



HAL
open science

Elaboration d'une procédure standardisée d'harmonisation des données de cytométrie en flux dans le cadre d'études multicentriques

Lucas Le Lann

► **To cite this version:**

Lucas Le Lann. Elaboration d'une procédure standardisée d'harmonisation des données de cytométrie en flux dans le cadre d'études multicentriques. Médecine humaine et pathologie. Université de Bretagne occidentale - Brest, 2019. Français. NNT : 2019BRES0048 . tel-02890387

HAL Id: tel-02890387

<https://theses.hal.science/tel-02890387>

Submitted on 6 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat de

L'UNIVERSITE
DE BRETAGNE OCCIDENTALE
Comue Université Bretagne Loire

Ecole Doctorale n° 605
Biologie Santé
Spécialité : Immunologie

Par

Lucas LE LANN

Elaboration d'une procédure standardisée d'harmonisation des données de cytométrie en flux dans le cadre d'études multicentriques

Thèse présentée et soutenue à Brest, le 26/09/2019
Unité de recherche : UMR 1227, Lymphocyte B et Auto-immunité

Rapporteurs avant soutenance :

Michelle ROSENZWAJG
Guillaume MONNERET

MCU-PH, UMR S 959, Sorbonne université, Paris
PU-PH, EA 7426, Université Claude Bernard Lyon 1, Lyon

Composition du Jury :

Président : Michelle ROSENZWAJG
Examineurs : Guillaume MONNERET

MCU-PH, UMR S 959, Sorbonne université, Paris
PU-PH, EA 7426, Université Claude Bernard Lyon 1, Lyon

Dir. de thèse : Jacques-Olivier PERS
Co-dir. de thèse : Christophe JAMIN

PU-PH, UMR 1227, Université de Bretagne Occidentale, Brest
MCU-PH, UMR 1227, Université de Bretagne Occidentale, Brest

Table des matières

1	INTRODUCTION	5
1.1	Cytomètre en flux	5
1.1.1	Introduction à la cytométrie en flux	5
1.1.2	Fluorochromes	5
1.1.3	Cytomètre	8
1.1.4	Compensation	12
1.1.5	Fichiers de cytométrie en flux	15
1.2	Analyse des données de cytométrie en flux	18
1.2.1	Analyse manuelle	18
1.2.2	Outils d'analyse automatisée	20
1.2.3	Analyses automatisées non-supervisées	21
1.2.4	Analyses automatisées supervisées	23
1.2.5	Analyses multi-centriques	24
1.2.6	Réponses aux problématiques multi-centriques, analyse de la littérature existante	24
1.3	Contexte de l'étude	30
1.3.1	Populations cellulaires	30
1.3.2	Maladies auto-immune systémiques	32
1.3.3	le projet PRECISESADS	33
1.3.4	La cytométrie en flux dans PRECISESADS	34
1.3.5	Combinaisons de fluorochromes	35
1.3.6	Analyses dans le cadre de PRECISESADS	36
1.3.7	Questions soulevées durant la thèse	40
2	RÉSULTATS	41
2.1	Préparation des données	41
2.2	Calibration	41
2.2.1	Harmonisation	41
2.2.2	Calibration au cours du temps	44
2.2.3	Billes 8 pics	45
2.3	Normalisation	46
2.3.1	Normalisation existante	46
2.3.2	Normalisation Intra-Centre	47
2.3.3	Validation de la normalisation	48
2.4	automatisation de la normalisation	52
2.4.1	clustering 2D et 1D	53
2.4.2	Script de Normalisation	53
2.5	Analyse des données	55
2.5.1	Compensation	55
2.5.2	Analyse manuelle	56
2.5.3	Observation des résultats de 300 patients	57
2.5.4	Comprendre l'effet centre	60
2.6	Automatisation de l'analyse par Altrabio	66
2.6.1	Validation des automates	68

2.6.2	Fichiers aberrants	70
2.6.3	Découplage de fluorochromes	72
2.6.4	Observation de l'effet lot et de l'effet centre	75
2.7	Effet lot	76
2.7.1	Calcul du coefficient lot	76
2.7.2	Observation de l'effet centre après correction de l'effet lot . . .	77
2.8	Effet centre	78
2.8.1	Calcul du coefficient de l'effet centre	79
2.8.2	Calcul du coefficient inter-center version 2	81
3	DISCUSSION	85
3.1	Principaux résultats de mon travail	85
3.2	Les méthodes de calibration	86
3.3	Mise en place d'un essai multi-centrique prospectif	86
3.3.1	Les points à considérer	86
3.3.2	Les scripts développés pour palier à chacun des aspects critiques	89
3.4	Conclusion générale	91
A	Annexe	98
A.1	Cytométrie report	98
A.2	Effet Lot et correction	100
A.3	Article en soumission	111
A.4	Article Harmonisation	129
A.5	Article PRECISESADS	137
A.6	Standard operating procedure	154
A.7	SCRIPTS	258
A.7.1	Script R de normalisation intra-centre	258
A.7.2	Script Python de correction de l'effet lot	261
A.7.3	Script Python de correction de l'effet centre	271

Acronymes

CD Cluster de différenciation

LED Lupus érythémateux disséminé

ScS Sclérodémie systémique

PR Polyarthrite rhumatoïde

SGS Syndrome de Gougerot-Sjögren

SAPL Syndrome des anti-phospholipides

MCTD Connectivite mixtes / Mixed Connective Tissue Disease

UCTD Connectivite indifférenciée/ Undifferentiated Connective Tissue Disease

SOP Procédure Opérationnelle Standard / Standard Operating Procedure

fcs flow cytometry standard (format de fichier)

PMT Tube Photo-multiplicateur

IMF Intensité médiane de fluorescence

CV Coefficient de Variation

FITC fluorescein isothiocyanate

PE Phycoerythrine

PC5.5 Phycoerythrine Cyanine 5.5

PC7 Phycoerythrine Cyanine 7

APC Allophycocyanine

APC-AF750 Allophycocyanine Alexa Fluor 750

PB Pacific Blue

KO Krome Orange

SADs Maladies autoimmunes systémiques / Systemic Autoimmune Diseases

ACP Analyse de composantes principales

PBMC Cellules mononuclées du sang périphérique / Preipheral blood Mononuclear Cells

Bit Unité de mesure encodé en une succession de 0 et de 1

Byte Formé de 8 bits permettant de passer d'une série de 0 et de 1 à du text

1 INTRODUCTION

1.1 Cytomètre en flux

1.1.1 Introduction à la cytométrie en flux

La cytométrie en flux est une technique employée dans les laboratoires d'analyse médicale que cela soit lors du diagnostic ou du suivi des patients [1]. C'est une méthode relativement rapide pour l'analyse des cellules sanguines et qui a pour avantage d'être non invasive. La puissance de la cytométrie vient de sa capacité de caractériser individuellement des milliers de cellules en suspension d'un échantillon. Pour cela, les cellules doivent être préalablement marquées à l'aide d'anticorps couplés à des fluorochromes.

1.1.2 Fluorochromes

Les fluorochromes sont des substances chimiques qui peuvent absorber l'énergie des photons provenant de sources lumineuses externes. Une fois l'énergie absorbée, le fluorochrome est dans un état dit "excité". Pour retourner à son état basal, le fluorochrome va émettre un photon.

L'excitation d'un fluorochrome atteint un pic optimal à certaines longueurs d'ondes. Pour atteindre ce pic, l'excitation se fait souvent avec l'aide de lasers permettant d'atteindre ces longueurs d'ondes précises. L'émission, par le fluorochrome, d'un signal lumineux couvre une partie du spectre lumineux mais atteint un pic d'intensité à une longueur d'onde donnée. Ce sont ces longueurs d'ondes des pics d'excitation et d'émission qui caractérisent les fluorochromes (*Figure 1*).

Par exemple, l'isothiocyanate de fluorescéine (FITC) peut être excitée par un laser émettant à 488nm, ce qui est proche de son pic d'excitation qui est de 495nm. Suite à cette excitation, le fluorochrome va émettre à son tour de la lumière autour de 521nm, son pic d'émission. Similairement, la R Phycoerythrin (PE) peut être excitée par le même laser à 488nm car son pic d'excitation est proche à 480nm. Mais son pic d'émission sera différent de la FITC à 578nm.

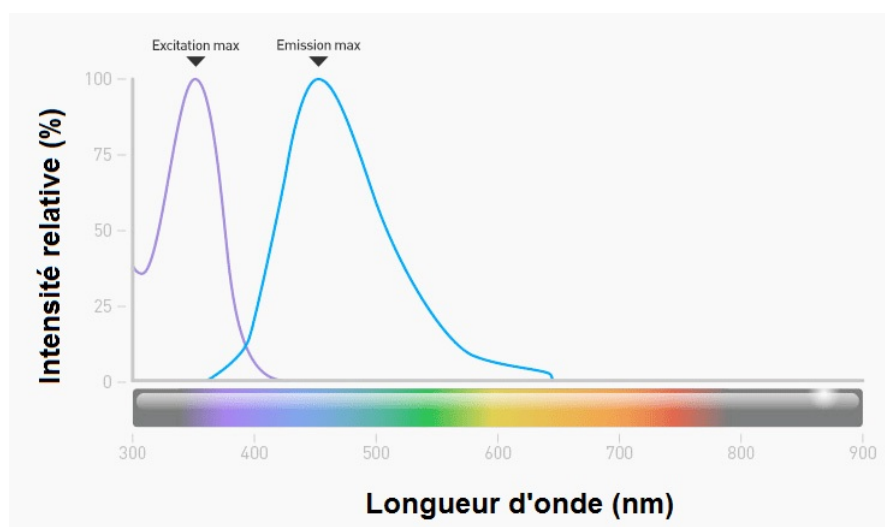


FIGURE 1: **Illustration du spectre d'excitation et d'émission.** Exemple d'un spectre d'excitation(violet) et d'émission(bleu) d'un fluorochrome. Source : d'après (<https://www.thermofisher.com/fr/fr/home/life-science/cell-analysis/cell-analysis-learning-center/molecular-probes-school-of-fluorescence/imaging-basics/fundamentals-of-fluorescence-microscopy/physical-properties-that-define-fluorescence.html>).

Grâce à ces différences dans les pics d'émission, il est possible de combiner plusieurs fluorochromes pouvant être excités par le même laser et éviter un chevauchement trop important dans leurs spectres d'émission respectifs (*Tableau 1*).

Fluorochromes	Abréviation	Excitation (nm)	Émission (nm)
Isothiocyanate de Fluorescéine	FITC	488	525
R Phycoerythrine	PE	488	575
R Phycoerythrine -Cyanine 5.5	PC5.5	488	692
R Phycoerythrine -Cyanine 7	PC7	488	770
Allophycocyanine	APC	633/635	660
Allophycocyanine -AlexaFluor750	APC-AF750	633/638	775
Pacific Blue	PB	405	455
Krome Orange	KO	405	528

TABLEAU 1: Description des différents fluorochromes qui seront utilisés lors de l'étude PRECISESADS avec leur abréviations ainsi que leurs pics d'excitation et d'émission.

L'intérêt des fluorochromes en cytométrie apparaît lorsque ceux-ci sont liés à des anticorps car ces derniers ont la capacité de se fixer spécifiquement à une cible, telle que les protéines membranaires. Ainsi, il est possible de marquer la cellule avec des anticorps associés à des fluorochromes (*Figure 2*). Ce marquage se faisant de manière spécifique, l'absence d'émission de lumière par le fluorochrome indiquera une absence de la protéine.

L'avantage de cette spécificité est la capacité de pouvoir marquer plusieurs protéines simultanément. Pour ce faire, il est nécessaire d'avoir des anticorps dirigés contre des cibles différentes et de les lier à des fluorochromes dont le spectre d'émission est suffisamment différent. Pour élargir la gamme des pics d'émission disponibles, tout en gardant les mêmes lasers, il existe des fluorochromes dit en tandem. La particularité des fluorochromes en tandem est leur capacité de transférer de l'énergie entre des molécules fluorescentes ou Förster resonance energy transfer (FRET). Lorsque les 2 fluorochromes sont couplés, un premier fluorochrome va pouvoir être excité par le laser, puis par le signal de FRET, il va pouvoir exciter le deuxième fluorochrome, qui va émettre aux longueurs d'ondes qui lui sont propres. Les fluorochromes et leurs tandems fournissent un plus large éventail de combinaisons fluorochromes/anticorps utilisable. Cet éventail est très utile pour l'identification approfondie des populations et des sous populations cellulaires en permettant d'observer simultanément la présence de 20 marqueurs pour certaines machines [2].

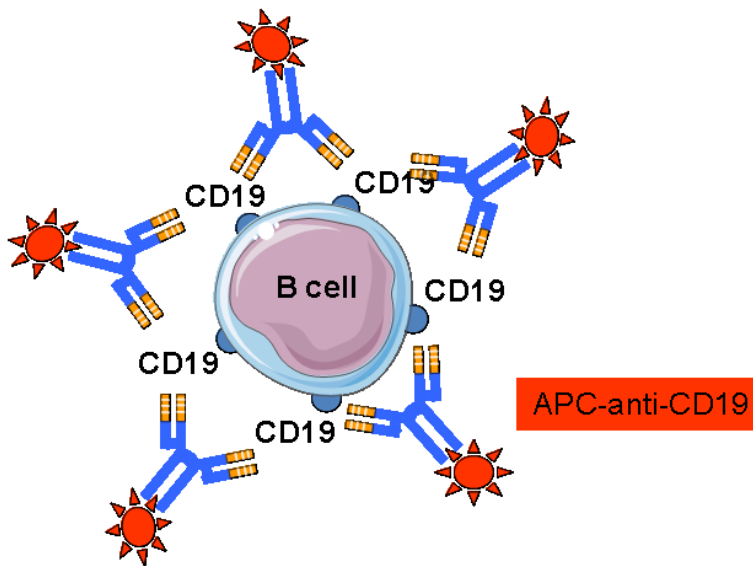


FIGURE 2: **Marquage de cytométrie.** Illustration du marquage d'une cellule (Lymphocyte B) par des anticorps anti-CD19 couplés au fluorochrome APC.

La préparation d'un échantillon, ou marquage, permet la fixation des complexes anticorps/fluorochromes aux cellules. Lors de cette étape l'échantillon de sang est mis en contact avec un cocktail d'anticorps/fluorochromes préalablement sélectionné puis incubé pour permettre la fixation sur les cellules. L'étape suivante est la lyse des globules rouges et la fixation des autres cellules. Cette fixation va permettre d'effectuer l'analyse des heures après le marquage sans perte du signal. Une fois ces différentes étapes réalisées, l'échantillon peut être acquis au cytomètre en flux.

1.1.3 Cytomètre

Le cytomètre en flux est une machine permettant d'analyser plusieurs paramètres simultanément sur des milliers de cellules en suspension par seconde. Le fonctionnement d'un cytomètre en flux peut être séparé en 3 parties : fluïdique, optique et électronique (*Figure 3*).

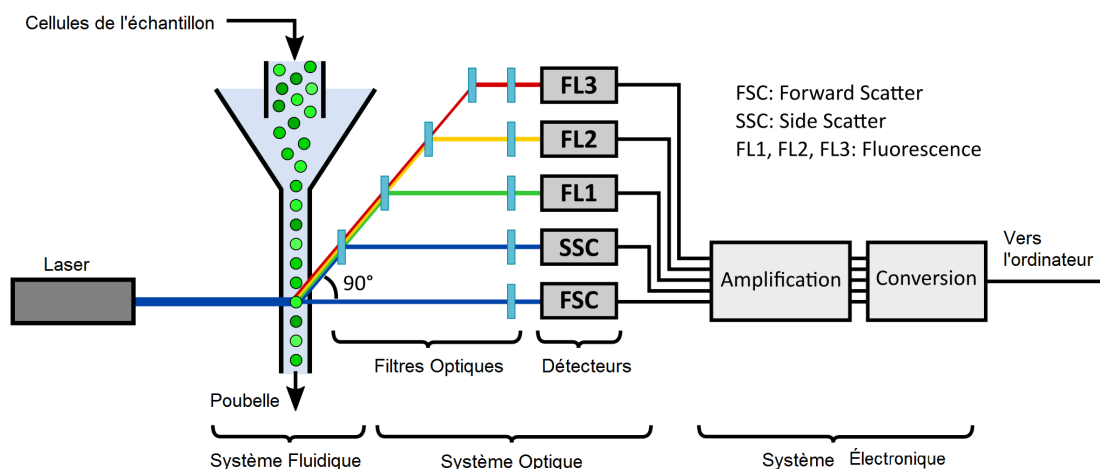


FIGURE 3: **Cytomètre en flux**. Schéma du fonctionnement d'un cytomètre en flux. Source d'après (<https://benchling.com/pub/tabor-flowcal>).

