.
7. Wash cells with 500 μL of BPB and centrifuge for 5 min 800 g 4°C.

   Note: At this point barcodes can be thawed as described in step 10.

8. Decant the supernatant and vortex the pellet gently.
9. Resuspend in 400 μL of BPB.
10. Prepare barcodes.
    a. Thaw the barcodes at RT.
    b. Spin for 10 s in microcentrifuge (full speed).
11. Barcode the samples.

    Note: Barcode ID per each sample is shown in Table 6.

    a. Using a multichannel pipette resuspend first four barcodes in 100 μL of ice-cold BPB.
    b. Mix well and transfer 100 μL to the FACS tubes containing the cells. Vortex gently.

    Note: Pipetting can lead to cell loss, avoid it.

    ⚠ CRITICAL: Annotate the correspondence of each barcode to its staining tube as shown in Table 6.

    c. Repeat steps a and b for the rest of the barcodes.
    d. Incubate for 20 min at RT.

    Note: During this step surface and cytokine antibody cocktails can be thawed as shown in step 15.

12. Wash the tubes with 2 mL CSB, 5 min at 800 g, 4°C. Discard the supernatant and vortex the pellet.
13. Repeat washing step twice.
14. Pool the barcoded samples in a single FACS tube.
    a. Resuspend cells in each tube in 120 μL of CBS and pool all the samples to a new single FACS tube.

Table 6. Metadata example

Barcode	ID	STIM	Date Exp	Date aqusition	Batch
2	p1	RSQ	181011	181017	day1
3	p1	UNS	181011	181017	day1
4	p1	IMQ	181011	181017	day1
5	p1	CPG	181011	181017	day1
6	p1	LPS	181011	181017	day1
9	p2	RSQ	181011	181017	day1
10	p2	UNS	181011	181017	day1
11	p2	IMQ	181011	181017	day1
12	p2	CPG	181011	181017	day1
13	p2	LPS	181011	181017	day1
15	p2	REF	181011	181017	day1
6	p1	RSQ	181011	181024	day2
7	p1	UNS	181011	181024	day2
8	p1	IMQ	181011	181024	day2
9	p1	CPG	181011	181024	day2
10	p1	LPS	181011	181024	day2
12	p2	RSQ	181011	181024	day2
13	p2	UNS	181011	181024	day2
14	p2	IMQ	181011	181024	day2
15	p2	CPG	181011	181024	day2
16	p2	LPS	181011	181024	day2
17	p2	REF	181011	181024	day2
8	p1	RSQ	181011	181025	day3
9	p1	UNS	181011	181025	day3
10	p1	IMQ	181011	181025	day3
11	p1	CPG	181011	181025	day3
12	p1	LPS	181011	181025	day3
14	p2	RSQ	181011	181025	day3
15	p2	UNS	181011	181025	day3
16	p2	IMQ	181011	181025	day3
17	p2	CPG	181011	181025	day3
18	p2	LPS	181011	181025	day3
19	p2	REF	181011	181025	day3

b. Add additional 120 μL of CSB to wash the original tubes and transfer all the remaining cells to the same FACS tube.

c. Centrifuge 5 min at 800 g, 4°C, decant the supernatant and vortex gently.

Optional: Other barcoding methods can be used, as for example, CD45 or β2-Microglobulin-based barcoding (Hartmann et al., 2018; Lai et al., 2015; Mei et al., 2015). However, β2-Micro-globulin resistance to fixation needs to be checked. We recommend to keep 3 barcodes per sample, to detect and exclude doublets. If surface-based barcoding is used, protocol needs to be adjusted.

Note: Barcodes can be titrated to ensure correct intensity. This protocol is adapted from Cell-ID 20-Plex Pd Barcoding Kit (Standard BioTools).

Surface, intracellular and iridium staining

Ⓣ Timing: 3 h

In this step staining of cell surface and intracellular markers is carried out and followed by DNA staining. On the contrary to flow cytometry, in mass cytometry no side and forward scatter parameters are available, therefore nucleated cells are defined using DNA intercalator. In this protocol we use Cell-ID™ Intercalator- Iridium (Ir) and Maxpar® Fix and Perm Buffer for preparing Ir-staining solution.

15. Prepare the antibody cocktail.
    a. Thaw one aliquot of the antibody cocktail at 4°C.

    Note: Both cocktails (surface and intracellular cytokines) can be thawed at this moment.

    b. Centrifuge 15,000 g for 10 min to sediment antibody aggregates. Keep the intracellular cocktail on ice until needed.
16. Stain cell surface.
    a. Add 298.6 μL of CSB to the cells and 31.4 μL of antibody cocktail as calculated in Table 3, vortex gently.

    Note: Avoid touching the bottom of the tube containing the cocktail, as antibody aggregates can generate staining artifacts.

    Note: If unspecific staining is observed FcR blocking can be done.

    b. Incubate for 30 min at 4°C.
17. Wash for intracellular staining.
    a. Wash cells with 3 mL of Perm-S buffer, centrifuge 5 min at 800 g, and 4°C, discard supernatant, vortex gently.
    b. Repeat Perm-S wash.
18. Stain intracellular cytokines.
    a. Add 309.6 μL of Perm-S to the cells and 20.4 μL of antibody cocktail as calculated in Table 4, vortex gently.

    Note: Avoid touching the bottom of the tube containing the cocktail, as antibody aggregates can generate staining artifacts.

    b. Incubate for 30 min at 4°C.

    Note: The staining time can be extended depending on the cytokine expression.

    c. Wash cells with 3 mL of CSB, centrifuge 800 g, 5 min at 4°C, decant supernatant, vortex gently.
    d. Repeat CSB wash.
19. Stain DNA with Ir.
    a. Calculate the volume of Maxpar® Fix and Perm Buffer needed for Ir staining using the rule $5*10^7$/mL (see Table 7).
    b. Dilute Ir (500 μM) to a final concentration of 0.125 mM in Maxpar® Fix and Perm Buffer.

    Note: Ir needs to be freshly prepared each day due to its instability.

    c. Add calculated volume of Ir solution to the cells.
    d. Vortex gently.
    e. Incubate for 30 min at RT.

    Note: In this example we have stained 1.65×10^7 cells, the total volume of Ir was 1.65 mL.

Table 7. Iridium preparation

	1 sample	11 samples (1 acquisition batch)	48 samples (4 acquisition batches)	Total volume needed (+ ~20%)
Number of cells / sample	1.5×10^6 cells	16.5×10^6 cells	66×10^6 cells	
Staining volume	150 μL	1,650 μL	7,200 μL	8,640 μL
Ir (stock 500 μM)	0.0375 μL	0.4125 μL	1.8 μL	2.16 μL

Preparation of Ir solution for one sample and scaled to 1 batch and a multibatch experiment (4 batches).

20. Centrifuge 5 min at 800 g and 4°C, decant supernatant, vortex gently.
21. Wash with 3 mL of CST.
22. Centrifuge 5 min at 800 g and 4°C, decant supernatant, vortex gently.
23. Prepare 2% PFA from Pierce™ 16% Formaldehyde stock.

   Note: This agent needs to be freshly prepared each day due to its low stability.

24. Resuspend the cells in 2% PFA at 1×10^7/mL. Vortex gently and store at 4°C until acquisition.

   ⏸ **Pause point:** The cells can be stored at 4°C in PFA for up to 72 h.

   Note: The antibody thawing protocol is adapted from (Schulz et al., 2019) and the staining protocol from www.fluidigm.com.

Sample acquisition in a CyTOF device

   ⏱ **Timing:** ~10 h depending on the amounts of cells needed

In this step the cells are prepared for CyTOF2/HELIOS acquisition. It is mandatory to first correctly tune the instrument, as recommended in the instrument manufacturer guidelines, https://www.fredhutch.org/content/dam/www/shared-resources/fc-cytof/CyTOF-Helios-User-Guide.pdf.

It is important to use the same acquisition template throughout the whole project to avoid mislabeling of antibody names. We also strongly advise to prepare a unique naming scheme (see Note below) for the samples and antibody panel acquired in a given study.

25. Prepare the EQ™ Four Element Calibration Beads (beads) working solution.
    a. Sonicate the beads for 10 min.
    b. Calculate the volume needed for the acquisition of the cells, according to Table 8, take 2.1 mL of 1× beads.
    c. Dilute 1× beads in MilliQ in 18.9 mL of water to obtain 0.1× beads solution, see Table 8.
    d. Vortex vigorously every time before use.

   Note: It is recommended to prepare fresh bead solution each time.