PARTIE FLUIDIQUE

La première étape dans un cytomètre en flux est la séparation des cellules de l'échantillon afin de les présenter individuellement devant les lasers. Pour cela, un jeu de pression permet d'aspirer l'échantillon de sang du tube de cytométrie. Puis cet échantillon va être dirigé, par l'intermédiaire d'une gaine fluidique, en une file unique de cellules qui seront alignées avec le rayon laser (*Figure 4*).

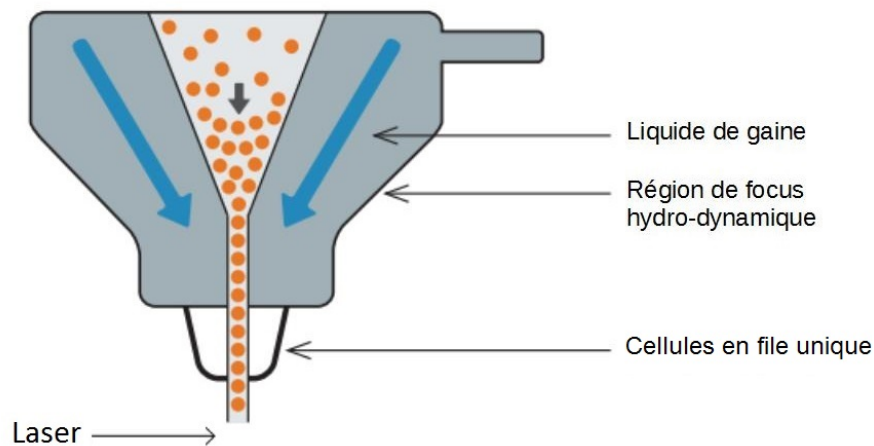


FIGURE 4: **Fluidique.** Schéma du fonctionnement de la partie fluidique d'un cytomètre en flux. Source d'après (<https://www.bio-rad-antibodies.com/flow-cytometry-fluidics-system.html>).

Des modifications dans la pression de la fluidique permet d'accélérer l'acquisition de l'échantillon, mais ce nombre plus important de cellules passées par seconde arrive au prix d'une moins bonne résolution. Une fois la cellule présentée devant le rayon du laser, la partie OPTIQUE de la machine entre en jeu.

PARTIE OPTIQUE

Cette étape de l'analyse consiste en l'excitation des fluorochromes, en la détection du signal d'émission et en la transformation de ce dernier en un signal électrique. Lorsque le rayon laser frappe la cellule, les premières informations recueillies sont la taille et la structure relative des cellules. Puis les spectres émis suite à l'excitation des fluorchromes présents à la surface de la cellule sont captés.

Concernant la détermination de la taille des cellules, le cytomètre utilise la dispersion du rayon laser sur la surface de la cellule. Cette dispersion est détectée dans l'axe du laser (appelé le Foward Scatter ou FS). Pour la structure, la détection du signal se fait à un angle de 90 degré (appelé Side Scatter ou SS). À cette angle, le signal obtenu résulte de la diffraction du laser sur les organites contenus dans la cellule.

Ces 2 paramètres permettent de faire une première séparation morphologique des populations cellulaires. En effet, les différentes populations du système immunitaire n'ont pas toutes la même taille, ni la même structure. Ces différences permettent de caractériser les trois grandes populations de leukocytes (*Figure 5*) : les granulocytes (15-30 μm) , les monocytes (10-15 μm) et les lymphocytes (6-14 μm).

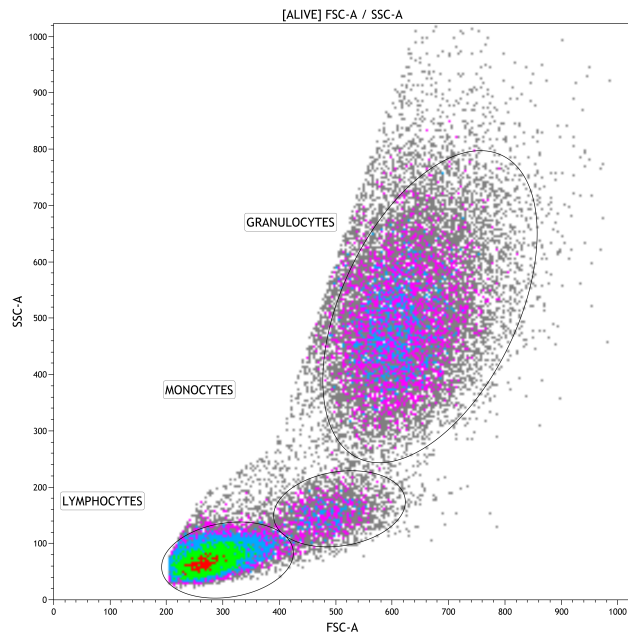


FIGURE 5: **Taille/structure.** Exemple de représentation biparamétrique avec la taille(FSC-A) et la structure(SSC-A). Les grandes populations immunitaires sont entourées et nommées.

Concernant les fluorochromes excités par le laser, le signal qu'ils émettent est filtré et dirigé vers les différents détecteurs. Cette filtration du signal est faite par une succession de miroirs dichroïques (*Figure 6*).

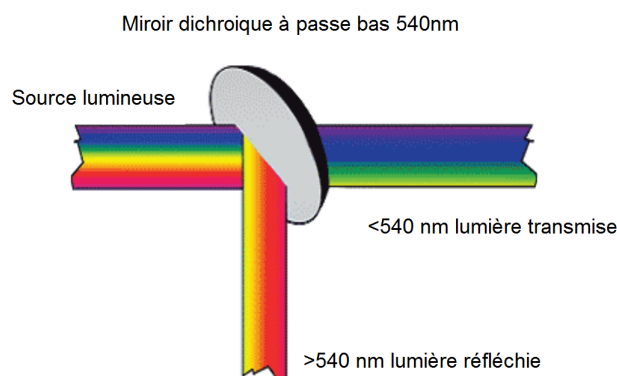


FIGURE 6: **Miroir dichroïque.** Illustration du fonctionnement d'un miroir dichroïque à faible bande passante. C'est à dire qu'il laisse passer les longueurs d'ondes inférieures à 540nm et renvoie les autres. Source d'après (<https://www.bio-rad-antibodies.com/flow-cytometry-optics-detection.html>).

Un miroir dichroïque est un filtre ne laissant passer qu'une partie du signal lumineux. Cette division peut se faire au dessus ou en dessous d'une certaine longueur d'onde. En plus de cela, ce filtre est recouvert d'un revêtement réfléchissant permettant de récupérer le reste du signal. La succession de ces miroirs dichroïques permet

de séparer les pics d'émission des différents fluorochromes pour les diriger vers différents tubes photomultiplicateurs en un seul passage.

Le tube photomultiplicateur (PMT) est capable de transformer un signal lumineux (photons) en un signal électrique (électrons). Cette transformation commence par l'arrivée d'un photon sur la photocathode qui va le convertir en un électron. Cette électron va ensuite être dirigé vers une série de dynodes. Ces dynodes vont permettre d'amplifier plus ou moins le signal électrique par la génération d'électrons secondaires en fonction du voltage appliqué au PMT (*Figure 7*).

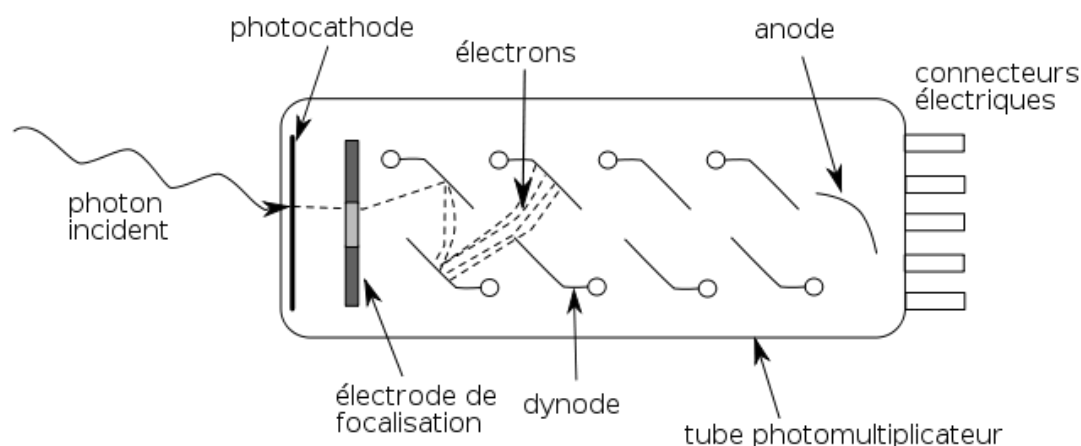


FIGURE 7: **Tube Photo-Multiplicateur.** Schéma du fonctionnement d'un tube photomultiplicateur(PMT) lors de l'arrivée d'un photon. Source d'après(Philippe Després)

Bien que les PMT permettent d'amplifier des signaux très faibles, ils ne possèdent pas tous la même sensibilité à certaines longueurs d'ondes à cause de l'utilisation de différents matériaux pour leur confection. Cette sensibilité est aussi dépendante du voltage appliqué au PMT pour l'amplification d'électrons. Ces deux aspects ont un impact sur l'intensité du signal électrique final.

PARTIE ÉLECTRONIQUE

Malgré l'amplification causée par le voltage sur les PMT, il est nécessaire d'amplifier le signal électrique par l'intermédiaire d'un gain puis de le convertir en un signal numérique.

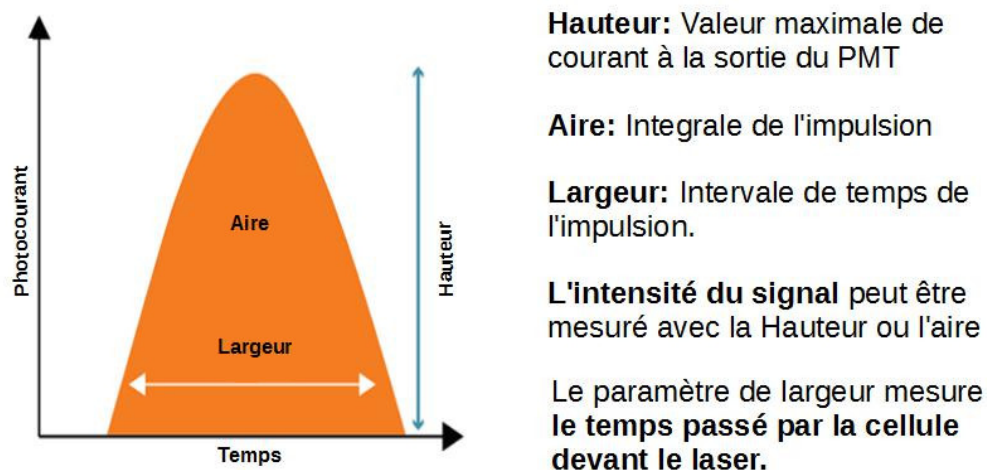


FIGURE 8: **Électronique.** Schéma du fonctionnement de la partie fluïdique d'un cytomètre en flux. Source d'après (<https://www.bio-rad-antibodies.com/flow-cytometry-fluidics-system.html>).

Ce signal numérique est enregistré en bits dont le nombre varie en fonction des fabricants. La valeur maximale est de 18 bits soit 262144 pour les cytomètres provenant de chez Becton Dickinson. Concernant les cytomètres provenant de Beckman Coulter, ceux-ci enregistrent les données avec comme maximum 24 bits soit 1048576.

Après cette conversion, les paramètres de chaque évènement (cellule) sont enregistrés. Il s'agit de l'intensité de fluorescence pour chacun des canaux et des valeurs en FS et SS mesurées comme l'aire du pulse de photocourant formé lorsque la cellule est passée devant le laser (*Figure 8*). Ces données sont stockées dans un format de fichier particulier suivant le standard de cytométrie en flux (Flow Cytometry Standard, FCS).

Toutes ces étapes permettent d'obtenir les informations pour caractériser les différentes cellules de l'échantillon. Mais d'autres modifications doivent encore être apportées au fichier de cytométrie pour pouvoir exploiter les résultats. En effet les différents spectres d'émission des fluorochromes peuvent déborder sur des canaux qui ne leurs sont pas dédiés. Les signaux doivent être compensés afin de s'assurer de la spécificité du signal mesuré.

1.1.4 Compensation

Ce "débordement" des différents fluorochromes dans les canaux voisins est une problématique importante pour l'exploitation des fichiers. En effet, le spectre d'émission de certains fluorochromes peut arriver au niveau des canaux dédiés à d'autres fluorochromes (*Figure 9*). Cette superposition de 2 fluorochromes engendre un faux

signal positif qu'il est nécessaire de corriger (*Figure 10*).

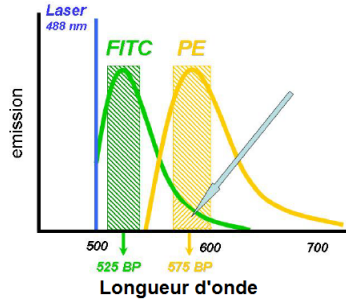


FIGURE 9: **Compensation.** Illustration du phénomène de chevauchement entre 2 fluorochromes, FITC (en vert) et PE (en jaune), tous les deux excités par un laser à 488 nm, leur spectre d'émission se chevauche dans la bande passante dédiée au pic de PE. Source d'après (<http://flow-book.denovosoftware.com/sites/default/files>).

La correction, ou compensation, consiste en la soustraction des signaux parasites provenant de multiples fluorochromes pour n'obtenir que le vrai signal.

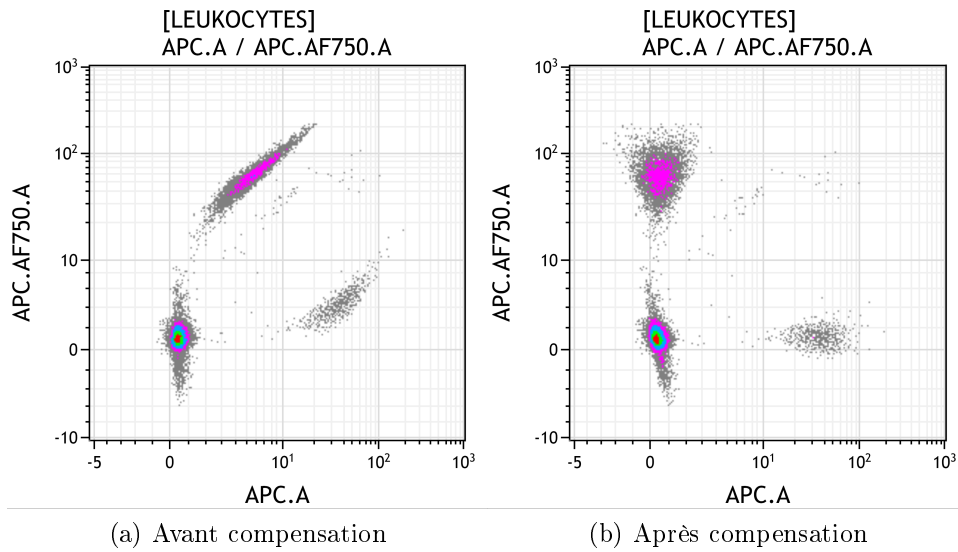


FIGURE 10: **Application de la compensation.** Exemple de l'impact de la compensation sur des populations cellulaires, ici les cellules APC.AF750 positives et les cellules APC positives. ".A" indique que c'est l'aire de l'impulsion électrique qui est utilisé comme valeur pour les intensités de fluorescence.

Par exemple, dans le cas où on utilise 4 canaux (FITC, PE, APC et KO) pour obtenir le vrai signal du canal FITC, on utilise la formule suivante :

$$\text{FITCvrai} = \text{FITCmesuré} \times \text{CompensationFITC} + \text{FITCmesuré} \times \text{CompensationPE} + \text{FITCmesuré} \times \text{CompensationAPC} + \text{FITCmesuré} \times \text{CompensationKO}$$

Différentes méthodes existent pour déterminer les valeurs de compensation nécessaires aux fichiers de cytométrie.

Fluorescence Minus One : Cette méthode consiste en un marquage "classique" de la suspension cellulaire mais avec comme particularité d'enlever un des marqueurs du cocktail de fluorochromes. L'idée derrière cette méthode est par l'absence de ce fluorochrome d'obtenir toutes les valeurs parasites du reste des fluorochromes. Ce marquage doit être réalisé pour chacun des fluorochromes du cocktail.

Kit de compensation : Une autre méthode est d'utiliser un kit de compensation avec des billes de capture. Ces billes permettent de définir une population négative et une population positive en fixant les fluorochromes présents en solution. Ces billes sont couplées avec un seul anticorps possédant chacun 1 seul des fluorochromes. Il y a autant de tube "mono-marqué" qu'il y a d'anticorps dans le cocktail général de marquage.

Une fois le marquage réalisé, les signaux obtenus dans les canaux autres que celui attendu permettent de définir la compensation nécessaire. L'avantage de cette méthode est la stabilité des billes réduisant l'influence biologique et reflète plus précisément la réalité physique des fluorochromes.

Détermination visuelle : La dernière méthode est d'utiliser des outils logiciel comme Kaluza (Beckman Coulter) et de visuellement déterminer la compensation. Pour cela, on réalise toutes les combinaisons bi-paramétriques des différents marqueurs. Ensuite, l'idée est de modifier les valeurs de compensation pour centrer l'IMF des populations négatives autour de 0.

Cette méthode a deux inconvénients, le premier est qu'elle est sensible à l'opérateur, le deuxième est qu'elle nécessite du temps pour déterminer la compensation sur les bi-paramétriques différents. Mais, en contre partie, elle est applicable à tous les fichiers de cytométrie et évite la nécessité de réaliser une des méthodes précédemment évoquée à chacune des inclusions.

Toutes ces méthodes permettent de générer les valeurs pour toutes les combinaisons de fluorochromes qui sont stockées dans une matrice de compensation pour les fichiers de cytométrie (*Tableau 2*).

TABLEAU 2: Exemple de Matrice de compensation

	FITC	PE	APC	KO
FITC	1	-0.002	-0.02	0
PE	0	1	0	0
APC	0	0.6	1	0
KO	0	0.4	0.5	1

Pour l'exploitation des intensités de fluorescence, les valeurs de la matrice doivent être appliquées aux intensités de fluorescence mesurées pour obtenir le vrai signal. Dans l'exemple du tableau 2 :

$$\text{FITCvrai} = \text{FITCmesuré} \times 1 + \text{FITCmesuré} \times -0.002 + \text{FITCmesuré} \times -0.02 + \text{FITCmesuré} \times 0$$

La matrice ainsi que les intensités brutes de fluorescence sont stockées dans un format de fichier particulier à la cytométrie en flux permettant ainsi de modifier la compensation après acquisition.

1.1.5 Fichiers de cytométrie en flux

L'acquisition d'un échantillon par un cytomètre en flux génère des fichiers qui suivent le format "Flow Cytometry Standard" (FCS). Ce format est géré par la société internationale pour l'avancement de la cytométrie (International Society for Advancement of Cytometry ISAC).

Ce format a premièrement été défini comme FCS 1.0 en 1984 [3] et mis à jour en 1990 (FCS 2.0) [4] et en 1997 avec FCS 3.0 [5]. Puis, par la dernière version en 2008 3.1 avec quelques améliorations mineures [6].

Le format FCS est composé de 4 parties : l'entête, le texte, les données et l'analyse.

L'entête, en Code américain normalisé pour l'échange d'information (American Standard Code for Information Interchange ou ASCII), renseigne que le document est un fichier fcs avec la version utilisée et les délimitations en bytes des différentes parties (*Figure 11*).

```
FCS3.0          58      3637      3638      332617      0 0
```

FIGURE 11: **Entête FCS**. Exemple de format d'un fichier .fcs produit par les cytomètres. "FCS3.0" correspond à la version dans lequel a été enregistré le fichier suivit des positions (en bytes) des différentes parties, "58-3637" correspond à la partie texte, "3637-332617" pour la partie données et "0-0" car il n'y a pas d'analyse enregistré sur ce fichier fcs.

Le texte, en ASCII, contient une suite de clefs-valeur, \$P1N -> "FITC.A". Ces valeurs permettent d'obtenir des informations cruciales sur les paramètres de la machine lors de l'acquisition (la date, les voltages employés,...) mais aussi la matrice de compensation (*Figure 12*).

```

\ $BEGINANALYSIS\0\ $BEGINDATA\3638\ $BEGINSTEXT\0
\ $BYTEORD\4,3,2,1\ $DATATYPE\F\ $ENDANALYSIS\0
\ $ENDDATA\332617\ $ENDSTEXT\0\ $MODE\L\ $NEXTDATA\0
\ $P10B\32\ $P10E\0,0\ $P10N\PC7.A\ $P10R\1048576
\ $P11B\32\ $P11E\0,0\ $P11N\APC.A\ $P11R\1048576\...

```

FIGURE 12: **Text FCS**. Exemple de format de la description de paramètres d'un fichier fcs produit par les cytomètres. Il correspond en une succession de clefs (défini par la présence du "\$") / valeurs délimitées par un séparateur ":

Les données, stockées en bytes, sont ordonnées en un tableau regroupant tous les évènements (majoritairement des cellules) assez gros pour être détectés par le cytomètre. Ces évènements se retrouvent en ligne avec en colonne tout les canaux morphologiques (FS,SS,...) et non morphologiques. Les canaux non morphologiques regroupent les différents fluorochromes ("FITC.A", "PE.A",...) utilisés lors de l'acquisition (*Figure 13*).

FS-H	FS-A	FS-W	SS-H	SS-A	SS-W	FITC.A	PE.A	PC5.5.A	PC7.A	APC.A	APC.AF750.A	PB.A	KO.A	TIME
13831	65573	69	85152	384657	65	8088	16091	2996	1799	50971	20639	39131	16570	211
13154	66083	72	71552	328010	66	289	793	259	134	1816	1134	2866	1099	212
13456	63511	68	84848	380662	64	64123	138989	31299	16642	401086	177856	289408	181929	214
13568	64622	69	85680	383239	65	49	137	62	25	117	436	799	160	229
12015	58045	70	54064	253647	67	1913	4107	791	349	13150	6519	7296	5043	239
12896	61370	69	81104	363832	65	61939	136508	29855	17231	388305	178193	298367	185305	242
289	8109	702	532	5112	670	138	345	60	90	76	493	503	329	243
13447	65828	70	79424	361828	65	2422	3923	633	823	15700	5958	13909	5182	250
286	9012	692	506	3763	503	85	183	97	139	161	694	713	347	256
272	6399	550	488	2942	460	158	219	116	90	223	961	831	154	259
391	7868	382	2469	17703	90	148	396	164	77	352	577	566	213	263
291	11038	762	287	2389	696	105	353	124	273	164	1396	271	146	264
13424	65770	71	76800	348877	66	137044	325983	100741	56519	673814	346419	595527	582989	272
13258	62900	69	83600	374826	65	926	1917	329	211	4778	2984	5283	1991	285

FIGURE 13: **Données FCS**. Exemple de la partie données, convertie en entiers, d'un fichier fcs produit par les cytomètres. Les paramètres FS (taille) et SS (structure) sont définis par plusieurs moyens, avec leur aire "A" mais aussi la largeur "W" et la hauteur "H" de l'impulsion électrique. Les canaux non morphologiques ("FITC.A" à "KO.A") sont définis uniquement par l'aire "A" du signal électrique. Le dernier paramètre est le temps (en seconde) de l'acquisition de l'évènement depuis le début de l'acquisition. Chaque ligne indique les paramètres pour 1 évènement : sont représentés ici 14 évènements d'un fichier fcs ayant acquis 10000 évènements.

L'analyse, en ASCII, est optionnelle et se compose d'informations complémentaires ajoutées après l'acquisition.

L'intensité de fluorescence, enregistrée pour chaque évènement, est sur une échelle linéaire. La valeur maximale varie en fonction des constructeurs des cytomètres. Les cytomètres provenant de chez Beckton Dickinson (BD) enregistrent les fichiers au format FCS 3.0 en 18 bits. Cela veut dire que les données sont comprises sur une échelle entre 0 et 262144 (2^{18}). Les cytomètres construits par Beckman Coulter, quant à eux, utilisent un autre format : List Mode Data (LMD). Ce format particulier contient en réalité deux jeux de données. Un premier est enregistré au format FCS 2.0 en 10 bits. Le deuxième, est au format FCS 3.0 en 20 bits. Les données sont comprises entre 0 et 1048576 (2^{20}).

1.2 Analyse des données de cytométrie en flux

1.2.1 Analyse manuelle

De manière usuelle, l'analyse des fichiers de cytométrie en flux est faite de façon non automatisée. Cette analyse manuelle est réalisée avec l'aide de logiciels comme KALUZA (Beckman-coulter, <https://www.beckman.com>) ou FlowJo (FlowJo LLC, <https://www.flowjo.com>). Par leur intermédiaire, un opérateur peut sélectionner des cellules qui ont une taille/structure ou une intensité de fluorescence similaire, le plus souvent à l'aide de graphes bi-dimensionnels. Cette délimitation de populations, ou "gating", est souvent réalisée de manière successive. Ainsi, la population sélectionnée dans un premier graphe servira de parent pour un second. L'analyse des données de cytométrie consiste à définir une stratégie dans la succession des "gates" (*Figure 14*).

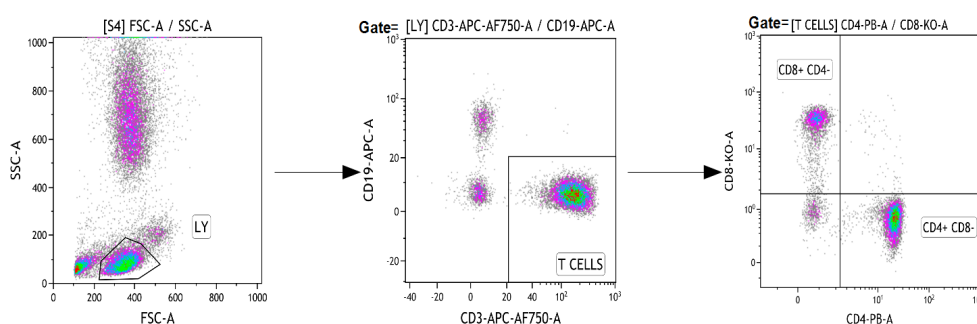


FIGURE 14: **Illustration de différentes gates.** Illustration de l'analyse manuelle avec la sélection de la population lymphocytaire T sur un graphe biparamétrique FCS-A/SSC-A, puis à partir de cette population la sélection des CD3+ sur un graphe biparamétrique CD3/CD19. Ensuite, à partir de cette population la sélection des population CD4+ et CD8+. Réalisé avec Kaluza (<https://www.beckman.com> , Beckman-coulter).

La stratégie employée pour l'analyse des données commence par une série de gates permettant l'exclusion des doublets, cellules mortes et autres débris. Les doublets apparaissent lorsque 2 cellules passent simultanément devant les lasers. Il en résulte un profil anormal dans la largeur, hauteur ou aire de la taille et/ou de la structure. Ces éléments n'apportent aucune information utilisable car ils sont susceptibles de "contaminer" les cellules uniques et peuvent être éliminés par une succession de gates. Après ces premières gates dites de "nettoyage", il est possible de séparer les grandes populations de leucocytes en utilisant toujours la taille/structure. Pour finir par la sélection d'une population en fonction de l'expression, ou de l'absence d'expression, de diverses combinaisons de marqueurs (*Figure 15*).

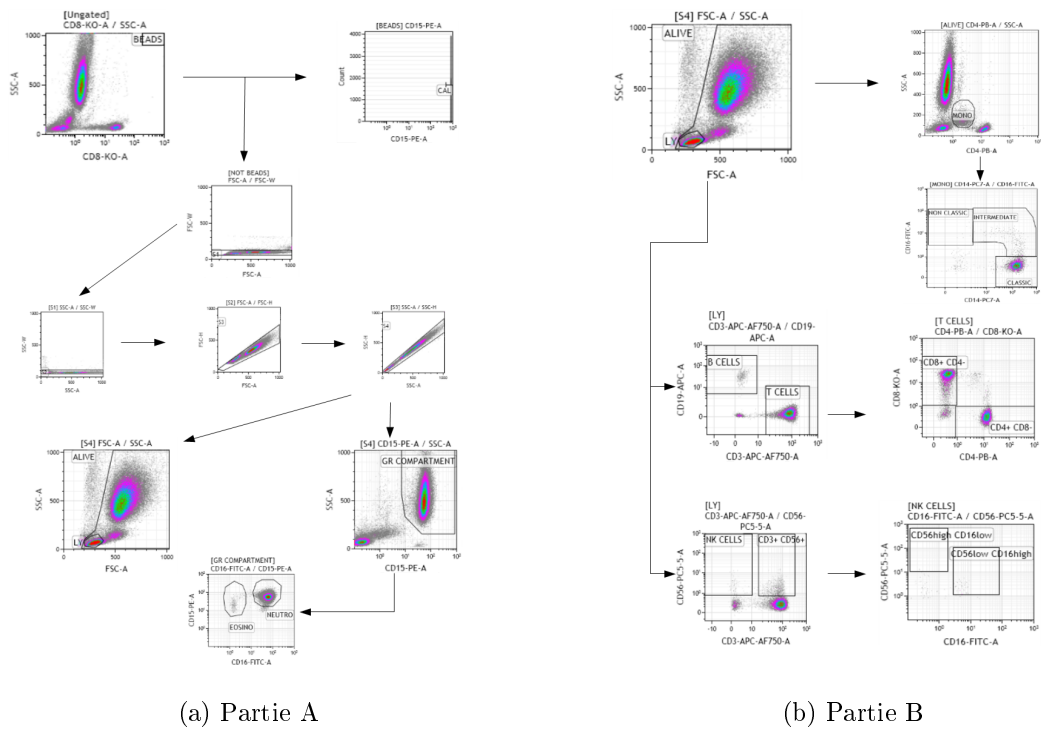


FIGURE 15: **Stratégie d'analyse.** Exemple de stratégie d'analyse manuelle. **PARTIE A** : La première étape consiste en la sélection des billes de comptage flowCount, puis par 4 gate d'exclusion des débris et doublets (S1, S2, S3 et S4). A partir de la gate S4, les granulocytes sont définis comme les évènements positifs au marqueur CD15 PE. Une dernière séparation est faite sur les granulocytes sur la présence ou non de CD16 FITC, permettant de distinguer les éosinophiles et les neutrophiles. La gate S4 permet d'enlever les dernier débris et doublets par la gate ALIVE , réunissant l'ensemble des Leukocytes, et aussi de définir la gate des lymphocytes sur leur taille structure. **PARTIE B** : La gate des lymphocytes permet la définition des lymphocytes B et T de par l'expression des marqueurs CD19 et CD3 respectivement. Ainsi que les sous populations des lymphocytes T en fonction de l'expression du CD4 et du CD8 parmi les CD3+. Cette gate des lymphocytes permet de définir aussi les Natural Killer (NK), ses sous populations et les NK-like. La dernière grande population est les monocytes avec une expression faible du CD4, ainsi que les 3 sous-populations de monocytes classique, intermédiaire et non classique. Ces sous-populations sont distinguées par les différences dans l'expression du CD14 et CD16, avec CD14+ CD16- pour les monocytes classiques, CD14+ CD16+ pour les monocytes intermédiaires et CD14 faible CD16+ pour les monocytes non-classique.

L'analyse manuelle, à l'aide des logiciels tels que Kaluza ou FlowJo, est le gold standard de la cytométrie en flux et ce bien qu'elle soit très consommatrice en temps. En effet, il est préférable que l'analyse manuelle soit réalisée par le même opérateur car des variations ont pu être observées lorsque plusieurs opérateurs réalisent l'analyse [7, 8].

Malgré la centralisation et la réalisation de l'analyse par un opérateur unique, l'analyse manuelle reste sujette à des variations, surtout si elle concerne des milliers

de patients inclus sur plusieurs années. Pour obtenir une stratégie d'analyse stable, sans monopoliser un opérateur pendant plusieurs mois, des outils d'automatisation de l'analyse sont apparus depuis quelques années [9].

1.2.2 Outils d'analyse automatisée

Ces dernières années ont vu une augmentation importante du développement d'outils permettant de s'affranchir de l'analyse manuelle des données de cytométrie en flux [10, 9]. Il y a eu une augmentation du nombre des fluorochromes disponibles, ainsi qu'une augmentation du nombre de patients inclus dans les études. Ces aspects ne permettent pas d'exploiter complètement les données ainsi générées. En effet, l'analyse manuelle étant sensible à l'utilisateur et se limitant à une succession de gate à 1 ou 2 dimensions, l'automatisation devient plus que nécessaire pour pleinement exploiter les résultats de cytométrie en flux.