Table 8. Calibration beads preparation

# Of samples/batch	# Of cells/sample	Total stained cells/batch	Acquisition cells concentration	[a]Beads solution (0.1×) final volume	EQ four element calibration beads	MilliQ water
11	$1.5*10^6$	$16.5*10^6$	$0.8*10^6$	~21 mL	2.1 mL	18.9 mL

[a]the final volume of the beads solution (0.1×) is calculated as following $16.5*10^6 / 0.8*10^6$. Calculation regarding cell concentration is approximate since we do not account for cell loss during staining.

Alternatives: EQ™ Six Element Calibration Beads (Standard BioTools) could also be used instead of EQ Four Element Calibration Beads. However, these beads can be only used on Helios or CyTOF XT instruments.

26. Prepare sample for the acquisition and acquire.
    a. Resuspend the cells by vortex.
    b. Take an aliquot of 200 μL transfer to a 5 mL FACS tube containing 1.8 mL of CSB.
    c. Centrifuge 5 min at 800 g and 4°C, vortex.
    d. Wash with 2 mL of MilliQ water.
    e. Centrifuge 5 min at 800 g and 4°C, discard the supernatant, vortex.
    f. Add 2 mL of MilliQ water, vortex and take 10 μL of cell aliquot for cell counting.
    g. Centrifuge 5 min at 800 g and 4°C discard the supernatant, vortex.
    h. Resuspend cells in freshly prepared beads at 8×10^5/mL, since the acquisition speed should not exceed 400 events/s.
    i. Introduce the name of the sample in the CyTOF software.

Note: We encourage the users to be consistent with experiment naming across all batches. It is advisable to include batch number, the acquisition date and the experiment name in the name of the fcs file. This will facilitate the analysis using CytoQP package and experiment tracking. For example: day1_human_tlr_response_220622.fcs.

    j. Acquire every aliquot in a Helios mass cytometer for no longer than 1.5 h to avoid sample degradation.
    k. Repeat steps (a) to (i) until the target number of cells is acquired.
    l. Proceed to data analysis pipeline described in quantification and statistical analysis section for data preprocessing and cleaning.

Note: If an acquisition is longer than 1.5 h for 200 μL, reduce the volume of the aliquot taken.

Alternatives: In this example samples were washed and acquired using MilliQ water with narrow bore (NB) injector, however cell acquisition solution (CAS) with wide bore (WB) injector can also be used. Samples can be acquired for longer times with CAS solution, though we advise aliquot acquisition since it reduces the size of each fcs file generated and speeds up the analysis time.

EXPECTED OUTCOMES

At the end of this protocol several fcs files will be generated. In this particular example 3 experiments were acquired in 3 different days. In each day, 11 samples were barcoded, pooled, stained and acquired in aliquots. Therefore, in each day, around 12 aliquots were acquired, thus 12 fcs files were generated for each batch.

The generated aliquots need to be preprocessed in order to obtain clean and artifact-free data, as discussed (Rybakowska et al., 2020). To do this, files need to be first normalized using the beads spiked in the acquisition tubes (Finck et al., 2013). Next, flow rate and signal aberrations need to be removed (Meskas et al., 2022; Monaco et al., 2016), followed by staining quality check per batch (Amir et al., 2018). After that, files need to be debarcoded and aggregated in order to obtain individual specific fcs files (Zunder et al., 2015). For further analysis, cleaning of doublets, dead cells or debris needs to be performed, thus files are gated using MC-specific gating strategy (Olsen et al., 2019). After the removal of bad quality events, files can be subjected to another step of normalization based on the reference samples included in every batch (Gassen et al., 2020). The summary of the workflow can be found in Figure 3.

After preprocessing steps are completed, files can be subjected to further exploration and analysis using e.g., dimensional reduction, clustering algorithms or manual gating, as shown in (Rybakowska et al., 2021b).

The above-mentioned steps are further described in the section quantification and statistical analysis. The package called CytoQP (GitHub: https://github.com/prybakowska/CytoQP, Zenodo: https://doi.org/10.5281/zenodo.6947414) was developed based on the published workflow (Rybakowska et al., 2021b) and is used to clean the data.

The R code is described below, the script (CytoQP_script.R) is provided on GitHub (https://github.com/prybakowska/CytoQP/blob/master/CytoQP_script.R) and as a supplemental information, Data S1, CytoQP_script.R. Additionally, a data set described in the experimental part is available and can be downloaded from FlowRepository: www.flowrepository.org (Spidlen et al., 2010) with the accession number FR-FCM-Z3YR, thus readers can try to use the code on their own.

QUANTIFICATION AND STATISTICAL ANALYSIS

⏱ Timing: 1 h for installation and ∼ 30 min/file for preprocessing

Note: The time depends on the number of files, their size and the computer specification. The above time was calculated using intel core i7, 8th Gen, RAM 16 GB and Windows operating system.

This part includes all the major steps necessary to process and clean collected data. Here we use the R script and package called CytoQP that processes your data in a semi-automatic way, thus knowledge of R, RStudio and some programming skills are required. However, our package is prepared in such a way that minimal computational knowledge will be needed. To execute the script, we will use the data generated above. The workflow of data generation and its structure can be found in Figures 1 and 2.

Example data are acquired and analyzed in aliquots to speed up the analysis time and avoid working with big fcs files. However, if files were barcoded and acquired as single samples they can also be analyzed with some adjustments marked in the note. Although, if the computer power resources are limited, we recommend to split the data into smaller fcs files e.g., 700 000 counts per file, preprocess as shown in Figure 3 and aggregate. To split the files, function `split_big_flowFrames` can be used.

MC and to some extent flow cytometry data can be analyzed. However, for the purpose of this paper only MC data is used. Therefore, parameter data_type across functions is set as default: MC data.

In the script we divide markers into phenotypic and functional. The phenotyping markers are those used for clustering, included in Table 3 and functional markers are the cytokine markers for which MSI (median signal intensities) are extracted, Table 4.

To follow this part of the manuscript, download and install R (4.1.2) and RStudio (2021.9.2.382)., and all necessary files, including fcs files and meta data. To execute the code, copy and paste the code provided in the text box to a new R script in the R project.

The code is organized in blocks. The outcome of one step can be directly used in the next block, unless stated otherwise.

Download necessary libraries, data files and setup RStudio

⏱ Timing: 1 h

1. To download and install R go to: https://www.r-project.org/. The current pipeline was performed using R version 4.1.2.
2. To download and install RStudio go to https://www.rstudio.com/products/rstudio/. The current pipeline was performed using RStudio version 2021.9.2.382.

   ⯈⯇ **Pause point:** Since this part of the protocol follows a computational pipeline exclusively, by saving the executed steps the protocol can be paused at any time. Note that the necessary objects (matrices, data frame or file pathways would need to be read in before continuing the analysis).

3. Set R studio creating the project.
   a. Open RStudio and go to File > New Project > New Directory > New Project.
   b. Navigate to *Directory name* and create the name e.g., *CytoQP_analysis*.
   c. Navigate to *Create project as a subdirectory of:* and select the subdirectory.
   d. Click *Create Project*.

   Note: At this point a new folder called *CytoQP_analysis* will be created in the selected directory.

4. Download fcs files.
   a. Go to: FlowRepository: https://flowrepository.org/id/FR-FCM-Z3YR.
   b. Click on orange bottom *download*, next mark that you are not a robot and click *ZIP and Download Files*.

   Note: For the analysis to progress faster some selected files can be downloaded, however for the script to show its purpose at least two files from each day (*day1*, *day2*, *day3*) should be downloaded. The *gating_strategy.wsp* and *meta_data.csv* can be omitted.

   c. Unzip downloaded files directly to the previously generated folder *CytoQP_analysis*.

   Note: If needed download compression tool to unpack files, to do this go to: https://www.win-rar.com/start.html?&L=12, click download WinRAR ad follow the instructions.

5. Install CytoQP package (GitHub: https://github.com/prybakowska/CytoQP).
   a. Copy and paste the following code to R studio script or copy the script CytoQP_script.R from Data S1 or from GitHub: https://github.com/prybakowska/CytoQP/blob/master/CytoQP_script.R.