R est très utilisé pour naviguer dans toutes ces données de cytométrie en flux et d'optimiser leur exploitation. L'attrait de R est sa puissance pour les analyses statistiques et la production de graphes. De plus la cytométrie en flux possède une communauté importante et active dans le développement de packages en opensource sur bioconductor [11].

Un des packages les plus important est **FlowCore** [12] qui fournit les outils de base permettant la manipulation des fichiers de cytométrie sous R, aussi bien sous les formats FCS que LMD. Par l'intermédiaire de FlowCore, il est possible d'accéder aux valeurs brutes enregistrées lors du passage dans le cytomètre. C'est à dire que les valeurs observées sont dans leur échelle originale (18 ou 20 bits) et sans l'application de la matrice de compensation. Ces données sont stockées sous la forme d'un objet **FlowFrame**.

Le flowframe possède toutes les informations contenues dans le fichier de cytométrie. Que ça soit les parties descriptives, comme l'entête avec les noms des paramètres, ou les intensités de fluorescence qui sont sous forme d'un tableau avec les différents canaux (FS/SS inclus) en colonne et les événements en ligne. D'autres informations cruciales peuvent être retrouvées comme les voltages ou la matrice de compensation utilisée lors de l'acquisition du patient. Étant sous R, le flowframe peut être modifié quant au nom des paramètres ou aux valeurs d'intensité. Ces modifications peuvent être enregistrées en tant que nouveau fichier fcs. Le flowframe peut être aussi utilisé par d'autres packages. Cela donne accès à d'autres fonctionnalités complémentaires à flowCore. Tel que le package FlowViz [13], qui permet la visualisation 1-2 dimension, ou bien le package flowStats [14], qui fournit davantage d'algorithmes pour l'analyse statistique des données de cytométrie en flux.

De multiples packages existent pour l'analyse des données de cytométrie en flux sous R avec des performances variables. FlowCAP [9, 15] challenge a testé différents algorithmes de clustering sur différents jeux de données. Les résultats ont été

comparés à plusieurs analyses manuelles réalisées par des experts. Il en est ressorti que plusieurs algorithmes de clustering produisent des résultats proches de l'analyse manuelle. En plus de montrer que l'automatisation arrive à des niveaux proches de l'analyse manuelle, flowCAP a remis en question le gold standard de la cytométrie en flux. En effet, il a été observé qu'il existait une grande variabilité entre les différents résultats obtenus par les experts pour les mêmes fichiers de cytométrie. Depuis FlowCAP, de nouveaux algorithmes d'automatisation du gating ont été proposés dans différents packages R. La stratégie employée pour ces algorithmes de clustering peut être divisée en 2 catégories : les algorithmes non supervisés et les algorithmes supervisés.

1.2.3 Analyses automatisées non-supervisées

Concernant les algorithmes non-supervisés, la stratégie est de partir sans a priori et de chercher tout les clusters possibles. Cette stratégie d'analyse nécessite souvent une partie de catalogage (associer une population à un cluster ou à un groupe de clusters) pour pouvoir interpréter la signification des résultats. L'avantage de cette méthode est l'utilisation de toutes les dimensions (canaux) dans la définition des populations. Elle n'est plus limitée à l'exploration des données par une succession de graphes à 2 dimensions.

Par exemple Flock, flow clustering without k, permet de définir des populations en utilisant tout les paramètres. Pour cela, Flock partitionne l'espace en cellules, puis sélectionne les cellules de plus forte densité et les fusionne si elle sont assez proches [16].

Un autre exemple est flowtype [17], ce package R va définir les seuils entre les évènements positifs et négatifs pour tous les marqueurs. Dans un deuxième temps, il va réaliser toutes les combinaisons de marqueurs avec ces seuils. Les différentes combinaisons de marqueurs, produites par flowType (*Figure 16*), peuvent être envoyées à un deuxième package R, RchyOptimyx (*Figure 17*). Ce package va déterminer la meilleure stratégie d'analyse pour les populations sélectionnés dans les résultats de flowType [18].

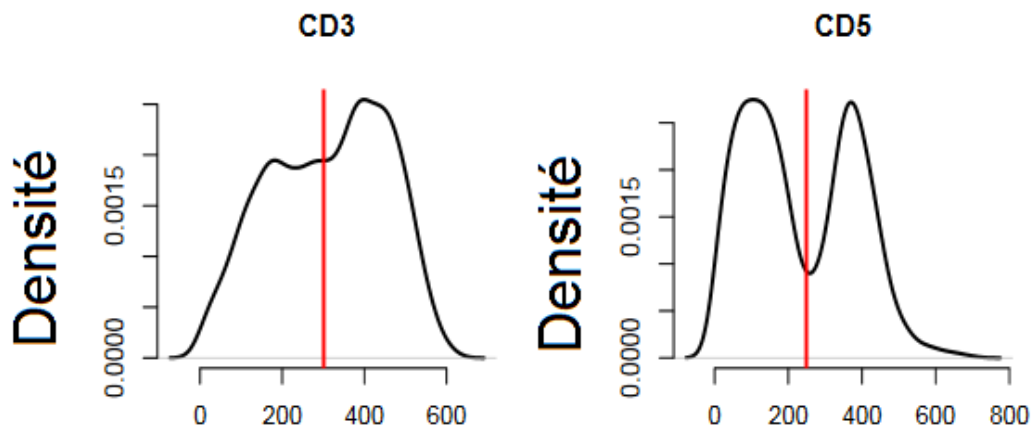


FIGURE 16: **FlowType**. Exemple de l'utilisation de l'algorithme de FlowType. Le trait rouge délimite les événements dit positifs au marqueur (à droite du trait) et ceux négatifs au marqueur (à gauche du trait). Cette délimitation est réalisée pour tous les marqueurs (ici le CD3 et le CD5). Source d'après (<https://bioconductor.org/packages/release/bioc/vignettes/flowType/inst/doc/flowType.pdf>).

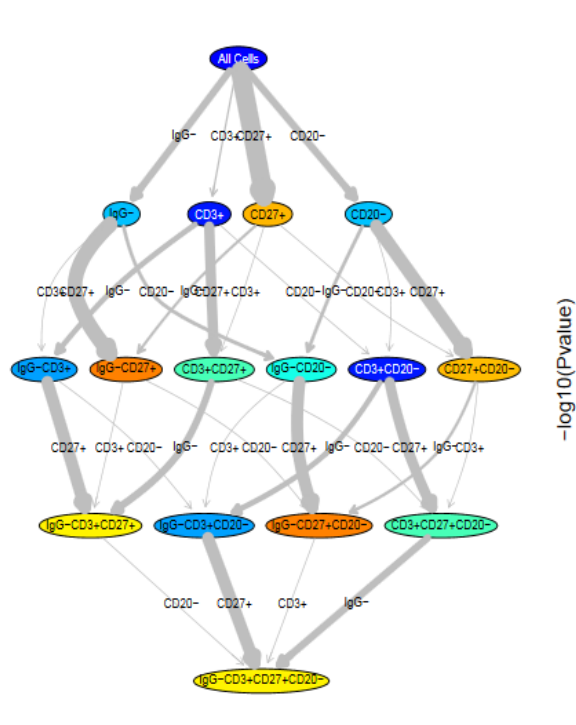


FIGURE 17: **RchyOptimyx**. Exemple de l'utilisation de l'algorithme de RchyOptimyx à partir de résultats pouvant provenir de flowType. Source (<https://www.bioconductor.org/packages/release/bioc/html/RchyOptimyx.html>).

1.2.4 Analyses automatisées supervisées

Les algorithmes pour le gating supervisé utilisent des données externes pour aider au clustering des populations. Leur but est le plus souvent de reproduire le gating manuel en utilisant des algorithmes de clustering sur une succession de graphes bi-paramétriques. En cela, ils permettent d'obtenir les données sur des populations connues.

C'est le cas pour `flowClust` (*Figure 18*) qui est un package R permettant de clusteriser des populations cellulaires en sélectionnant 2 paramètres et le nombre de clusters attendu [19].

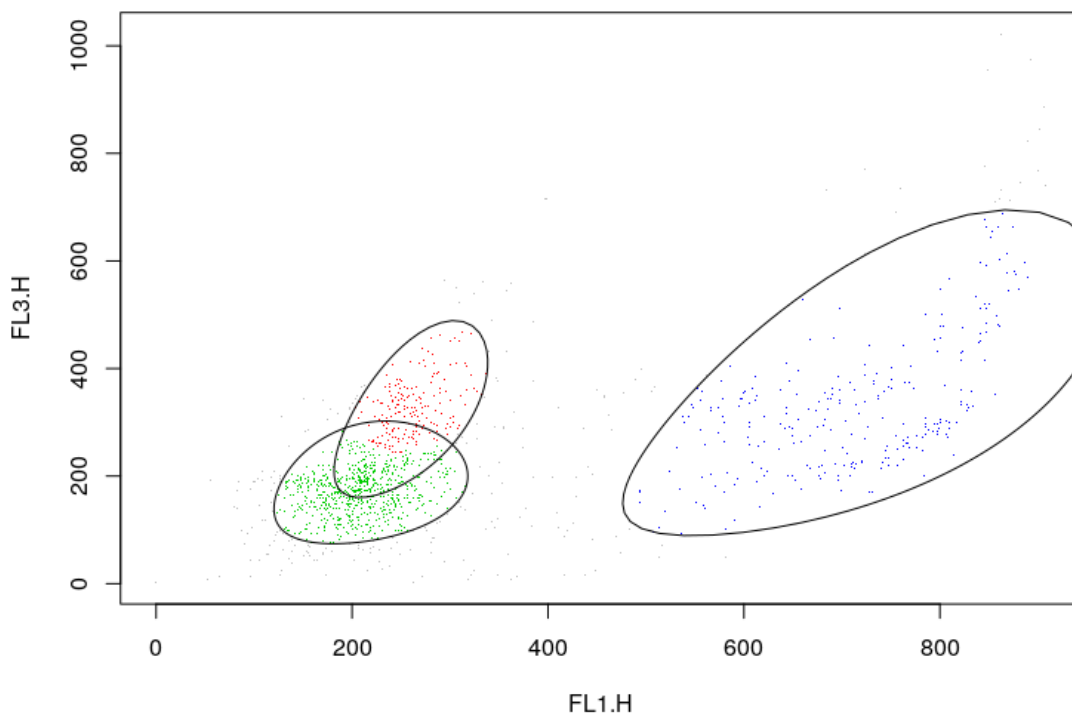


FIGURE 18: **flowClust**. Exemple de résultats obtenus par l'utilisation de l'algorithme de `flowClust`. Source (<https://bioconductor.org/packages/release/bioc/vignettes/flowClust/inst/doc/flowClust.html>).

Malgré les avancés dans l'automatisation du gating, de nombreuses difficultés peuvent être rencontrées lors de l'inclusion et de l'analyse d'un grand nombres d'échantillons. Particulièrement lorsque plusieurs centres sont concernés.

1.2.5 Analyses multi-centriques

Les analyses multi-centriques concernent souvent un nombre important d'inclusions. Avant même de comparer entre les centres, ce nombre d'inclusions est une source importante de variations entre les acquisitions. En effet, l'influence de plusieurs paramètres peut compromettre la comparabilité des résultats tel que les lots de réactifs, le protocole utilisé et le maintien de la sensibilité des instruments dans le temps.

C'est sans compter les erreurs de manipulation possibles, comme l'utilisation d'un mauvais anticorps ou une erreur dans le pipetage du réactif liquide, qui peuvent arriver lors de la préparation des échantillons. Les conséquences de ces erreurs rendent certains patients inutilisables dans le reste de l'étude ou peuvent induire des variations techniques dans les résultats produits.

Des variations dans les cytomètres ont un impact sur les résultats [8]. L'hétérogénéité dans les bancs optiques (laser, miroirs dichroïques, PMT) et dans l'électronique rend chaque cytomètre quasiment unique. Cette variabilité dans les caractéristiques des machines ne permet pas la comparabilité des données produites directement.

1.2.6 Réponses aux problématiques multi-centriques, analyse de la littérature existante

Des études ce sont intéressées aux problèmes que soulèvent la comparabilité des données de cytométrie en flux lors de l'utilisation de plusieurs cytomètres [7, 8, 20, 21, 22, 23, 24, 25, 26]. Chacune des études essaie de répondre à une ou plusieurs des difficultés rencontrées lors de la comparaison inter-centre des données.

1. **Semi-automated and standardized cytometric procedures for multi-panel and multi-parametric whole blood immunophenotyping [21].**

Cet article du Labex Milieu intérieur a pour but d'observer les variations immunophénotypiques de populations saines et d'identifier les facteurs à l'origine de cette hétérogénéité de la réponse immunitaire.

Les échantillons, provenant de donneurs sains, sont passés dans 2 cytomètres Miltenyi MACSQuant. Afin de réduire les variations, ils ont sélectionné 4 panels de 8 couleurs et développé des SOP. Un des avantages de cette étude est l'utilisation d'un automate pour réaliser le marquage des échantillons.

Pour valider leur étude, ils ont réalisé un test, effectué par un opérateur unique sur un seul cytomètre, de 6 échantillons de donneurs sains. Chaque échantillon est divisé en 5 aliquots. Les nombres absolus en résultant ont des CV inférieurs à 15%, sauf pour un des 6 donneurs pour la population des pDC (21,5%) et

pour les monocytes CD14+ pour 2 des 6 donneurs (16% et 23%).

Un deuxième test a été exécuté pour évaluer la reproductibilité dans le temps avec du sang stabilisé (Eurocell) sur une période d'un mois. Les résultats obtenus ont des CV inférieurs à 5.3% pour la concentration cellulaire et inférieurs à 5% pour la fréquence dans les lymphocytes.

2. Centrally Determined Standardization of Flow Cytometry Methods Reduces Interlaboratory variation (Percentages) [7].

Le but de cet article est de comparer 3 méthodes d'analyse des données de cytométrie en flux impliquant différents centres. La première (A) est de réaliser le gating dans chaque centre avec une stratégie propre à chaque laboratoire. La deuxième (B) est de déterminer une stratégie de gating commune puis que chaque centre réalise ce gating. La dernière (C) centralise à la fois la stratégie et l'analyse.

Les échantillons utilisés pour ce test sont des PBMC cryopréservés isolés à partir de 3 donneurs sains. Ils ont été envoyés aux 7 centres de l'étude qui les ont préparés et acquis en triplicate. La comparaison inter-centres est faite sur la médiane des fréquences de chaque individu.

Le CV est calculé comme :

$$\frac{\text{déviation standard}}{\text{moyenne}} * 100 \quad (1)$$

Lorsque l'on compare les CV des différentes méthodes : A, 4.4–102.1% ; B, 10.9–65.6% ; C, 1.8–20.9%. Ainsi, seule la méthode pleinement centralisée arrive à des CV inférieurs à 25%. C'est ce seuil d'acceptabilité à 25% qui a été déterminé en référence à un autre article [27].

3. Comparable Flow Cytometry Data can be Obtained with Two Types of Instruments Canto II, and Navios. A GEIL Study [22].

Dans le cadre du Groupe d'Etude Immunologique des Leucémies (GEIL), 5 centres, ayant accès à des cytomètres de type Navios et CantoII, ont travaillé sur du sang périphérique afin de voir si les résultats sont comparables.

Dès le début de l'analyse, les paramètres morphologiques sont exclus car présentant trop de différences entre les centres.

Concernant l'IMF, 2 calibrations ont été testées :

- La première est effectuée sur des billes 8 pics. Ces billes sont standardisées pour obtenir 8 pics stables à différentes intensités de fluorescence et ceci pour chacun des canaux.

L'IMF du 7^{ème} pic est déterminé pour tous les canaux dans le centre de référence et sert de cibles pour les autres centres. Pour cela les valeurs cibles sont transformées en fonction du cytomètre afin qu'elles soient à la même échelle.

- La deuxième calibration se base sur l'expression du CD16 des granulocytes dans le sang. Toujours avec un centre de référence et une mise à l'échelle des valeurs cibles entre les centres.

Estimation des différences entre les centres par la méthode de Bland et Altman :

$$\frac{\text{IMF référence} - \text{IMF centre}}{\frac{\text{IMF référence} + \text{IMF centre}}{2}} * 100 \quad (2)$$

Les résultats de la calibration utilisant les billes 8 pics, indiquent des différences inter-centres (par Bland et Altman) inférieures à 4% pour les fréquences et inférieures à 15% pour les IMF.

Concernant la méthode utilisant l'expression du CD16, les variations inter-centres sont inférieures à 30% pour les IMF.