   Note: This installation will automatically download and install CytoQP dependencies. These dependencies can be found in Table 10.

```
if(!require(devtools)){

install.packages("devtools") # If not already installed

}

devtools::install_github("prybakowska/CytoQP")

library(CytoQP)
```

Analyze the data using provided R code

⏱ Timing: 30 min/file

6. Perform bead normalization.

In CyTOF instrument signal drop caused by progressing instrument decalibration is frequently observed. It is especially visible when barcoded samples are collected for long time. In order to correct for it bead-based normalization was introduced (Finck et al., 2013). The algorithm uses beads spiked and acquired together with the sample, as shown in section "Sample acquisition in a CyTOF device" (steps 25 and 26)". This allows to track and correct signal changes during acquisition time.

Note: The bead normalization functions are based on CATALYST package (Crowell et al., 2022), however some adjustments were made. Instead of reading all the aliquots as a flowSet or aggregating them in one big file, sample aliquots are normalized one-by-one. This avoids generation of big R objects and speed up the analysis process. This step is necessary when tens of heavy files are analyzed.

Note: To allow aliquot normalization using a common transformation factor for all files, a base-line file is created by aggregation of all sample aliquots and used to obtain one transformation factor. Thus, all the sample aliquots are normalized to the mean bead intensity from the aggregated file. Maximum of 25 000 cells are aggregated form each fcs aliquot, so approximately 250 beads are taken per aliquot. Otherwise, one random aliquot can be used as a reference file, although user should make sure that this aliquot represents the rest of files.

Note: If cells were acquired in one big aliquot, we recommend to split them using function `split_big_flowFrames` and follow the script. Alternatively, files can be normalized one by one, however the number of cells to create *baseline_file* should be increased (Table 9) or one file can be used to create a baseline using `baseline_file` function. If the files are not split it can be computationally difficult and slow to preprocess them.

Note: The below script generates a folder called *BeadNorm*, where normalized files with suffix *beadNorm.fcs* can be found. These files do not contain beads anymore, as they were removed during the normalization. Additionally, in the subfolder *Plots_BeadNormalization* 2 panels are generated per aliquot as shown in Figure 3, step 1. Panel (A) allows to check the quality of beads gating and panel (B) allows to verify bead intensities before and after normalization. We therefore recommend to take a careful look at both panels and across fcs aliquots.

a. Navigate to the folder *BeadNorm > Plots_BeadNormalization* and verify if the normalization was correctly performed by checking beadGate.png and beadLines.png plots.

```
# ————————————————————————————————————————————————————————————————————————————

# Bead normalization —————————————————————————————————————————————————————————

#—————————————————————————————————————————————————————————————————————————————

# define full pathway to the files that you want to normalize

files_raw <- list.files(path = getwd(),

          pattern = ".FCS",

          full.names = TRUE)

# create baseline file to which all the files will be normalized

set.seed(2)
```

```
ref_sample <- baseline_file(fcs_files = files_raw,

                beads = "dvs")
# Normalize files
files_beadnorm <- bead_normalize(files = files_raw,

                cores = 1,

                non_mass_channel = NULL,

                norm_to_ref = ref_sample,

                to_plot = TRUE,

                remove_beads = TRUE,

                k = 80,

                markers_to_keep =

                  c("CD", "HLA", "IgD", "TCR", "Ir",

                    "Viability","IL", "IFNa",

                    "TNF", "TGF", "MIP", "MCP", "Granz"))
```

Note: Although the above function generates useful diagnostic plots, we believe that it is important to verify marker behaviors upon normalization. Thus, we recommend to plot all the markers that were bead normalized and verify their expression before and after the normalization.

Alternatives: Premessa package (GitHub: https://github.com/ParkerICI/premessa#bead-based-normalization) with graphical user interface (GUI) could also be used. However, the workflow would need to be adjusted.

b. (optional) visualize the behavior of all the markers before and after normalization executing the code below and check the following file *Marker_distribution_across_aliquots_and_batches.pdf* in *BeadNorm > Plots_BeadNormalization.* The example visualization of marker CD11c-Tb159 can be found in Figure 5A.

```
# —————————————————————————————————————————————————————————
# Visualized files after bead normalization——————————————————————————————
#——————————————————————————————————————————————————————————
# Define batch id and sample id for each file
batch_pattern <- stringr::str_match(basename(files_raw),

                "(?i).*(day[0-9]*).*.FCS")[,2]
plot_marker_quantiles(files_after_norm = files_beadnorm,

                files_before_norm = files_raw,

                batch_pattern = batch_pattern,

                plot_name = "Marker_distribution_across_aliquots.pdf",

                arcsine_transform = TRUE,

                remove_beads = TRUE,
```

```
        bead_channel = "140",

        uncommon_prefix = "_beadNorm.fcs|.FCS",

        markers_to_plot = c("CD", "HLA", "IgD", "IL",

                "TNF","TGF", "GR", "IFNa"),

        manual_colors = c("darkorchid4", "darkorange",

                "darkgreen"))
```

Note: If only one sample was acquired per batch, split the files using function `split_big_-flowFrames()` and visualize them.

7. Clean flow rate and signal instability.

Sample acquisition in CyTOF can suffer from sudden changes in the flow rate and clogging, which affect data quality. Hence, it is important to detect and clean signal anomalies. To do this we use two algorithms *flowAI* (Monaco et al., 2016) to spot flow rate irregularities and *flowCut* (Meskas et al., 2022) to track signal instability. Briefly, both algorithms divide cell events into bins/segments and compare them to detect outliers. The functions from these packages are adapted to mass cytometry data and written to wrapper functions `files_clean`.

```
# ———————————————————————————————————————————————————————————————————

# Signal Cleaning ————————————————————————————————————————————————————

# ———————————————————————————————————————————————————————————————————

files_clean <- clean_files(files = files_beadnorm,

        cores = 1,

        to_plot = "All",

        data_type = "MC",

        Segment = 1000,

        arcsine_transform = TRUE,

        non_used_bead_ch = "140")
```

Note: If samples were acquired in one big aliquot, parameter adjustments may be required. See Table 9 for more details.

Note: The above step will generate a new folder called *Cleaned* where new *cleaned.fcs* files will be stored. Additionally, two subfolders are created called *FlowRateCleaning* and *SignalCleaning* where 1 plot per file is generated, as shown in Figure 3, step 2 (A) flow rate, (B) signal cleaning. These plots are handy to check the quality of the acquisition rate and signal across the markers, but also the level of the cleaning achieved. Therefore, we recommend to verify if the bad quality events are removed from the data. If not, the parameters can be adjusted. Please refer to Table 9 for parameter adjustments.

Alternatives: PeacoQC (Emmaneel et al., 2022) could also be used to clean the flow rate and signal instability, however this package is currently not available in CytoQP package.