4. **Standardizing Flow Cytometry Immunophenotyping Analysis from the Human ImmunoPhenotyping Consortium [23].**

L'idée initiale de cette étude est que la standardisation pré-analytique permet de réduire les variations techniques mais qu'il en reste sur la partie analytique. Pour cette partie, la standardisation peut être complétée par l'utilisation d'algorithmes lors de l'analyse des données de cytométrie.

L'étude utilise des PBMC cryopréservés, des CytoTrol (Beckman Coulter, Miami, FL), et des PBMC lyophilisés qui servent de témoins. Ces échantillons sont acquis dans 9 sites possédant des cytomètres Fortessa ou LSR (tous les 2 de la société BD).

Concernant la standardisation pré-analytique, elle est réalisée à l'aide de SOPs. La partie analytique est faite en comparant l'analyse manuelle centralisée, flow-Density et Opencyto. FlowDensity est un algorithme de clustering supervisé, disponible sous R, se basant sur la densité des distributions des événements

pour réaliser un gating séquentiel. Opencyto est un package R permettant l'utilisation de plusieurs algorithmes de clustering afin de réaliser une analyse supervisée.

Afin de déterminer l'efficacité de cette analyse automatisée, les fichiers ont été analysés manuellement dans un seul centre. Les conclusions tirées de cette comparaison sont que les variations en intra-centre sont plus faibles qu'en inter-centres et que l'analyse automatisée fait preuve d'une meilleure stabilité que l'analyse manuelle. Mais, de plus fortes variations sont observées pour les populations rares, mettant en avant une difficulté pour les algorithmes de gating d'identifier ce genre de populations.

5. **Basophil Activation Test : Implementation and Standardization Between Systems and Between Instruments [24].**

L'étude s'intéresse à l'observation des basophiles, plus particulièrement à leur activation, et à la comparaison de 2 kits de test, FlowCast(Buhlmann, Schönenbuch, Switzerland) et BasoflowEx (Exbio, Praha, Czech Republic). Ces kits sont testés sur des prélèvements de sang frais provenant d'individus sains.

La standardisation entre les 2 cytomètres utilisés dans l'étude, Navios et CantoII, est exécutée par la définition de cibles, pour les canaux FITC et PE, dans le NAVIOS avec l'utilisation de billes de calibration (Cytocal beads, DISTRILABVR The Netherlands Multifluor Calibrator, Thermo Scientific, Carlsbad, CA). Ces cibles ont ensuite été converties pour être utilisées dans le Canto II.

Pour tester cette calibration, 16 échantillons ont été analysés sur les 2 instruments. L'estimation des différences entre les centres est faite par un graphique de Bland et Altman mais surtout par l'observation de la corrélation entre les 2 cytomètres utilisés lors de cette étude.

Les corrélations obtenues pour les différentes populations ont un r^2 supérieur à 0,90 entre les deux instruments dans les différentes conditions.

6. **NIST/ISAC Standardization Study : Variability in Assignment of Intensity Values to Fluorescence Standard Beads and in Cross Calibration of Standard Beads to Hard Dyed Beads [25].**

Le but de cette étude est de déterminer la variabilité entre différents centres par l'utilisation de 2 types de billes : billes avec de multiples fluorophores et billes de compensation. Ces billes, fournies par différents producteurs, sont acquises dans 133 cytomètres, provenant des constructeurs BC et BD.

La calibration se fait sur les billes de compensation avec comme cible une IMF proche de 1000 (sur une échelle entre 1 - 10000). En utilisant les paramètres ainsi déterminés, les billes avec de multiples fluorophores sont acquises.

La comparaison des différentes machines se fait à l'aide d'un ratio $\frac{\text{billes avec de multiples fluorophores}}{\text{billes de compensation}}$ qui est normalisé sur la médiane de tous les instruments.

Plus précisément, les CV des ratios sont calculés pour les groupes de même machine. Les CV ainsi générés sont supérieurs à 10%. Les auteurs de cette étude en concluent que l'utilisation de billes 8 pics, ne partageant pas les fluorochromes des panels, pour la calibration doit être faite avec précautions.

7. **Standardization of whole blood immune phenotype monitoring for clinical trials : panels and methods from the ONE study [8].**

Cette étude veut standardiser le phénotypage du système immunitaire par cytométrie en flux. Pour réaliser cela, l'étude utilise une série de SOPs. Le travail s'effectue sur les fréquences et les concentrations cellulaires issues d'échantillons de sang.

Ils ont testé les différences causées par différents opérateurs, les différences intra-centre (CV entre 0.05% et 20%) et les différences inter-centres 0.05% à 30% de variation entre les centres.

Pour conclure, les auteurs préconisent de réaliser le marquage de cytométrie dans les 4h après le prélèvement. L'acceptabilité de la variabilité, inter-centres, est inférieure à 30% pour la plupart des populations cellulaires.

8. **Quality assurance of intracellular cytokine staining assays : analysis of multiple rounds of proficiency testing [26].**

L'étude s'intéresse à la comparaison de résultats de marquage de cytokines intra-cellulaires avec l'utilisation de PBMC cryopréservés.

L'étude s'est déroulée sur 7 séries d'expériences. Le nombre de centres varie en fonction de la série, allant de 10 centres lors du début de l'étude, jusqu'à atteindre 16 centres à partir de la 5^{ème} série.

Le CV est calculé comme $\frac{\text{déviation standard}}{\text{moyenne}} * 100$ sur les fréquences cellulaire. Pour comparer les différences entre expériences, le test de Kruskal-Wallis est utilisé, suivi par le test de Dunn pour les comparaisons multiples.

La recommandation de l'étude est que les variations sont acceptables, en inter-centres, si le CV des fréquences est inférieur à 35%.

9. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols [20].

L'étude Euroflow a pour but de standardiser les instruments et les procédures liées à l'utilisation de la cytométrie en flux dans différents laboratoires. Pour cela, ce projet a sélectionné une combinaison optimale de 8 fluorochromes pour différents panels.

Ces panels sont couplés avec le développement de différentes SOPs. Parmi ces SOPs, il y a une description du paramétrage des différentes machines. Ce paramétrage utilise les billes 8 pics plus particulièrement le pic de plus forte intensité de fluorescence. L'acceptabilité de la calibration inter-centre est atteinte si l'IMF est dans les $\pm 15\%$ de l'IMF cible et si les CV des pics sont inférieurs à 6%.

Afin d'évaluer la comparabilité des données, euroflow a testé 2 approches. La première est d'envoyer un échantillon de sang stabilisé dans huit différents centres. La deuxième est le marquage d'échantillons comparables dans les différents centres (n=30 échantillons de sang).

Les CV des IMF résultantes de la première approche se retrouvèrent inférieurs à 44% (et même inférieur à 16% pour 4/8 des fluorochromes). Concernant la deuxième approche, les CV des IMF sont similaires dans le sens où les mêmes fluorochromes (4/8) se retrouvent avec de plus faibles variations.

Une des particularité de l'étude Euroflow est qu'elle propose de rassembler les panels d'un patient en un seul fichier fcs. Pour cela les événements sont fusionnés en fonction de leurs distances pour les marqueurs en commun.

Globalement, ces différentes études ont essayé de répondre à différents aspects de la comparaison des données entre cytomètres. Les solutions proposées se recoupent sur certains points, comme l'utilisation de SOPs.

Il en résulte que la comparaison entre plusieurs centres nécessite : une calibration qui est réalisée à l'aide de billes standardisées. Une sélection des fluorochromes qui prend en compte les capacités des différents cytomètres. Des SOPs qui décrivent les diverses procédures pré-analytiques, ainsi qu'une analyse centralisée par un opérateur unique, voir même par l'automatisation du gating par différents algorithmes tel que Opencyto.

Tout cela engendre une réduction des variations observées entre les différentes machines et ainsi permet de se rapprocher de la comparabilité des données de cyométrie en flux.

1.3 Contexte de l'étude

Notre système immunitaire est une machinerie complexe qui nous permet de lutter contre les invasions d'une pléthore de pathogènes tel que les bactéries ou les virus. La réponse de ce système peut être divisée en 2 types : l'inné et l'adaptatif.

La réponse immunitaire innée est la réponse la plus rapide dans la lutte contre les pathogènes. Le pathogène doit en premier passer les barrières physiques, comme la peau ou les muqueuses. Une fois ces barrières franchies, des cellules spécialisées prennent le relais dans la lutte en détruisant le pathogène par phagocytose ou bien en le piégeant par l'intermédiaire de la netose [28].

Le système immunitaire adaptatif, plus lent dans sa mise en place, permet une réponse plus spécifique au pathogène. De plus, ce dernier a une forme de mémoire immunitaire qui permet une réponse plus rapide lors de prochaines rencontres avec le pathogène.

Les deux systèmes peuvent communiquer, par exemple avec les cellules présentatrices d'antigènes comme les cellules dendritiques et les macrophages qui peuvent présenter des antigènes aux lymphocytes [29].

1.3.1 Populations cellulaires

Toutes les cellules du système immunitaire ont pour origine des cellules souches issues de la moelle osseuse ou du foie foetal. Ces cellules souches pluripotentes peuvent se différencier en plusieurs populations cellulaires (*Figure 19*). Une fois différenciées, ces cellules vont pouvoir circuler pour assurer la protection de l'ensemble de l'organisme.

Différentes sous-populations cellulaires ont été décrites assumant différents rôles au sein de la réponse immunitaire (*Tableau 3*). L'identification de ces populations et sous-populations se base sur des critères physique comme la taille ou la structure interne des cellules. Mais aussi sur l'expression de certaines protéines ou complexes protéiques présent à la membrane externe appelé *Cluster de Différenciation (CD)*.

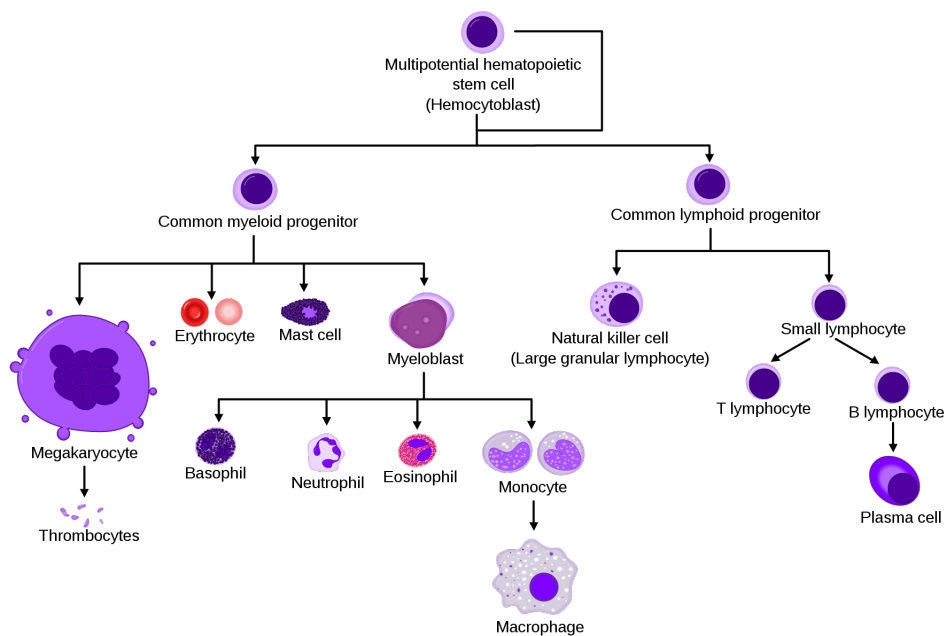


FIGURE 19: **Hématopoïèse**. Arbre représentant l'origine et les différentes étapes de maturation des populations de cellules à partir d'une cellule pluripotente. Source d'après (Mikael Häggström and A. Rad).

POPULATION	RÔLE	Classe de Différenciation
B (LB)	Production d'anticorps	CD19+
T helper (<i>Th</i>)	aide les macrophages et les LB → Th1 activation des macrophages, LT cytotoxiques → Th2 aide la production d'anticorps des LB	CD3+ & CD4+
T cytotoxique	Élimination de cellules infectées ou cancéreuses par libération de perforine et de granzyme pour induire l'apoptose	CD3+ & CD8+
Natural Killer NK	Élimination des cellules infectées par virus	CD56+ & CD3-
Cellule Dendritique DC	Cellule présentatrice d'antigène	CD11b+ ou CD123+
Monocytes	se différencie en macrophage →phagocytose →cellule présentatrice d'antigènes →reparation tissus	CD14+
Neutrophiles	phagocytose Neutrophil extracellular traps (NETs)	CD15+ & CD16+
Eosinophiles	Lutte contre parasites Cellules présentatrice d'antigènes Destruction de cellules tumorales	CD15low & CD16+
Basophiles	Contribue a la réponse inflammatoire	CD123+ HLADR-

TABLEAU 3: Principales populations du système immunitaire présentes dans le sang

Lors du développement des populations du système immunitaire adaptatif, il arrive que certaines de ces cellules puissent reconnaître des éléments du soi comme pathogènes. Dans ce cas, des mécanismes de contrôle permettent le tri entre les cellules auto-réactives ou non. Ces contrôles permettent la destruction (par apoptose) ou l'inactivation de la cellule (anergie) [30].

Lorsque ces étapes du contrôle de l'auto-réactivité des cellules immunitaires échouent, cela peut engendrer des maladies auto-immunes. Classiquement, ces maladies sont divisées en 2 groupes, les maladies dites spécifiques d'organes et celles dites systémiques. C'est ce deuxième groupe de maladies qui va nous intéresser plus particulièrement.

1.3.2 Maladies auto-immune systémiques

Les maladies auto-immune systémiques (SADs) rassemblent de multiples classes de maladies hétérogènes [31] dont le point commun est la perte de tolérance vis à vis de multiples organes. Malgré une forte hétérogénéité dans les organes impactés, les différentes maladies auto-immunes systémiques partagent un grand nombre de caractéristiques tels que des symptômes cliniques et possèdent des mécanismes immunopathologiques similaires. De plus, les patients appartenant à la même classe de maladie peuvent avoir des expressions très différentes dans leurs symptômes.

Lupus érythémateux disséminé Le Lupus Érythémateux Disséminé (LED) est caractérisé par des plaques rouges en masque de loup sur le visage, une atteinte des articulations, des lésions de la peau, une atteinte des reins, du coeur, des poumons, des vaisseaux sanguins et du cerveau [32, 33].

Sclérodémie Systémique La Sclérodémie Systémique (ScS) a comme symptômes cliniques des lésions de la peau entraînant un épaissement et un durcissement de cette dernière (à cause d'une augmentation du collagène). Cette maladie peut aussi entraîner une atteinte des poumons avec une pneumopathie interstitielle, des manifestations touchant le coeur, les intestins et les reins [33, 34].

Polyarthrite Rhumatoïde La Polyarthrite Rhumatoïde (PR) est, comme de nombreuses maladies auto-immune systémiques, une maladie chronique qui peut être handicapante. En effet, cette maladie systémique attaque les articulations induisant douleur, raideur et gonflement. Cette attaque peut aller jusqu'à la destruction de l'articulation [35, 36, 33].

Syndrome de Gougerot Sjögren Le Syndrome de Gougerot Sjögren(SGS) est une infiltration lymphocytaire des glandes exocrines (glandes salivaires et lacrimales) par la production d'auto-anticorps. De cet infiltrat résulte une destruction des glandes, ce qui entraîne des complications telles qu'une sécheresse buccale et un manque de larmes. Le SGS peut être considéré comme secondaire si une autre maladie auto-immune lui est associée [37, 33].

Syndrome des anti-phospholipides Le Syndrome des anti-phospholipides primaire (SAPL primaire) peut être diagnostiqué lors de thromboses. Son impact peut toucher plusieurs régions du corps, comme la peau ou les reins, conséquence de la thrombose. Si le SAPL est en co-occurrence avec une autre maladie auto-immune, similairement au SGS, il est considéré comme un Syndrome des anti-phospholipides Secondaire [38].

Connectivite Mixte La Connectivite Mixte (CM) possède plusieurs symptomes caractéristiques de plusieurs maladies auto-immunes systémiques, comme un mélange de LED et de PR [39].

Connectivite Indifférenciée La Connectivite Indifférencié (CI) est définie par une série de symptomes communs aux maladies auto-immunes systémiques mais pas de symptomes qui permettraient de la classer clairement [40].

Des similitudes entre les différentes manifestations clinique des maladies auto-immune systémiques posent un réel problème lors du diagnostic par le corps médical. La difficulté peut aussi provenir de la variabilité au sein d'une même classe de maladie.