8. Detect aliquot outliers.

Table 9. Parameter adjustment for selected functions

Function name and parameter	Adjustment description
baseline_file	
ncells – number of cells to be aggregated per each file, default 25000	The number of cells is 25 000 so around 250 beads will be detected per file, it is recommended to increase it when one big aliquot is analyzed, to obtain better representation of the beads across file.
clean_files	
alpha – the statistical significance level used to accept anomalies, default 0.01. Applies to flow rate cleaning.	Decrease or increase for less or more strict cleaning level
Segment – specifies the number of events in each segment to be analyzed. Default 1000. Applies to signal cleaning.	We work with ~700 000 events per file, thus default setting works well for us. However, this parameter is sensitive to the number of events collected in fcs file. The more cells are acquired the more cells should be analyzed per segment, keeping in mind the total number of segments to obtained for good statistics. The same apply to the removal of small or big changes in the mean signal, the bigger the segment (the more events) bigger mean change can be removed. Should be adjusted according to the number of events and the type of signal disturbances.
MaxPercCut – the maximum percentage of cells to be removed, default 0.5. Applies to signal cleaning.	This parameter depends on the quality of the data and the number of events that the user can lose. The higher value will increase the number of bad events removed, however more cells will be discarded from further analysis.
UseOnlyWorstChannel – enable automated detection of worst channel to be cleaned. Default TRUE. Applies to signal cleaning.	Setting this parameter to TRUE ensures strict cleaning, otherwise bad quality events can be missed when taking into consideration the statistics of all channels together. Can be set to FALSE if too many events are lost.
AllowFlaggedRerun – specify if flowCut will run second time. Default TRUE. Applies to signal cleaning.	It can happen that the signal in other channels is also affected but it will not get cleaned with predefined channel. Setting this parameter to TRUE ensures strict cleaning, and can be set to FALSE if too many events are lost.
AlwaysClean – the file will be cleaned even if it is relatively stable. The segments that are 7 SD further from the mean of all segments are removed. Default TRUE. Applies to signal cleaning.	Setting this parameter to TRUE ensures strict cleaning, and can be set to FALSE if too many events are lost. We did not notice any excessive cleaning when keeping this parameter set to TRUE.
file_quality_check	
file_batch_id – Character vector with batch label for each fcs_file, the order and the length needs to be the same as in fcs_files. If only one batch or one file is processed the parameter should be left as NULL (default).	When one aliquot per experiment is acquired or only one batch is analyzed this parameter should be left as default. In this case the AOF scores and quality score will be calculated treating fcs files as they are coming from one experiment.
sd – number of standard deviations allowed for file outlier detection. Default is set to 3.	Can be lowered or increased if more files need to be discarded or kept, respectively. It is advisable to increase when one aliquot per experiment is analyzed.
nClus – numeric, as in FlowSOM, Exact number of clusters for meta-clustering. Default = 10.	Can be increased if more cell populations need to be detected
Debarcode_files	
min_threshold – Logical, if the minimal threshold for barcoding should be applied. Default TRUE	Can be set to FALSE, however user will not be notified if bad quality barcoding is spotted
threshold – Numeric, value for the minimum threshold for debarcoding, default is set to 0.18.	Works only if min_threshold set to TRUE. If increased, the higher quality events will be taken, but the lower amounts of cells will be debarcoded and vice versa, if decreased.
barcode_key – debarcoding scheme as in CATALYST. Binary matrix with sample names in the row names and barcode names in the column names or a vector of numeric masses corresponding to barcode channels (in this case the barcode scheme will be created internally)	Needs to be adjusted if other than 6-choose-3 scheme was chosen, or less than 20 barcodes were selected.
gate_intact_cells	
tinypeak_removal_head – Numeric from 0-1, as in deGate to exclude/include tiny peaks in the head of the density distribution curve for both Iridium channels. Default 0.8	The modification of this parameter will affect the inclusion or exclusion of Ir low events and its modification is useful when high content debris (dense peaks) are present. In our experience the setting to 0.8 is enough to remove Ir low content events.
tinypeak_removal_tail – the same as *tinypeak_removal1* but for the tail in the density distribution curve. Default 0.8	The same as above but regarding the events in the tail (high content of Ir)
alpha_head – Numeric, 0–1, as in deGate specify the significance of change in the slope being detected at the head of the density distribution curve. Default 0.05.	If increased, a more tide gate, will be drawn at the head of the distribution.
alpha_tail – the same as *alpha_head* but for the tail in the density distribution curve. Default 0.1	If increased, a more tide gate, will be drawn at the tail of the distribution.

Table 9. *Continued*

Function name and parameter	Adjustment description
gate_live_cells	
viability_channel – character, the channel name used for viability staining.	If other viability stain as for example Rh103 is used, the parameter needs to be adjusted, by writing Rh103Di.
* *Parameters: tiny_peak_removal_viability, alpha_viability, tinypeak_removal_Iridium, alpha_Iridium* – as above	Works the same as described above, but for described channels.
gate_singlet_cells	
n_mad - Numeric, number of MADs to detect outliers. Default set to 2.	If increased, a more tide gate, will be drawn.
train_REF_model	
nQ - Numeric, as in CytoNorm, number of quantiles to use. Default = 101, which results in quantiles for every percent of the data.	Although the default parameter is 101 we use 2 quantiles (0.05 and 0.95), as it was reported that quantile normalization can introduced artifacts. If the modeling of marker distributions is no adequate, users can switch to default setting.
limit - Numeric, as in CytoNorm, these values are modeled to map onto themselves by the spline.	We set it to 0–8, as recommended by the authors to avoid the introduction of extreme values, like negative values that are normally absent in MC data. This is especially useful when nQ=101, however can cause introduction of artifacts due to forced spline fitting in between the limits.
quantileValues - If specified, it should be a vector of length nQ with values between 0 and 1, giving the percentages at which the quantiles should be computed. If NULL (default), the quantiles will be evenly distributed, including 0 and 1.	We choose percentile 0.05 and 0.95, as a most common and robust normalization values. Can be adjusted according to the user's needs.
Goal - Goal distribution. Default "mean", can also be nQ numeric values or one of the batch labels.	We kept this parameter at default. However, can be adjusted if needed.
norm_with_clustering - Logical, if data should be normalized using clustering algorithm, FlowSOM. Default set to FALSE.	In this setting we perform data normalization without clustering, however it can be needed if population-specific batch effect occurred.
extract_pctgs_msi_per_flowsom	
nCells - Numeric, number of cells to be cluster per each file, default is set to 50 000	Can be increased or decreased depending if rare cells are analyzed. User should ensure that all cells can be represented in aggregated file at a reasonable amount.

In this pipeline, barcoded samples are acquired, hence long-time acquisition is done. For this reason, changes in signal intensity or sudden problems with instruments, like unexpected shutdown or additional tuning, can happen. Thus, it is safe to verify the channel intensities across acquired aliquots in each batch. To do this, we take advantage of Average Overlap Frequency (AOF) algorithms (Amir et al., 2018, 2019) that detect potential staining problems, and the cytutils package. (https://github.com/ismms-himc/cytutils). As an input we use FlowSOM (Gassen et al., 2015) algorithm with the default setting to gate the data. We also provide useful diagnostic plots that mark problematic markers and aliquots. All of this is written in one wrapper function called `file_quality_check`. Please refer to Table 9 for parameter adjustment. If other clustering parameters are to be used, we recommend to run functions `fsom_aof` separately, followed by function `aof_scoring` and `scaled_aof_score`.

```
# ————————————————————————————————————————————————————————————————

# File outliers detection ————————————————————————————————————————

# ———————————————————————————————————————————————————————————————

# Define batch_id for each file

file_batch_id <- stringr::str_match(basename(unlist(files_clean)),

                "(day[0-9]*).*.fcs")[,2]

file_scores <- file_quality_check(fcs_files = files_clean,

                file_batch_id = file_batch_id,

                phenotyping_markers = c("Ir","CD",
```

```
                    "HLA", "IgD")

        arcsine_transform = TRUE,

        nClus = 10,

        sd = 3)
```

Note: This code generates a dotplot (Figure 3, step 3A) with bad quality aliquots highlighted in green. Furthermore, a data frame with aliquots scored as *good* and *bad* quality is produced and saved as *Quality_AOF_score.RDS*, hence bad quality files can be easily removed from further analysis in the next step. Additionally, the following diagnostic plots are generated: heatmap for scaled AOF scores per marker for each sample, FlowSOM tree and t-SNE colored by metaclusters to verify the quality of the clustering (Figure 3, step 3B and C, respectively). Moreover, heatmap for AOF scores per marker is also plotted (not shown).

Note: The AOF algorithm is designed to verify the staining quality of bimodal makers, as it calculates the separation between positive and negative populations. Thus, at this point of analysis we only use the phenotyping markers. We assume that if some markers scored low, the analysis of that particular sample will be affected. However, users can include any marker for AOF scoring using parameter *markers_to_score*.

Note: We use default parameters of FlowSOM (10 metaclusters) in order to validate marker expression in main leukocyte populations. We assumed that a lower amount of metaclusters will be less sensible to potential batch effects, thus similar cells will be assigned to the same group across the aliquots and AOF score will be calculated for the same types of cells.

Table 10. CytoQP dependencies

| Package name	Publication	Source
CATALYST	(Crowell et al., 2022)	https://www.bioconductor.org/packages/release/bioc/html/CATALYST.html
dplyr	(Wickham et al., 2022)	https://cran.r-project.org/web/packages/dplyr/index.html
flowCore	(Ellis et al., 2022)	https://bioconductor.org/packages/release/bioc/html/flowCore.html
flowCut	(Meskas and Wang, 2022)	https://bioconductor.org/packages/release/bioc/html/flowCut.html
flowDensity	(Malek and Taghiyar, 2022)	https://bioconductor.org/packages/release/bioc/html/flowDensity.html
FlowSOM	(Van Gassen et al., 2015)	https://bioconductor.org/packages/release/bioc/html/FlowSOM.html
ggplot2	(Wickham et al., 2021)	https://cran.r-project.org/web/packages/ggplot2/index.html
ggpubr	(Kassambara, 2020)	https://cran.r-project.org/web/packages/ggpubr/index.html
graphics	(R Core Team, 2013)	R Core Team
grDevices	(R Core Team, 2013)	grDevices
grid	(R Core Team, 2013)	R Core Team
gridExtra	(Auguie and Antonov, 2017)	https://cran.r-project.org/web/packages/gridExtra/index.html
pheatmap	(Kolde, 2019)	https://cran.r-project.org/web/packages/pheatmap/index.html
RColorBrewer	(Neuwirth, 2022)	https://cran.r-project.org/web/packages/RColorBrewer/index.html
S4Vectors	(Pagès et al., 2022)	https://bioconductor.org/packages/release/bioc/html/S4Vectors.html
stats	(R Core Team, 2013)	R Core Team
stringr	(Wickham and RStudio, 2019)	https://cran.r-project.org/web/packages/stringr/index.html
SummarizedExperiment	(Morgan et al., 2022a)	https://bioconductor.org/packages/release/bioc/html/SummarizedExperiment.html
utils	(Bengtsson, 2021)	https://cran.r-project.org/web/packages/R.utils/index.html
uwot	(Melville et al., 2021)	https://cran.r-project.org/web/packages/uwot/index.html
CytoNorm	(Gassen et al., 2020)	https://github.com/saeyslab/CytoNorm
cytutils	(Amir et al., 2018)	https://github.com/ismms-himc/cytutils
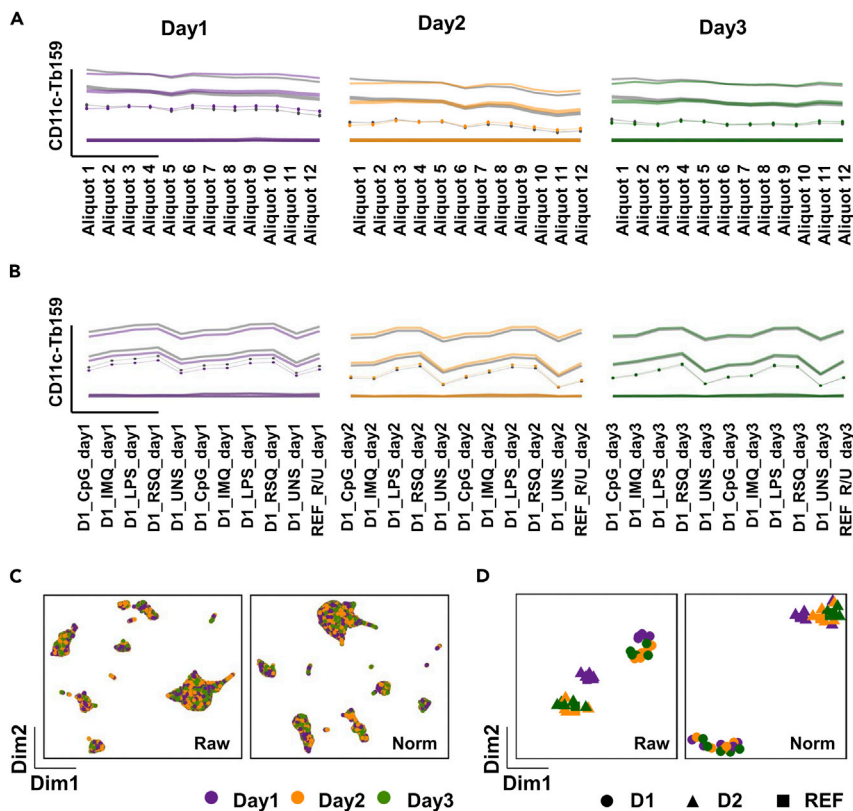
Figure 5. Visualization after normalization

(A) All aliquots are plotted for each acquisition day for CD11c-Tb159. The lines represent percentiles, dots represent median values. The thicker lines represent 25 and 75 percentiles. The thinner lines show the 1 and 99 percentiles. Gray lines represent the percentiles before normalization, and colored lines, after normalization.

(B) Data represented as in A for aggregated and deconvoluted, donor-specific fcs files. R/U (RSQ/UNS).

(C) Scatter dot plots representing two dimensions obtained from the UMAP analysis. UMAP was built on aggregated files (1,000 cells per fcs file) using default parameters and phenotyping markers (Table 3). Data are shown before (Raw) and after, normalization (Norm) with the reference sample. Cells are colored according to the staining day. (D) Data were clustered using FlowSOM, metacluster frequencies and MSI for functional markers (Table 4) were extracted. Next, dimensional reduction using UMAP was performed for metacluster frequencies (not shown) and MSI for all markers to verify sample distribution. Data are represented as in C. The left panel represents sample distribution before normalization and the right panel after normalization. Samples are colored by staining batch. Donors and reference sample are represented with shapes.

Note: If the data were acquired as one big file we advise to divide it into segments (using `split_big_flowFrames()` function) and proceed as suggested in this script. Alternatively, the quality score can be calculated treating fcs files as they are coming from the same acquisition batch. In this case the parameter *file_batch_id* and *sd* should be adjusted as suggested in Table 9. It should be noted that by this change, the whole experiment can be marked as an outlier, and rather than discarding it from further analysis we strongly advise to further investigate the quality of the staining. It can happen that the staining discrepancies come from the donor-specific differences and not from the experimental setup. In our setting we apply AOF to detect abnormalities across aliquots which have equal cell and marker distribution. Thus, removing one or few aliquots does not remove any cell population.

9. Debarcode files.

To obtain sample-specific fcs files the aliquots are debarcoded. To do this, we take advantage of *CATALYST's* function. But as in the bead-based normalization part, we work with one flow frame

at a time. Additionally, we include a minimal separation cutoff, which ensures safe separation threshold and shows files with unclear distribution. To speed up the analysis and avoid generation of heavy fcs files, the debarcoding step is taken before the file's aggregation. Thus, for each aliquot, multiple debarcoded files are generated and the total number of debarcoded files is equal to the total number of barcodes used. Additional file named "unassigned.fcs" is also generated and contains the cells that were not assigned to any barcode. At this point the meta data file containing the sample assignation to the barcodes needs to be provided. The example of the meta data table used here is shown in Table 6.

```
# ——————————————————————————————————————————————————————————————————

# Files debarcoding ————————————————————————————————————————————————

#———————————————————————————————————————————————————————————————————

# Define file batch ID for each file

file_batch_id <- stringr::str_match(basename(unlist(files_clean)),

                "(day[0-9]*).*.fcs")[,2]

# Load metadata or read in

md <- meta_data

# read in or create barcode key

sample_key <- CATALYST::sample_key

# Extract information about barcodes used in each batch

barcodes_list <- list()

for (batch in unique(file_batch_id)){

idx <- md[md[,"BATCH"] == batch, "BARCODE"]

barcodes_list[[batch]] <- rownames(sample_key)[idx]

}

# Debarcode files

files_debarcoded <- debarcode_files(fcs_files = files_clean,

                file_score = file_scores,

                min_threshold = TRUE,

                barcodes_used = barcodes_list,

                file_batch_id = file_batch_id,

                less_than_th = TRUE,

                barcode_key = sample_key)
```

Note: The above script generates a folder called *Debarcoded* with subfolders for each acquisition day. In this case, 3 subfolders called *day1*, *day2*, and *day3* are generated and contained fcs files with the barcode name as a prefix and *_debarcoded.fcs* as suffix. Furthermore, the code generates two plots as shown in Figure 3, step 4, one for the distribution of barcode separation and cell yield upon debarcoding (A) and the other with normalized intensities for selected barcode and separation threshold (B). Additionally, the names of the files with separation threshold below 0.18 are stored in the RDS file called *files_with_lower_debarcoding_-threshold.RDS* if the parameter less_than_th is left at default, TRUE. The plots and RDS file are stored in the *Debarcoded* folder.

Note: In this protocol we use 6-choose-3 palladium barcoding scheme. If other scheme is used, the *sample_key* and *barcodes_list* would need to be redesigned.

Note: If one fcs file of one batch is analyzed, the code does not require any adjustment.

Alternatives: premessa package (GitHub: https://github.com/ParkerICI/premessa#bead-based-normalization) with graphical user interface (GUI) could also be used. However, the workflow would need to be adjusted.

10. Aggregate and deconvolute the names of debarcoded files.

To combine debarcoded aliquots into individual-specific fcs files, we perform file aggregation using the function *aggregate_files.* To do this, the same metadata as in section Debarcode files and Table 6 needs to be provided.