Tout cela permet de mettre en lumière des difficultés rencontrées par le corps médical pour rendre un diagnostic de ces maladies ainsi que pour les entreprises pharmaceutiques dans le développement et la vente de médicaments adaptés. En effet, des problèmes peuvent être engendrés en raison de l'accessibilité aux traitements. Certains patients reçoivent des médicaments pour leur classe de maladie mais avec très peu de bénéfices à cause de l'hétérogénéité au sein de cette même classe de maladie. Alors que d'autres patients se voient refuser l'accès à des traitements destinés à une autre classe de maladie, mais qui seraient potentiellement efficaces car ils partagent des mécanismes immunopathologiques similaires.

Fort de ce constat, un projet européen impliquant des laboratoires de recherche et des entreprises pharmaceutiques a été lancé : le projet IMI **PRECISESADS**.

1.3.3 le projet PRECISESADS

PRECISESADS a pour but de reclassifier les maladies auto-immunes systémiques. Pour réaliser cette reclassification, il est évident que l'on ne peut pas se contenter uniquement des critères cliniques des SADS car c'est sur ces critères que se base la classification actuelle. Il faut donc définir de nouvelles signatures biologiques en se basant sur des marqueurs nouvellement découverts. Pour trouver ces marqueurs, il est nécessaire d'employer différents types d'analyses et ce sur un grand nombre de patients provenant de différentes classes de maladies. L'ampleur du projet nous offre une meilleure vision des SADS avec pour objectif final de permettre

au patient l'accès à une médecine de précision.

Pour atteindre cet objectif, PRECISESADS a retenu 7 maladies autoimmunes systémiques (LED, ScS, PR, SGS primaire, SAPL primaire, CM et CI). Des patients atteints d'une de ces 7 SADS, ainsi que des contrôles, sont inclus dans plusieurs centres en Europe. Les données sont générées à partir de différentes techniques dites "Omiques" (Génomique, Transcriptomique, Epigénétique, Protéomique, Métabolomique, Cytométrie en flux). Toutes ces données sont regroupées et analysées à l'aide d'outils bioinformatiques par l'équipe de bioinformatique de PRECISESADS. C'est à partir de cette analyse que l'on va chercher de nouveaux biomarqueurs et ainsi réaliser la nouvelle classification des SADS.

L'inclusion des patients et des témoins sains se fait dans 2 cohortes différentes. Pour l'étude, 3167 patients et témoins sains sont répartis dans une des 2 cohortes :

- La cohorte **Transversale** inclut des patients et des témoins sains. Cette première cohorte est elle-même divisée en 2 phases :

PHASE I : Cette phase a toutes les analyses omiques et inclut 303 patients (environ 50 patients par maladie auto-immune systémique avec les SAPL primaire, CM et CI comptabilisé comme un seul groupe) et 54 témoins sains.

PHASE II : Lors de cette phase se fait l'inclusion de 2320 patients (environ 400 patients par maladie auto-immune systémique et 563 témoins sains). Cette phase est plus "légère" en terme d'analyses effectuées après le prélèvement pour permettre l'inclusion d'un plus grand nombre de patients.

- La cohorte **PROSPECTIVE** où 215 patients nouvellement diagnostiqués sont inclus puis suivis à 6 et 14 mois. A chacune des visites, M0/M6/M14, ces patients bénéficient d'une analyse similaire à ceux de la phase I.

Les différentes cohortes sont une source importante de données. Parmi la multitude des techniques employées, la cytométrie en flux est celle qui va particulièrement nous intéresser pour le reste de ce document. La cytométrie en flux est une méthode puissante qui permet de rapporter des informations qualitatives et quantitatives sur un échantillon de sang.

1.3.4 La cytométrie en flux dans PRECISESADS

Pour le projet PRECISESADS, 9 panels de cytométrie ont été sélectionnés pour couvrir un spectre large des populations cellulaires de l'immunité (*Figure 20*). Ils sont utilisés pour chaque patient de la phase I et pour chaque visite de la cohorte prospective. Concernant la phase II, qui inclut un nombre important d'individus, seulement les 2 premiers panels sont utilisés.

1.3.5 Combinaisons de fluorochromes

Chaque panel possède différentes combinaisons, ou cocktails, de 8 fluorochromes associés aux anticorps (*Figure 20*). Les panels 1 et 2 correspondent aux grandes populations du système immunitaire. Les panels 3 et 4 s'intéressent aux sous populations des lymphocytes T. Les panels 5 et 6 se concentrent sur les sous populations de lymphocytes B. Et les panels 7, 8 et 9 s'intéressent aux Fc gamma récepteurs et à des protéines recepteurs du complément.

	LASER 1 488 nm		LASER 2 638 nm		LASER 3 405 nm			
	FITC	PE	PC5.5	PC7	APC	APC AF750	PB	KO
Panel 1 Leukocytes	CD16	CD15	CD56	CD14	CD19	CD3	CD4	CD8
Panel 2 Cellules dendritiques et basophiles	CD1c	Lin	CD141	CD11c	CD123	DRAQ7	HLA DR	
Panel 3 Sous-populations des lymphocytes T #1	CD57	CD45RA	CD27	CD62L	CD38	CD3	CD4	CD8
Panel 4 Sous-populations des lymphocytes T #2	TCR g-d	TCR a-b	CD25	CD127	CD69	CD3	CD4	CD8
Panel 5 Sous-populations des lymphocytes B #1	IgD	CD267	CD5	CD27	CD19	CD24	CD38	
Panel 6 Sous-populations des lymphocytes B #2	CD43	CD69	CD5	CD27	CD11b		CD20	CD3
Panel 7 Neutrophiles et recepteurs FC gamma	CD64	CD32	CD18	CD14	CD19	CD11b		CD16
Panel 8 Protéines du complément #1	CD35	CD59		CD14	CD19	CD3	CD21	CD16
Panel 9 Protéines du complément #2	CD46	CD55		CD14	CD19	CD3		CD16

FIGURE 20: **Panels de cytométrie.** Présentation des différentes combinaisons anticorps/fluorochromes présentes dans chacun des 9 panels du projet PRECISESADS.

Les anticorps sont sélectionnés en fonction de leur cible, majoritairement des glycoprotéines membranaires. La combinaison de la présence et/ou de l'absence de certains de ces anticorps permet de caractériser les populations formant le système immunitaire. La nomenclature utilisée pour décrire ces cibles est le cluster de différenciation (CD) suivi d'un numéro.

Par exemple, les cellules exprimant le marqueur CD3, sont classifiées comme des lymphocytes T. De plus ces derniers peuvent exprimer d'autres marqueurs comme le CD4 ou le CD8. Ainsi, seront caractérisés plusieurs populations cellulaires en fonction de la combinaison des marqueurs, comme la population CD3+ CD4- CD8+ définissant les lymphocytes T cytotoxique (*Tableau 4*).

TABLEAU 4: **Clusters de différenciation.** Liste des Clusters de différenciation utilisés dans les 9 panels de l'étude PRECISESADS avec leur rôle et la population concernée.

CD	RÔLE	POPULATION
CD1c	protéine présentatrice d'antigènes	T et NK T
CD3	Complexe associé au TCR, transduction du signal	T
CD4	co-récepteur du TCR	T helper , monocytes, macrophages et cellule dendritique(DC)
CD5	récepteur	T et sous-population de LB
CD8	co-récepteur du TCR	T et NK
CD11b	inhibition cellulaire	Monocytes, macrophages, granulocytes et NK
CD11c	activation cellulaire	DC,monocytes, macrophages, granulocytes et B
CD14	liaison avec le lipopolysaccharide bactérien	Monocytes, macrophages,granulocytes,B et DC
CD15	adhésion pour phagocytose	Granulocytes
CD16	activation de la dégranulation , phagocytose	Granulocytes ,NK, monocyte et macrophages
CD18	adhésion et signale	Leukocytes
CD19	co-récepteur du BCR	B
CD20		B
CD21	Récepteur du complement	B
CD24	adhésion	B et granulocytes
CD25	IL2 récepteur	DC, T et B activés
CD27	Récepteur du facteur de nécrose tumorale	T
CD32	Récepteur	granulocytes, monocytes et B
CD35	Récepteur du complement	Granulocytes, monocytes, DC et B
CD38	adhésion et transduction du signale	T ,B et NK
CD43	adhésion	Leukocytes (sauf B)
CD45RA	signal	T, B, granulocytes et monocytes
CD46	régulation du complement	Leukocytes
CD55	régulation du complement	Leukocytes
CD56	adhésion	Nk, T
CD62L	adhésion	T, B, granulocytes, NK et monocytes
CD64	ADCC(antibody dependent cell mediated cytotoxicity)	macrophages et monocytes
CD67 (CD66b)	CGM6 (CEA gene membre 6)	Granulocytes
CD69	marqueur d'activation	T ,B et NK
CD123	récepteur	
CD127	IL7 récepteur	T
CD141	récepteur cellulaire en lien avec la coagulation	monocytes et DC
CD267	Récepteur du facteur de nécrose tumorale	B
DRAQ7	marqueur de viabilité	
HLA DR	présentation d'antigènes	macrophages, B et DC
IgD	activation des cellules B	B
TCR g-d	reconnaissance d'antigènes	T
TCR a-b	reconnaissance d'antigènes	T

La combinaison de ces 39 marqueurs permet de caractériser plus de 140 populations et sous-populations cellulaires connues grâce à l'utilisation de différents panels. La sélection des fluorochromes et leurs combinaisons avec des anticorps a été faite en prenant en compte la compatibilité avec les cytomètres des centres du projet, mais ce n'est pas les seules choses à prendre en compte lorsque l'on compare différents cytomètres.

1.3.6 Analyses dans le cadre de PRECISESADS

Dans le cadre de PRECISESADS, l'assurance de résultats comparables entre les différents patients est réalisée à l'aide de méthodes similaires aux autres études. C'est à dire avec la sélection d'une combinaison de fluorochromes qui sont utilisables par les différents cytomètres. La définition de plusieurs SOPs détaillant la calibration des différentes machines, la définition des différents panels, l'utilisation des billes 8 pics au quotidien, la préparation des échantillons et l'envoi des fichiers sur un serveur de

l'équipe de bioinformatique sont décrits en (*Annexe A.6, page 154*).

Un autre aspect de ce travail multi-centrique est la mise en place d'une collecte de données complémentaires en ligne (flow cytometry report) permettant de renseigner des informations complémentaires comme l'heure de la réalisation du prélèvement, les voltages des instruments au moment de l'analyse ou bien les différents lots utilisés pour chaque inclusion (*Annexe A1, A2, page 98 , 99*).

Là où se démarque PRECISESADS est la suppression d'erreur de pipetage et d'utilisation de réactifs liquides grâce à la technologie duraclone. Un autre point où ce démarque l'étude est une plus grande précision exigée lors de la calibration des différentes machines avec une tolérance dans la variation de l'IMF de $\pm 5\%$ de la valeur cible alors que d'autres études acceptent des variations à 15% et jusqu'à 35% pour certaines.

Duraclone tubes : Ces tubes, développés par Beckman-Coulter, possèdent tous les couples fluorochromes/anticorps d'un panel sous forme sèche. Cela permet de réduire plusieurs problèmes : dans la gestion des lots entre les centres, des problèmes liés à la manipulation en ne nécessitant pas de pipetage de réactifs. De plus, ces tubes peuvent être conservés à température ambiante, à l'abri de la lumière.

Harmonisation : Une phase de calibration a été mise en place s'appuyant sur des billes de capture et des billes 8pics, permettant la mise en miroir des différents cytomètres et le maintient dans le temps. Elle a fait l'objet d'une publication [41] et est détaillée plus loin dans ce document (*Annexe A.4, page 129*).

Une fois la mise en miroir des différents cytomètres faite et l'inclusion des patients standardisée, il reste l'analyse des milliers de fichiers FCS.

Cette analyse des fichiers de cytométrie en flux a pris en compte les différences de résolution entre les constructeurs. La solution est de mettre tous les fichiers sur la même échelle, qui est 0 - 1024, lors de l'analyse et pour l'exploitation de l'IMF. Pour cela les données en 18 bits sont divisées par un facteur 256 et les données en 20 bits sont divisées par un facteur de 1024. Cette échelle a été préférée car elle correspond à celle utilisée dans le logiciel d'analyse KALUZA. Cela permet de comparer les résultats obtenus manuellement et ceux obtenus avec des outils comme R.

Concernant PRECISESADS, le choix a été fait de se concentrer dans un premier temps sur les variations dans les populations cellulaires connues. De ce fait, une liste de populations et de mesures, telles que la fréquence ou l'IMF, ont été sélectionnées pour chaque panel. Pour obtenir tous ces résultats, une analyse automatisée supervisée a été préférée pour l'analyse des milliers de fichiers de cytométrie générés.

Un des algorithmes testés fut OPENCYTO [42]. La particularité de celui-ci est de permettre l'utilisation de plusieurs algorithmes de clustering différents. Cet éventail

d'algorithmes disponibles permet de varier la méthode de clustering en fonction de la "gate" désirée. Par exemple, Opencyto permet d'utiliser flowclust pour la définition de populations. Ces algorithmes de clustering sont déterminés par l'utilisateur dans un tableau où la stratégie de gating est définie (*Figure 21*).

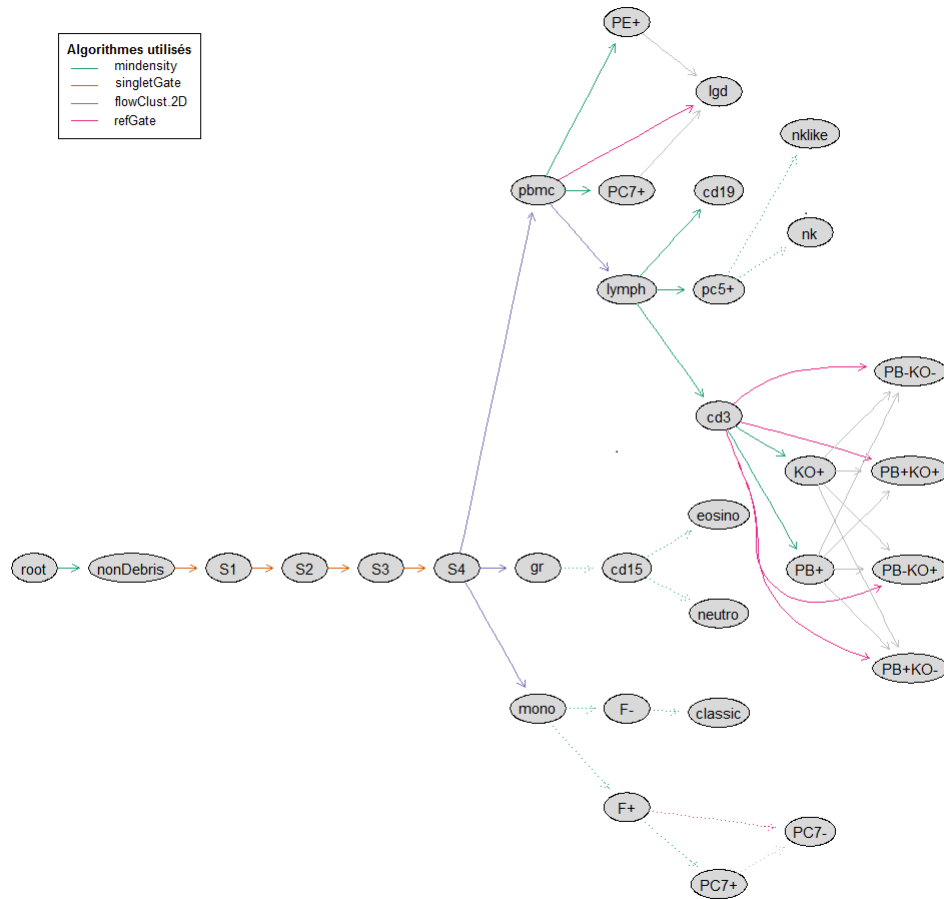


FIGURE 21: **Stratégie d'analyse avec OPENCYTO**. Schéma de la stratégie d'analyse réalisé pour le Panel 1. Les populations (nodes) sont reliées par les différents algorithmes de clustering utilisés (edge) pour définir la population enfant.

L'automatisation de l'analyse, par l'intermédiaire d'OPENCYTO, a permis d'obtenir des résultats encourageant. Mais la détermination des différents algorithmes utilisés prend du temps car implique beaucoup d'essais et d'erreurs. C'est pourquoi, nous avons finalement décidé de soustraire cette analyse avec la société Altrabio (Lyon). Leur méthode utilise du machine-learning pour imiter le gating manuel et produire les données nécessaires à PRECISESADS.

Une fois les données analysées, elles sont envoyées aux membres de PRECISESADS responsables de l'analyse bioinformatique pour reclassifier les SADs. Dans ce

but, il est impératif de leur fournir le plus de données comparables possible entre les différents centres (*Tableau 5*).

	Transversale		Prospective			Total
	PHASE I	PHASE II	M0	M6	M14	
Panel 1	301	2319	215	191	184	3210
Panel 2	301	2319	215	191	184	3210
Panel 3	301		215	191	184	891
Panel 4	301		215	191	184	891
Panel 8	301		215	191	184	891
Panel 6	301		215	191	184	891
Panel 7	301		215	191	184	891
Panel 8	301		215	191	184	891
Panel 9	301		215	191	184	891
Nombre total de fichiers FCS						12657

TABLEAU 5: Tableau récapitulatif du nombre de fichiers de cytométrie générés pour chaque panel dans les 3 cohortes (Transversale phase I, transversale phase II et prospective à M0, M6, M14) de PRECISESADS.

C'est sans compter sur les différences dans les équipements des différents centres. En effet, dans ce projet il y a une grande variation dans les cytomètres disponibles dans les 11 centres (*Tableau 6*). Au total ce sont 6 cytomètres différents provenant de 2 constructeurs (Beckman-Coulter BC et Becton Dickinson BD) qui sont utilisés dans le projet.

TABLEAU 6: Centres impliqués dans le projet PRECISESADS

CENTRE	ACRONYME	PAYS	CYTOMETRE	CONSTRUCTEUR
Université de Bretagne Occidentale	UBO	FRANCE	Navios	Beckman-Coulter
Université de Genève	UNIGE	SUISSE	Navios	Beckman-Coulter
Centro Hospitalar do Porto	CHP	PORTUGAL	Navios	Beckman-Coulter
Institut d'Investigacion Biomedica de Bellvitge	IDIBELL	ESPAGNE	Gallios	Beckman-Coulter
Deutsches Rheuma-Forschungszentrum Berlin	DRFZ	ALLEMAGNE	Canto II	Beckton Dickinson
Istituto di Ricovero e Cura a Carattere Scientifico	IRCCS	ITALIE	Canto II	Beckton Dickinson
Université Catholique de Louvain	UCL	BELGIQUE	Canto II	Beckton Dickinson
Medizinische Hochschule Hannover	MHH	ALLEMAGNE	Canto II	Beckton Dickinson
Parque Tecnológico de Ciencias de la Salud	FPS	ESPAGNE	VERSE	Beckton Dickinson
Hospital Universitario Reina Sofia	SAS	ESPAGNE	Aria	Beckton Dickinson
Katholieke Universiteit Leuven	KUL	BELGIQUE	Fortessa	Beckton Dickinson

Au vu de l'ampleur de l'étude PRECISESADS, avec une grande hétérogénéité dans les cytomètres associée à une inclusion d'un nombre non négligeable d'individus et ce sur une période de 5 ans, des problématiques se sont posées sur la comparabilité des données.

1.3.7 Questions soulevées durant la thèse

Pour faire face à ces problématiques de standardisation des résultats de cytométrie en flux entre les différents centres et sur toute la durée de l'étude, différentes solutions ont été appliquées :

- standardisation des réactifs et des protocoles.
- Mise en miroir des machines et maintien de cette mise en miroir dans le temps.

Mon travail de thèse a consisté à mettre en œuvre des solutions pour assurer la comparabilité des données.

2 RÉSULTATS

2.1 Préparation des données

L'assurance de la comparabilité des données est une étape cruciale de l'étude PRECISESADS. Cela implique de relever de multiples défis pour comparer différentes sources d'informations provenant de populations hétérogènes. Le plus grand défi est celui de comparer des résultats de machines de différents constructeurs mais aussi de maintenir cette comparabilité sur plusieurs années.

La réponse choisie, dans le cadre de PRECISESADS, est de mettre en place une phase de calibration avant l'acquisition des différents patients. Cette calibration est très similaire à celles qui ont pu être décrites dans plusieurs articles précédemment évoqués dans l'introduction [22][24][25][20].

2.2 Calibration

La première étape de la calibration est l'harmonisation des différentes machines entre elles. C'est à dire déterminer les paramètres des cytomètres pour qu'ils produisent, pour un même échantillon, des intensités de fluorescence similaires entre les machines. La deuxième étape est de maintenir cette mise en miroir au quotidien pour contrer les variations qui peuvent survenir avec l'utilisation normale d'un cytomètre.

2.2.1 Harmonisation

Pour l'harmonisation des différents cytomètres, il a été choisi d'utiliser les VersaComp Capture Beads (Beckman coulter). Ce kit de billes est composé de 2 types de billes, dites positives et négatives. Les billes de capture (positives) accrochent les anticoprps couplés aux fluorochromes en suspension et permettent d'obtenir des intensités de fluorescence médianes pour chacun des 8 canaux. Les billes négatives n'accrochent rien et permettent d'obtenir une fluorescence similaire à celle de cellules non marquées.

L'attrait de ces billes est leur stabilité et leur capacité à être spécifiques du cocktail de fluorochromes utilisé pour l'acquisition des patients. Dans le cadre de l'harmonisation, ces billes sont utilisées avec un kit de compensation Duraclone. Ce kit, fourni avec l'achat des tubes duraclones multi-couleurs, est composé de 8 tubes duraclone où chacun des tubes possède un seul des 8 fluochromes du panel 1.

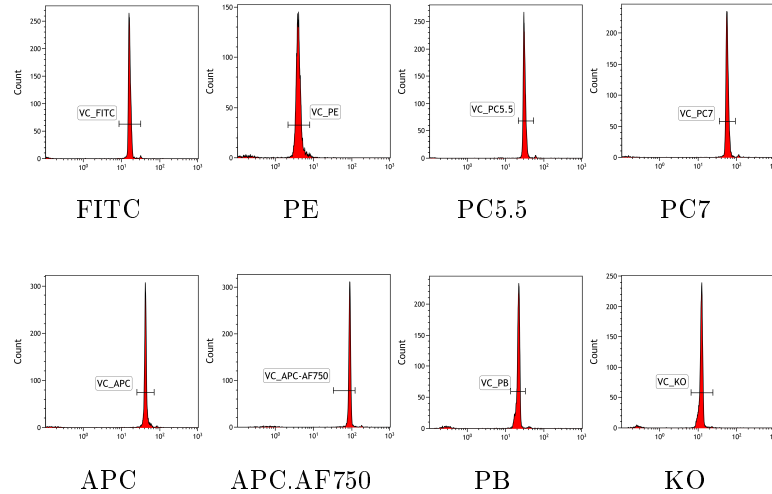


FIGURE 22: **Références de calibration.** Définition des valeurs de référence faite au centre référent pour chaque fluorochrome (a-h) en utilisant des billes VersaComp de capture marquées avec le kit de compensation duraclone du panel 1.

Dans un premier temps, la valeur cible en IMF pour chaque canal est déterminée en passant les versaComp dans le cytomètre du centre UBO (NAVIOS). C'est ce centre qui sert de centre de référence pour l'ensemble du projet (*Figure 22*).

Une fois ces valeurs déterminées, la calibration est réalisée dans chacun des centres avec les mêmes lots de réactifs (*Figure 23*). Cette fois, chaque centre doit modifier les voltages des PMT pour obtenir une IMF dans une fourchette acceptable et cela pour chacun des 8 canaux. La fourchette acceptable a été définie arbitrairement à $\pm 5\%$ de la valeur cible du centre de référence afin d'éviter au maximum les variations entre les centres (*Figure 24*).

Par exemple, pour le canal FITC, la valeur cible est une IMF de 16,22 déterminée à l'UBO. Les différents centres doivent modifier leurs voltages pour obtenir une valeur en IMF comprise entre 15,40 et 17,03. Pour les centres possédant un cytomètre de chez BD les cibles ont été transformées pour correspondre à leur échelle, pour cela les valeurs ont été multipliées par 256. Cela nous donne une fourchette pour la IMF comprise entre 3942,4 et 4359,68.

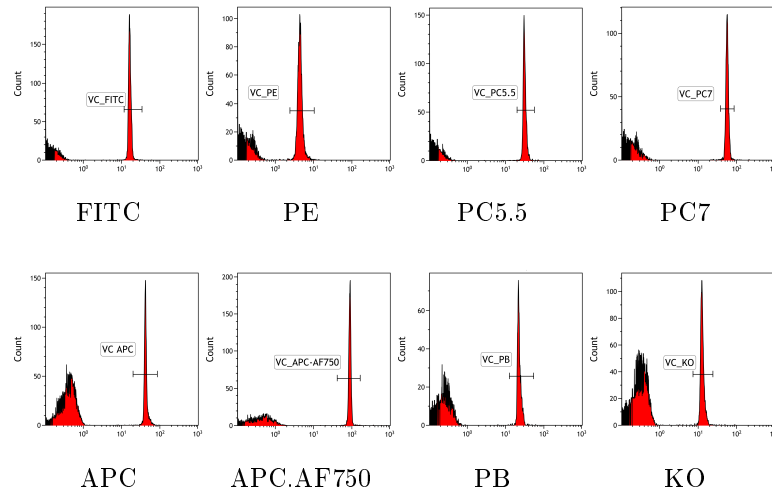


FIGURE 23: **Résultat de la calibration.** Exemple de calibration réalisée dans le centre DRFZ pour chaque fluorochrome (a-h) en utilisant les mêmes billes Versa-Comp de capture marquées avec le même kit de compensation duraclone du panel 1 que le centre référent UBO.

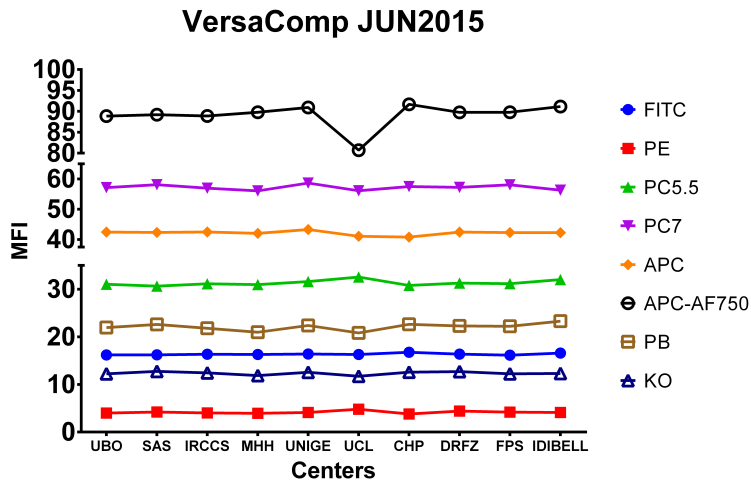


FIGURE 24: **Première calibration.** Médianes des intensités de fluorescence pour chaque fluorochrome et chaque centre lors de la première calibration inter-centre. Le centre UCL a échoué à calibrer le canal APC-AF750.

	FITC	PE	PC5.5	PC7	APC	APC-AF750	PB	KO
Coefficient of variation	1.13%	6.65%	1.85%	1.55%	1.71%	3.47%	3.42%	2.75%

Une fois les voltages des PMT définis par les billes de capture pour l'ensemble des canaux, des billes à 8 pics de fluorescence sont passées. Les IMF de ces billes 8 pics servent de référence propre à chaque centre pour maintenir cette calibration dans le temps (*Figure 25*).

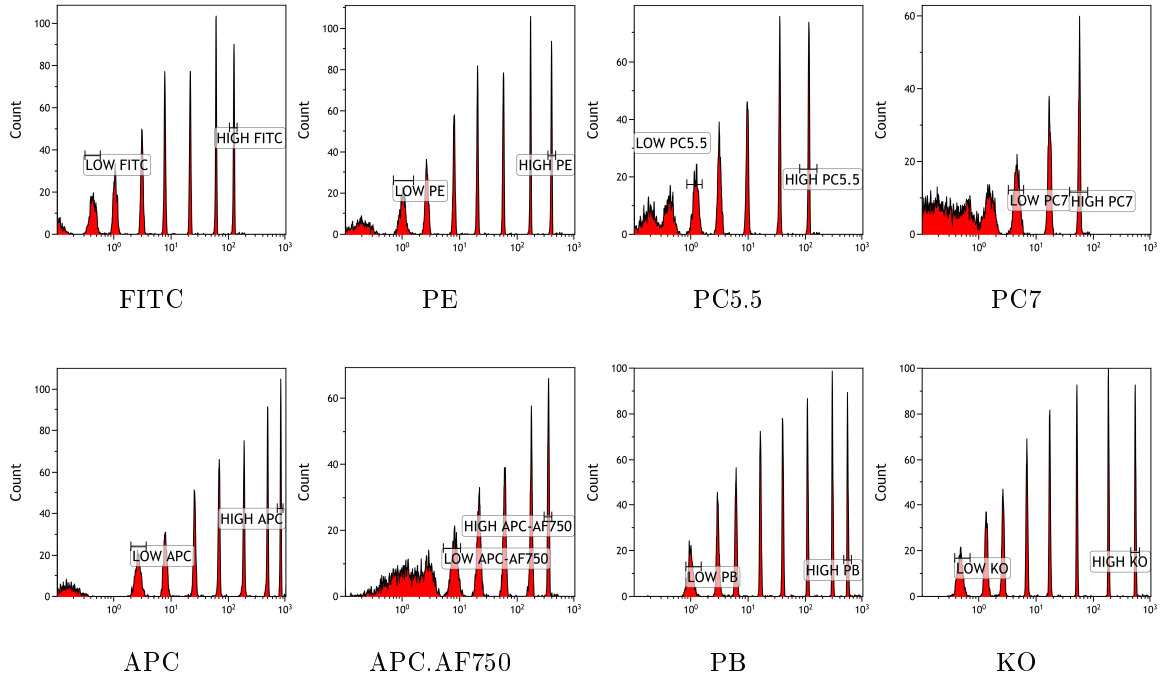


FIGURE 25: **Billes à 8 pics de fluorescence.** Exemple de profils de billes à 8 pics de fluorescence passées dans le centre UBO après la première calibration. Les pics sélectionnés comme "low" et "high" sont utilisés comme références pour les acquisitions des patients de PRECISESADS.

Les valeurs de référence des billes 8 pics au quotidien correspondent aux pics "low" et "high" pour chaque canal (*Figure 25*).

Une fois la première harmonisation des différents cytomètres validée et les références pour les billes 8 pics définies pour chaque instrument, l'inclusion des patients et témoins sains pour PRECISESADS a pu commencer.

2.2.2 Calibration au cours du temps

Une calibration des différentes machines a eu lieu tous les 4 mois environ et ce tout au long des 5 années de l'étude (*Figure 26*).

Moyenne des IMF des billes de capture versacomp des 11 centres de PRECISESADS

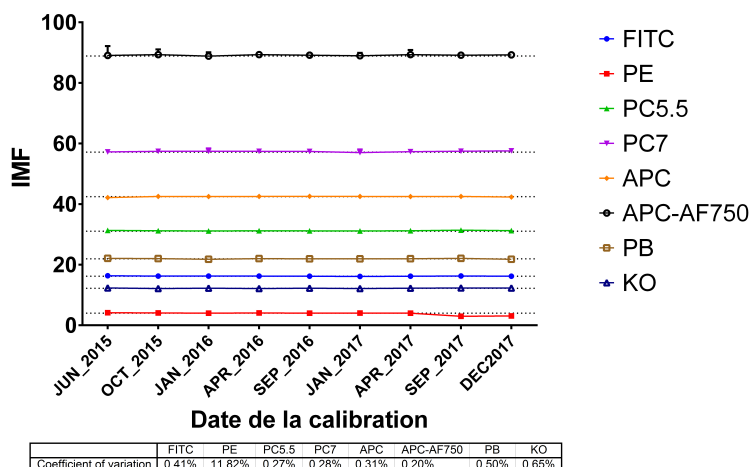


FIGURE 26: La calibration au cours du temps. Évolution des différentes calibrations au cours du temps, représentées sous forme de moyenne des médianes des intensités de fluorescence entre les centres et leur déviation standard, pour chaque fluorochrome.

Les différents centres de l'étude ont réalisé de multiples calibrations en gardant cette fourchette d'acceptabilité de $\pm 5\%$ de l'IMF cible. Cela nous permet d'avoir plusieurs points dans le temps auxquels sont rattachés des références pour les billes 8 pics propres à chaque centre. Le plein potentiel de ces références est atteint lors de l'acquisition d'un patient.