Note: This step is required even if the data does not require the aggregation, as the new, deconvoluted name will be set at this point. No function adjustment is required.

```
# ———————————————————————————————————————————————————————————————
# Files aggregation and file name deconvolution ————————————————————
# ———————————————————————————————————————————————————————————————
# Assign barcodes names
md$barcode_name <- paste0(rownames(CATALYST::sample_key)[md$BARCODE])
# Assign new sample names specifying patient id and its batch name
md$fcs_new_name <- paste0(md$ID, "_", md$STIM, "_", md$BATCH, ".fcs")
# Aggregate and deconvolute file names
files_agg <- aggregate_files(fcs_files = files_debarcoded,
            md,
            barcode_column = "barcode_name",
            batch_column = "BATCH",
            cores = 1,
            write_agg_file = TRUE)
```

Note: The aggregated files will be stored in a folder called *Aggregated* and they will contain sample specific names, provided in the meta data.

11. Gate the files.

In this step we perform fcs file gating to select intact, singlets and viable cells. This part of the code is based on *flowDensity* package.

```
# ———————————————————————————————————————————————————————————————
# Files gating ————————————————————————————————————————————————————
# ———————————————————————————————————————————————————————————————
```

```
# Create directory to store plot

gate_dir <- file.path(getwd(), "Gated")

if(!dir.exists(gate_dir)){dir.create(gate_dir)}

# Gate the files and plot the gating strategy for each file

n_plots <- 3

png(file.path(gate_dir, paste0("gating.png")),

  width = n_plots * 300, height = length(files_agg) * 300)

layout(matrix(1:(length(files_agg) * n_plots),

ncol = n_plots, byrow = TRUE))

for (file in files_agg){

  print(file)

  ff <- flowCore::read.FCS(filename = file,

              transformation = FALSE)

  ff <- gate_intact_cells(flow_frame = ff,

          file_name = basename(file),

          save_gated_flow_frame = FALSE)

  ff <- gate_singlet_cells(flow_frame = ff,

          channels = "Event_length",

          file_name = basename(file),

          save_gated_flow_frame = FALSE)

  ff <- gate_live_cells(flow_frame = ff,

          viability_channel = "Pt195Di",

          save_gated_flow_frame = TRUE,

          file_name = basename(file), suffix = "_gated")

}

dev.off()
```

Note: The scheme of the gating strategy for one file is shown in in Figure 3, step 6. This gating is generated for each file and stored as *.png* figure in *Gated* folder. We advise to carefully verify the gating quality.

Note: The setting for gating functions can be modified, see Table 9.

Note: This gating strategy is designed for detecting Ir-stained human nucleated leukocytes, however it could also be applied to other cell types, such as mouse splenocytes. If non-nucleated cells are of interest the gating strategy should be redesigned.

12. Normalize files using reference sample.

Bead normalization helps to correct day-to-day variation caused by the CyTOF instrument, unfortunately it does not consider variation introduced during experimental sample manipulation.

Therefore, normalization using a reference sample that is present in every batch during sample staining is advisable.

In this step we take advantage of a reference sample and perform batch normalization using the CytoNorm package (Gassen et al., 2020). To do this, we compute 5% and 95% percentile across all the reference samples for each marker and obtain the markers specific goal distribution. This information is then used for batch normalization, see Figure 3, step 7.

```
# ———————————————————————————————————————————————————————————————————

# Normalization using reference sample ——————————————————————————————

#———————————————————————————————————————————————————————————————————

# Set input directory

gate_dir <- file.path(getwd(), "Gated")

# Define files for normalization including reference sample

files_ref <- list.files(gate_dir, pattern = "*_gated.fcs$",

          full.names = TRUE, recursive = TRUE)

df <- data.frame("file_paths" = files_ref,

        "batch_labels" = stringr::str_match(

        files_ref, "day[0-9]*")[,1],

        "ref_ids" = grepl("REF", files_ref))

# Build normalization model

model <- train_REF_model(df = df,

            markers_to_normalize =

              c("CD", "HLA", "IgD","IL", "TN",

                "MCP", "MIP","Gran", "IFNa", "TG"),

            arcsine_transform = TRUE,

            nQ = 2, limit = c(0,8),

            quantileValues = c(0.05, 0.95),

            goal = "mean",norm_with_clustering = FALSE,

            save_model = TRUE)

# Normalize files

normalize_REF(model = model, df = df, arcsine_transform = TRUE,

            norm_with_clustering = FALSE)
```

Note: This code generates fcs files with prefix *Norm*. The normalized files are stored in the folder called *CytoNormed* together with diagnostic plots. Plot examples are shown in Figure 3, step 7. We recommend investigating the generated panels and adjusting the parameters if necessary, for the parameter setting refer to Table 9 and *CytoNorm* package.

Note: If no reference sample was acquired this step should be skipped.

As a proof of concept for our protocol we compared cell population and cytokine positive population frequencies as well as cytokine MSIs, using spearman correlation. We assume that the correlation will be high if the data are properly curated. These results can be found in Figure S1.

13. Plot batch effect.

Although the above function generates valuable diagnostic plots, we believe that it is important to verify markers' behavior upon normalization. Thus, we recommend to plot all the markers that were normalized with CytoNorm and verify their expression before and after the normalization using the function `plot_marker_quantiles`. Additionally, we advise to visualize batch effects using dimensional reduction methods and the function `plot_batch`, which uses UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) (McInnes et al., 2018). We suggest to extract cell frequencies relative to live cell gate and median signal intensities (MSI) for subsequent visualization across batches. To do this we take advantage of FlowSOM algorithm and function extract `extract_pctgs_msi_per_flowsom`.

```
# ——————————————————————————————————————————————————————————————
# Plot batch effect ————————————————————————————————————————————
#——————————————————————————————————————————————————————————————
# Define files before normalization
gate_dir <- file.path(getwd(), "Gated")
files_before_norm <- list.files(gate_dir,pattern = ".fcs", full.names = TRUE)
# Define files after normalization
norm_dir <- file.path(getwd(), "CytoNormed")
files_after_norm <- list.files(norm_dir, pattern = ".fcs",full.names = TRUE)
# files needs to be in the same order, check and order if needed
test_match_order(x = basename(gsub("Norm_","",files_after_norm)),
basename(files_before_norm))
batch_labels <- stringr::str_match(basename(files_before_norm), "day[0-9]*")[,1]
# Plot batch effect
set.seed(789)
plot_batch(files_before_norm = files_before_norm,
      files_after_norm = files_after_norm,
      batch_labels = batch_labels,
      cores = 1, out_dir = norm_dir,
      clustering_markers = c("CD", "IgD", "HLA"),
      manual_colors =
        c("darkorchid4", "darkorange", "chartreuse4"))
# Plot quantiles
batch_pattern <- "day[0-9]*"
plot_marker_quantiles(files_after_norm = files_after_norm,
          files_before_norm = files_before_norm,
```

```
            batch_labels = batch_labels,

            arcsine_transform = TRUE,

            markers_to_plot =

              c("CD", "HLA", "IgD", "IL", "TNF",

               "TGF", "GR", "IFNa", "MCP", "MIP"),

            manual_colors =

              c("darkorchid4", "darkorange", "darkgreen"),

            out_dir = norm_dir,

            plot_name = "Marker_distribution_across_batches.pdf")
# Extract cell frequency and MSI

files_list <- list("after" = files_after_norm, "before" = files_before_norm)

results <- list()

for (name in names(files_list)){

  files <- files_list[[name]]

  mx <- extract_pctgs_msi_per_flowsom(files = files,

                   nCells = 50000,

                   phenotyping_markers =

                     c("CD", "HLA", "IgD"),

                   functional_markers =

                     c("MIP", "MCP", "IL",

                       "IFNa", "TNF", "TGF",

                       "Gr"),

                   xdim = 10, ydim = 10, n_metaclusters = 35,

                   out_dir = norm_dir,

                   arcsine_transform = TRUE,

                   save_matrix = TRUE, file_name = name,

                   seed = 343, impute_0_values = TRUE)

  results[[name]] <- mx

}
# create the list to store the plots

plots <- list()

for (name in names(results[[1]])){

  df_plot <- prepare_data_for_plotting(frequency_msi_list = results,

                   matrix_type = name,

                   n_neighbours = 11, seed = 35)

batch <- stringr::str_match(rownames(df_plot), "day[0-9]*")[,1]

samples_id <- ifelse(grepl("REF", rownames(df_plot)),"REF",

          ifelse(grepl("p2", rownames(df_plot)), "p2", "p1"))
```

```
plots[[name]] <- plot_batch_using_freq_msi(df_plot = df_plot, fill = batch,

                    shape = samples_id, color = batch,

                    split_by_normalization = TRUE,

                    title = name,

                    manual_colors =

                      c("darkorchid4", "darkorange", "darkgreen"))
}

gg_a <- grid_arrange_common_legend(plot_lists = plots, nrow = 2, ncol = 2,

                    position = "right")
ggplot2::ggsave(filename ="batch_effect_frequency_MSI.png",

      device = "png",path = norm_dir,plot = gg_a,

      units = "cm",width = 22, height = 14, dpi = 300)
```

Note: This code generates plots in the *CytofNorm* folder, which can be seen in Figures 5B, 5C and 5D.

LIMITATIONS

For the purpose of time this protocol recommends the use of an automated cell counter and a qualified technician to perform all the staining. Without these assumptions met, the protocols may take longer.

The blood samples are fixed upon drawing or stimulation, thus, fixation sensitive epitopes will be lost during the stabilization process. It is then recommended to screen for antibodies compatible with the PROT fixation method as in (Rybakowska et al., 2021a). Alternatively, antigens sensitive to fixation can be stained in the whole blood just before the fixation (see section Potential Solution 2). Of note, no post thawing stimulations can be performed as cells are already fixed.

This protocol is limited to whole blood studies, however small adaptations in the preservation protocol could be introduced and PBMC or mouse splenocytes could be successfully studied. In this case, the analysis pipeline will require some modification in the gating of intact cells. With PBMC, more samples could be processed each day as a smaller amount of cells is needed to analyze minor cell populations.

Working with whole blood requires long time CyTOF/Helios acquisition (10^6 cells/1 h). In order to analyze minor cell populations such as dendritic cells, a long-time access to the core facility and CyTOF is required. It should be noted that sample acquisition can be also split into 2 up to 3 days. It should consider that, if a lot of experiments will be acquired, the acquisition process can take up to months.

The samples were acquired on a CyTOF2 upgraded to HELIOS with the NB injector using water as carrier. Hence, to avoid long exposure to water and subsequent labeling loss but also large fcs file generation, they were acquired in aliquots. Therefore, the above analysis pipeline is prepared to process multiple aliquots per experiment. The script can be adjusted if one big file was generated.

The analysis protocol is an R-based code that requires basic knowledge in R language and programming. However, in this protocol we present a step-by-step analysis pipeline with detailed description of each step, thus researchers new to the R environment can learn and execute the code.

TROUBLESHOOTING

Problem 1

Cell aggregation/clumps are formed during the thawing process (see step 3), affecting proper cell lysis.

Potential solution

Immediately after thawing is completed, add the lysis buffer, pipet vigorously up and down and filter using 100 μm strainer. Add an additional lysis step using 10 mL of lysis buffer, incubate for another 10 min and centrifuge 800 g 5 min 4°C.

Problem 2

Some of the surface antibody clones do not work with the fixation, see Figure 4.

Potential solution

If phenotyping experiments are performed, stain the cells for a given marker before the fixation step and add the fixation buffer to the blood afterward. If the stimulation experiment is performed, stain the whole blood just before the incubation time is over.

Problem 3

Reference sample does not express all the measured markers (see section prepare reference sample, stimulate and preserve).

Potential solution

The reference sample can be stimulated with various simulation agents together or separately and joined just before fixed blood aliquoting, as long as the stimulation protocol is similar to the one used in the study cohort. Or using an alternative stimulation e.g., anti-CD3/CD28.

Problem 4

Unspecific antibody staining is observed in the eosinophil populations (conjugate and titrate antibody and confirm clone fixation resistance).

Potential solution

Block the whole blood with 100 U/mL of heparin in CSB for 20 min at RT just before and during the surface and intracellular antibody staining as described in (Burns et al., 2021; Rahman et al., 2016).

Problem 5

Sample acquisition could not be finished in a single day (sample acquisition in a CyTOF device).

Potential solution

Keep the sample at 4°C and follow with the acquisition the next day. Only the sample that was not previously washed with water or CAS can be stored for no longer than 72 h. If the instrument is out of service for a longer time, freeze unwashed sample at −80°C. To do this spin the cells 800 g, 5 min, and resuspend in 10%DMSO /FBS as shown in (Sumatoh et al., 2017). Alternatively, cells could be frozen directly after Ir staining (data not shown), as shown in https://assets.fluidigm.com/asset/260.

Problem 6

Signal or flowrate cleaning discards too many or too little events, Figure 3.

Potential solution

Adjust the algorithm parameters, following the instruction in the Table 9.

Problem 7

File_quality_check function marks many files as outliers, see step 8 in quantification and statistical analysis.

Potential solution

Check if the proper data transformation is applied if arcsine_tranform parameter is set to FALSE or TRUE. `File_quality_check` is a wrapper constructed from 3 different functions. Run each function separately to narrow down the problem. Verify the FlowSOM tree and its structure.

Problem 8

Samples are not normalized correctly when using the reference sample, see step 12 in quantification and statistical analysis.

Potential solution

Adjust parameters in QuantileNorm train function, changing quantilesValues or the number of quantiles nQ. Build de model using FlowSOM clustering as described in (Gassen et al., 2020).

Problem 9

Cannot load the files in a particular step.

Potential solution

Make sure that none of the steps was skipped or that proper naming of the file (suffix/prefix) is called.

Problem 10

Warnings are printed while using plot_marker_quantiles, see step 13 in quantification and statistical analysis.

```
"No match, please make sure that files are in the same order or define uncommon_prefix parameter"

or

''Perfect match but wrong order. Please order the files or define uncommon_prefix parameter''
```

Potential solution

These warnings will happen if files are either not ordered, or their names do not match. Make sure that the parameter uncommon_prefix is correctly defined. Check if after removing the uncommon prefix/suffix files are ordered and match. If yes, the warning can be ignored. If no, please select the correct files and order them.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Concepción Marañón (concepcion.maranon@genyo.es).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The protocol includes all the datasets generated during this study. The data can be downloaded from https://flowrepository.org/ using the accession number FlowRepository: FR-FCM-Z3YR.

The R package CytoQP and analysis script (CytoQP_script.R) can be accessed on GitHub: https://github.com/prybakowska/CytoQP and the code is provided in this manuscript and as a

supplementary R script called, Data S1 CytoQP_script.R. The original code has been deposited in Zenodo and is publicly available as of the date of publication, Zenodo: https://doi.org/10.5281/zenodo.6947414. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xpro.2022.101697.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Conceptualization, P.R. and C.M.; Methodology, P.R., C.M., and S.V.G.; Sample collection, P.R.; Investigation, P.R., S.V.G., J.M.M., and K.Q.; Resources, M.E.A. and C.M.; Writing – Original Draft, P.R. and C.M.; Writing – Review & Editing, C.M., M.E.A., S.V.G., K.Q., and Y.S.; Visualization, P.R. and S.V.G.; Supervision, Project Administration, and Funding Acquisition, C.M. and M.E.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Amir, E.A.D., Guo, X.V., Mayovska, O., and Rahman, A.H. (2018). Average overlap frequency: a simple metric to evaluate staining quality and community identification in high dimensional mass cytometry experiments. J. Immunol. Methods 453, 20–29. https://doi.org/10.1016/j.jim.2017.08.011.

Amir, E.A.D., Lee, B., Badoual, P., Gordon, M., Guo, X.V., Merad, M., and Rahman, A.H. (2019). Development of a comprehensive antibody staining database using a standardized analytics pipeline. Front. Immunol. 10, 1315. https://doi.org/10.3389/fimmu.2019.01315.

Auguie, B., and Antonov, A. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics (CRAN).

Bengtsson, H. (2021). R.utils: Various Programming Utilities (CRAN). https://github.com/HenrikBengtsson/R.utils.

Burns, M., Ostendorf, L., Biesen, R., Grützkau, A., Hiepe, F., Mei, H.E., and Alexander, T. (2021). Dysregulated CD38 expression on peripheral blood immune cell subsets in SLE. Int. J. Mol. Sci. 22, 2424. https://doi.org/10.3390/ijms22052424.

Crowell, H.L., Zanotelli, V.R.T., Chevrier, S., Robinson, M.D., and Bodenmiller, B. (2022). Catalyst: Cytometry dATa anALYSis Tools. Bioconductor version: Release (3.14). https://github.com/HelenaLC/CATALYST.

Ellis, B., Haal, P., Hahne, F., Meur, N.L., Gopalakrishnan, N., Spidlen, J., Jiang, M., Finak, G., and Granjeaud, S. (2022). flowCore: flowCore: Basic Structures for Flow Cytometry Data. Bioconductor version: Release (3.14).

Emmaneel, A., Quintelier, K., Sichien, D., Rybakowska, P., Marañón, C., Alarcón-Riquelme, M.E., Van Isterdael, G., Van Gassen, S., and Saeys, Y. (2022). PeacoQC: peak-based selection of high quality cytometry data. Cytometry A 101, 325–338. https://doi.org/10.1002/cyto.a.24501.

Finck, R., Simonds, E.F., Jager, A., Krishnaswamy, S., Sachs, K., Fantl, W., Pe'er, D., Nolan, G.P., and Bendall, S.C. (2013). Normalization of mass cytometry data with bead standards. Cytometry A 83, 483–494. https://doi.org/10.1002/cyto.a.22271.

Gassen, S.V., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, 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.

Gassen, S.V., Gaudilliere, B., Angst, M.S., Saeys, Y., and Aghaeepour, N. (2020). CytoNorm: a normalization algorithm for cytometry data. Cytometry A 97, 268–278. https://doi.org/10.1002/cyto.a.23904.

Hartmann, F.J., Simonds, E.F., and Bendall, S.C. (2018). A universal live cell barcoding-platform for multiplexed human single cell analysis. Sci. Rep. 8, 10770. https://doi.org/10.1038/s41598-018-28791-2.

Hartmann, F.J., Simonds, E.F., Vivanco, N., Bruce, T., Borges, L., Nolan, G.P., Spitzer, M.H., and Bendall, S.C. (2019). Scalable conjugation and characterization of immunoglobulins with stable mass isotope reporters for single-cell mass cytometry analysis. Methods Mol. Biol. 1989, 55–81. https://doi.org/10.1007/978-1-4939-9454-0_5.

Jamin, C., Le Lann, L., Alvarez-Errico, D., Barbarroja, N., Cantaert, T., Ducreux, J., Dufour, A.M., Gerl, V., Kniesch, K., Neves, E., et al. (2016). Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. Autoimmun. Rev. 15, 1038–1045. https://doi.org/10.1016/j.autrev.2016.07.034.

Kassambara, A. (2020). ggpubr: "ggplot2" Based Publication Ready Plots (CRAN). https://github.com/kassambara/ggpubr.

Kolde, R. (2019). pheatmap: Pretty Heatmaps (CRAN).

Lai, L., Ong, R., Li, J., and Albani, S. (2015). A CD45-based barcoding approach to multiplex mass-cytometry (CyTOF). Cytometry A 87, 369–374. https://doi.org/10.1002/cyto.a.22640.

Malek, M., and Taghiyar, M.J. (2022). flowDensity: Sequential Flow Cytometry Data Gating. Bioconductor version: Release (3.14).

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2015). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw 3. https://doi.org/10.21105/joss.00861.

Mei, H.E., Leipold, M.D., Schulz, A.R., Chester, C., and Maecker, H.T. (2015). Barcoding of live human PBMC for multiplexed mass cytometry. J. Immunol. 194, 2022–2031. https://doi.org/10.4049/jimmunol.1402661.

Melville, J., Lun, A., Djekidel, M.N., and Hao, Y. (2021). uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction (CRAN). https://github.com/jlmelville/uwot.

Meskas, J., and Wang, S. (2022). flowCut: Precise and Accurate Automated Removal of Outlier Events and Flagging of Files Based on Time Versus Fluorescence Analysis. Bioconductor version: Release (3.14).

Meskas, J., Yokosawa, D., Wang, S., Segat, G., and Brinkman, R. (2022). flowCut: An R package for automated removal of outlier events and flagging of files based on time versus fluorescence analysis. (Cytometry A), In press. https://doi.org/10.1002/cyto.a.24670.

Monaco, G., Chen, H., Poidinger, M., Chen, J., de Magalhães, J.P., and Larbi, A. (2016). flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. Bioinformatics 32, 2473–2480. https://doi.org/10.1093/bioinformatics/btw191.

Morgan, M., Obenchain, V., Hester, J., and Pagès, H. (2022a). SummarizedExperiment: SummarizedExperiment Container. Bioconductor version: Release (3.14). https://bioconductor.org/packages/SummarizedExperiment.

Morgan, M., Wang, J., Obenchain, V., Lang, M., Thompson, R., and Turaga, N. (2022b). BiocParallel: Bioconductor facilities for parallel evaluation (R package version 1.30.3). https://github.com/Bioconductor/BiocParallel.

Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palettes (CRAN).

Olsen, L.R., Leipold, M.D., Pedersen, C.B., and Maecker, H.T. (2019). The anatomy of single cell mass cytometry data. Cytometry A 95, 156–172. https://doi.org/10.1002/cyto.a.23621.

Pagès, H., Lawrence, M., and Aboyoun, P. (2022). S4Vectors: Foundation of Vector-like and List-like Containers in Bioconductor (https://bioconductor.org/packages/S4Vectors). Bioconductor version: Release (3.14).

R Core Team (2013). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Rahman, A.H., Tordesillas, L., and Berin, M.C. (2016). Heparin reduces nonspecific eosinophil staining artifacts in mass cytometry experiments. Cytometry A 89, 601–607. https://doi.org/10.1002/cyto.a.22826.

Rybakowska, P., Alarcón-Riquelme, M.E., and Marañón, C. (2020). Key steps and methods in the experimental design and data analysis of highly multi-parametric flow and mass cytometry. Comput. Struct. Biotechnol. J. 18, 874–886. https://doi.org/10.1016/j.csbj.2020.03.024.

Rybakowska, P., Burbano, C., Van Gassen, S., Varela, N., Aguilar-Quesada, R., Saeys, Y., Alarcón-Riquelme, M.E., and Marañón, C. (2021a). Stabilization of human whole blood samples for multicenter and retrospective immunophenotyping studies. Cytometry A 99, 524–537. https://doi.org/10.1002/cyto.a.24241.

Rybakowska, P., Van Gassen, S., Quintelier, K., Saeys, Y., Alarcón-Riquelme, M.E., and Marañón, C. (2021b). Data processing workflow for large-scale immune monitoring studies by mass cytometry. Comput. Struct. Biotechnol. J. 19, 3160–3175. https://doi.org/10.1016/j.csbj.2021.05.032.

Schulz, A.R., Baumgart, S., Schulze, J., Urbicht, M., Grützkau, A., and Mei, H.E. (2019). Stabilizing antibody cocktails for mass cytometry. Cytometry A 95, 910–916. https://doi.org/10.1002/cyto.a.23781.

Spidlen, J., Moore, W., Parks, D., Goldberg, M., Bray, C., Bierre, P., Gorombey, P., Hyun, B., Hubbard, A., Lange, S., et al. (2010). Data File Standard for Flow Cytometry, Version FCS 3.1. Cytometry A 77, 97–100. https://doi.org/10.1002/cyto.a.20825.

Sumatoh, H.R., Teng, K.W.W., Cheng, Y., and Newell, E.W. (2017). Optimization of mass cytometry sample cryopreservation after staining. Cytometry A 91, 48–61. https://doi.org/10.1002/cyto.a.23014.

Thrash, E.M., Kleinsteuber, K., Hathaway, E.S., Nazzaro, M., Haas, E., Hodi, F.S., and Severgnini, M. (2020). High-throughput mass cytometry staining for immunophenotyping clinical samples. STAR Protoc. 100055. https://doi.org/10.1016/j.xpro.2020.100055.

Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data: FlowSOM. Cytometry A 87, 636–645. https://doi.org/10.1002/cyto.a.22625.

Wickham, H., and RStudio. (2019). stringr: Simple, Consistent Wrappers for Common String Operations (CRAN). https://github.com/tidyverse/stringr.

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., and RStudio. (2021). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics (CRAN). https://github.com/tidyverse/ggplot2.

Wickham, H., François, R., Henry, L., Müller, K., and RStudio. (2022). dplyr: A Grammar of Data Manipulation (CRAN). https://github.com/tidyverse/dplyr.

Yan, L., Ma, C., Wang, D., Hu, Q., Qin, M., Conroy, J.M., Sucheston, L.E., Ambrosone, C.B., Johnson, C.S., Wang, J., and Liu, S. (2012). OSAT: a tool for sample-to-batch allocations in genomics experiments. BMC Genom. 13, 689. https://doi.org/10.1186/1471-2164-13-689.

Zunder, E.R., Finck, R., Behbehani, G.K., Amir, E.-A.D., Krishnaswamy, S., Gonzalez, V.D., Lorang, C.G., Bjornson, Z., Spitzer, M.H., Bodenmiller, B., et al. (2015). Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. Nat. Protoc. 10, 316–333. https://doi.org/10.1038/nprot.2015.020.