2.2.3 Billes 8 pics

Avant l'acquisition de chaque patient, les billes 8 pics sont passées dans le cytomètre (Figure 27) et, si besoin, les différents pics sont réalignés à la référence du centre. Ce réalignement se fait par la modification des voltages des PMT pour que l'IMF des pics "high" et "low" de chaque canal soit dans la fourchette des $\pm 5\%$ de la référence. Une fois validés, les nouveaux PMT ajustés sont utilisés pour l'acquisition des différents panels.

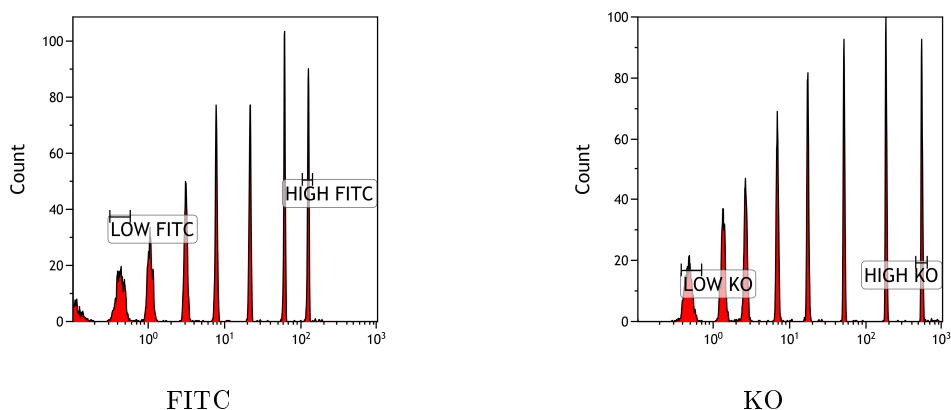


FIGURE 27: Définition de la référence des billes 8 pics. Exemple de la sélection des pics pour les billes 8 pics sur le FITC(a) et le KO(b) pour le centre UBO.

L'objectif de ce réaligement est de mettre le cytomètre dans les conditions d'IMF similaires à celles de la calibration. Mais, il arrive que l'alignement échoue et que les billes 8 pics soient enregistrées malgré un décalage par rapport à la référence dans un des canal. Ce décalage va entraîner un décalage similaire de l'IMF des panels pour le canal concerné (*Figure 28*).

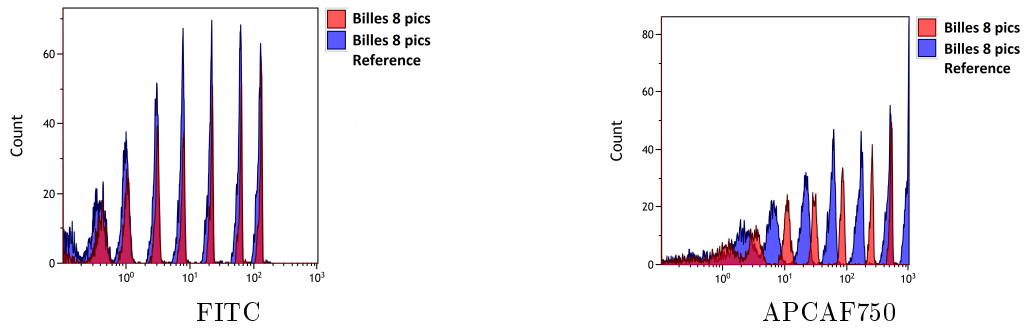


FIGURE 28: **Décalage dans les billes 8 pics.** Exemple d'un bon alignement entre les profils des billes 8 pics (en rouge) et leurs références (en bleu) pour le canal FITC (a) et d'un mauvais alignement avec un décalage pour le canal APC-AF750(b) pour le même patient.

Pour assurer l'exploitation des données, il est nécessaire de corriger dans les panels la variation observée par le décalage des billes 8 pics. Pour cela on réalise une normalisation des données.

2.3 Normalisation

2.3.1 Normalisation existante

Des méthodes de normalisation, comme `fdanorm` ou `gaussnorm`, sont déjà disponibles dans le package R `FlowStats`. Leur stratégie est de réaligner l'intensité de fluorescence des différentes populations entre elles dans chacun des canaux [43]. L'avantage de cette normalisation est de stabiliser la position des populations entre les individus pour l'analyse automatisée. Mais cela rend l'exploitation de l'IMF impossible car l'alignement des populations réduit toutes les variations, qu'elles soient d'origine technique ou biologique.

Une autre solution fut proposée par Schwartz [44] puis implémentée dans le package R `flowbeads` [45]. Cette solution combine l'utilisation de l'unité des molécules de fluorochrome d'équivalence soluble (MESF) et la régression linéaire. C'est en se basant sur cette normalisation avec MESF que j'ai développé la normalisation réalisée dans `PRECISESADS` mais en utilisant l'IMF des billes 8 pics.

2.3.2 Normalisation Intra-Centre

La normalisation intra-centre, développée pour PRECISESADS, utilise les billes 8 pics acquises quotidiennement et leurs références (déterminées à chaque étape de calibration). L'idée de départ est que, pour le même cytomètre, le même lot de billes doit produire la même IMF au cours du temps. Si des variations sont observées, elle doivent avoir pour origine un problème technique ou une erreur lors du paramétrage de la machine. Une solution pour corriger ces variations est de déterminer des paramètres permettant de ramener l'IMF des différents pics des billes 8 pics à celle de leur référence. Une fois ces paramètres de normalisation déterminés, ils peuvent être appliqués aux différents panels associés aux billes.

Le script de normalisation prend les données des billes 8 pics et de leur référence pour chaque canal (*Annexe A.7.1, page 258*). C'est à dire qu'il utilise l'IMF de tous les pics du canal à normaliser pour définir une droite de régression dont il est possible d'extraire les paramètres d'intercepte et de coefficient. Ce sont ces paramètres qui permettent de réaligner l'IMF des billes 8 pics avec celle de référence (*Figure 29*)

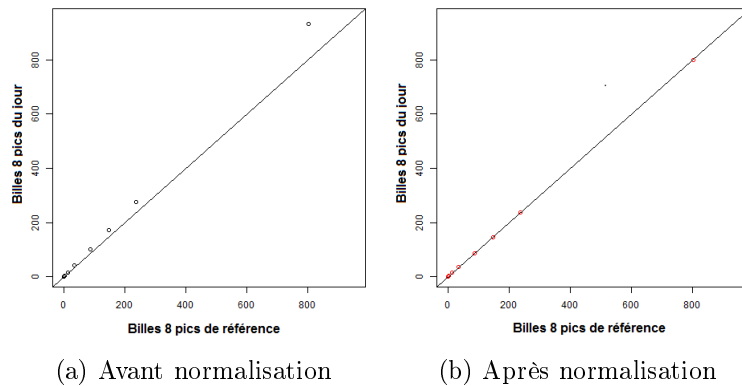


FIGURE 29: **Normalisation des billes 8 pics.** Exemple de la normalisation sur l'IMF des différents pics en APC entre des billes ayant un décalage avec leur référence (a) et une fois les valeurs normalisées en rouge (b). La droite représente les valeurs où $X = Y$.

Une fois les paramètres déterminés pour chaque canal, ils sont appliqués avec la formule suivante :

$$Y = \alpha X + \beta$$

avec X l'IMF à transformer, Y l'IMF normalisée et α le coefficient et β l'intercepte de la droite formée avec la référence et les billes 8 pics (*Figure 29a*).

L'application de la normalisation se fait directement sur les données brutes des fichiers de cytométrie et ceci avant toute compensation ou toute transformation.

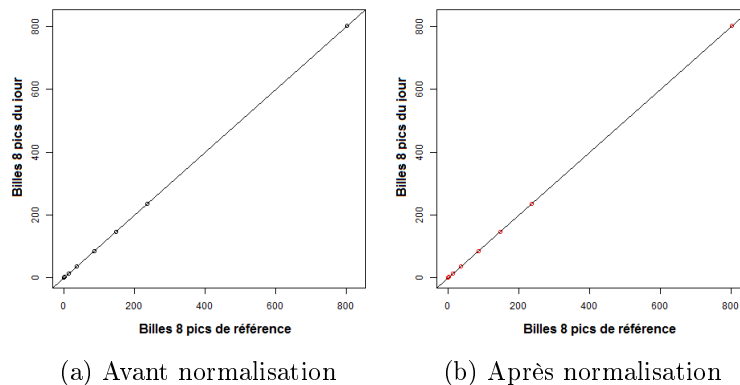


FIGURE 30: **Normalisation des billes 8 pics.** Exemple de la normalisation sur l'IMF des différents pics en APC entre des billes n'ayant pas de décalage. La droite représente les valeurs où $X = Y$.

Les premiers tests ont permis de ramener les billes 8 pics à des valeurs proches de leur référence (*Figure 29b*). La normalisation a aussi été testée sur des billes très proches de leurs valeurs de référence pour vérifier s'il était possible d'appliquer la normalisation à toutes les inclusions et ceci sans risque de modifier des canaux parfaitement alignés. Pour ces cas là, l'application du script de normalisation n'a pas changé les valeurs (*Figure 30*).

La normalisation est efficace pour corriger le décalage dans les billes 8 pics par rapport à la référence, mais un problème est apparu dans la validation de cette normalisation pour les panels. En effet, les billes 8 pics possèdent une référence sur laquelle on peut valider la normalisation, ce qui n'est pas le cas des panels.

2.3.3 Validation de la normalisation

Afin de tester l'efficacité de cette normalisation, il nous a semblé nécessaire de pouvoir comparer un échantillon en minimisant toutes les sources de variations. Pour cela, un échantillon de sang est passé dans la même machine plusieurs fois avec des voltages différents, modifiés intentionnellement.

L'idée de ce test de normalisation est, dans un premier temps, de déterminer plusieurs jeux de voltages différents avec les billes 8 pics. Un jeu de voltage où l'IMF des pics "low" et "high" de chaque canal arrive sous les 5% de variations acceptable, puis un autre entre 5 et 10% de variations et un dernier supérieur à 10% de variations par rapport à la référence. Chaque jeu de voltage des billes 8 pics est passé 3 fois au cytomètre. Notre attention se portera principalement au jeu de voltage de moins de 5% et supérieur à 10% dans la suite de ce document (*Figure 31*).

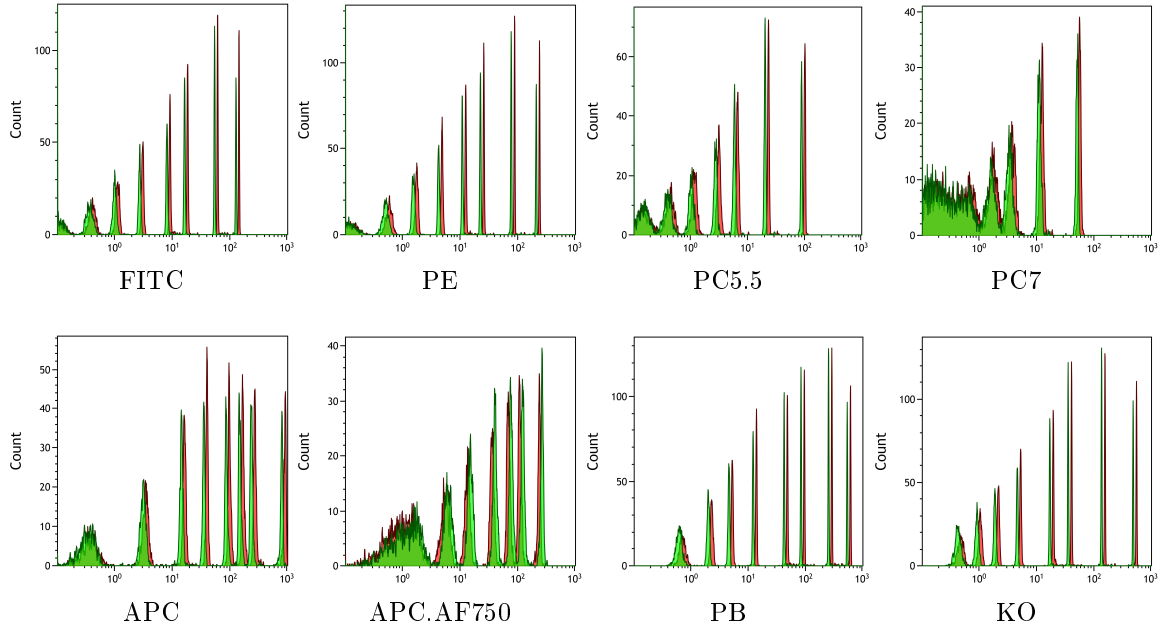


FIGURE 31: **Comparaison des billes 8 pics.** comparaison des aspects des billes 8 pics entre la référence (vert) et celles avec un décalage supérieur à 10% du centre UBO et ceci pour chacun des 8 fluorochromes.

CV(%)	FITC high	PE high	PC5.5 high	PC7 high	APC high	APC AF750 high	PB high	KO high
sous 05percent	0.1	0.18	0.06	0.27	0.14	0.17	0.18	0.11
sous 10percent	0.05	0.08	0.07	0.03	0.11	0.35	0.21	0.22
supérieur 10percent	0.66	0.55	0.54	0.8	0.49	0.65	0.69	0.56

TABLEAU 7: **Variation de l'IMF des billes 8 pics.** Coefficient de variation (CV), en pourcentage, des 3 passages pour le pic "high" de chacun des 8 canaux.

Les 3 passages de billes 8 pics à voltage constant nous permettent de voir que la variation entre différents passages à voltage constant est faible. Les variations observées sont inférieures à 1% en IMF (Maximum à 0.8 en PC7). Cela veut dire que les variations que l'on pourrait attendre de multiples passages est inférieure à 1% (Tableau 7).

Lorsque son comparées les IMF entre les différents jeux de voltage, on s'intéresse plus particulièrement au premier passage du jeu inférieur à 5% et à celui supérieur à 10%. Malgré l'acceptabilité des billes inférieures à 5%, il est à noter qu'elles possèdent quand même une variation par rapport à la référence. Cette variation peut aller jusqu'à 2.37% en PC7 dans l'exemple montré (Tableau 8).

En comparant les voltages des PMT, on s'aperçoit qu'une différence entre 8 et 10 volts suffit à passer d'un décalage acceptable ($<5\%$) à un décalage supérieur à 10% (*Tableau 8*).

Après ces premières observations, le script de normalisation a pu être appliqué sur les différentes acquisitions des billes 8 pics.

Canal	Référence	Moins de 5%	Variation(%)	PMT	Supérieur à 10%	Variation(%)	PMT	Supérieur à 10% normalisé	Variation(%)
FITC	127.33	129.27	1.52	421	145.67	14.40	429	127.32	-0.01
PE	210.55	212.37	0.86	426	241.98	14.93	435	210.51	-0.02
PC5.5	89.86	89.41	-0.50	466	102.29	13.83	476	89.92	0.07
PC7	50.3	51.49	2.37	556	57.07	13.46	565	50.32	0.04
APC	800.65	806.83	0.77	515	920.08	14.92	525	800.01	-0.08
APC-AF750	237.89	239.75	0.78	634	267.72	12.54	645	238.26	0.16
PB	532.1	542.65	1.98	412	618.2	16.18	421	531.96	-0.03
KO	481.47	487.94	1.34	372	558.72	16.04	380	481.46	0.00

TABLEAU 8: Normalisation de l'IMF des billes 8 pics. Valeurs des IMF, voltages et variations pour le pic "high" des 8 canaux pour le premier passage sous les 5 %, pour le premier passage supérieur à 10% et pour le premier passage supérieur à 10% après normalisation sur la référence.

La normalisation des billes 8 pics a réduit fortement la différence des IMF avec une variation supérieure à 10% qui passe à une variation inférieure à 1% (maximum à 0.16%). Par exemple, la variation passe de 12,54% en APC-AF750 à une variation de 0.16% après normalisation avec la référence (*Tableau 8*).

Une fois la normalisation validée sur les billes 8 pics, le sang d'un individu a été préparé avec le panel 1 et passé au cytomètre avec les 3 jeux de voltages déterminés précédemment. Les données sont analysées manuellement à l'aide du logiciel Kaluza.

	Lymphocyte	Monocytes	Granulocyte	T	T CD4+CD8-	T CD4+CD8+	B	Nkcells	eosino	neutro
sous 05percent 01	29.34%	5.95%	61.43%	21.13%	12.58%	6.70%	2.39%	5.06%	2.42%	58.98%
sous 05percent 02	28.98%	6.23%	61.84%	21.23%	12.47%	6.85%	2.19%	4.84%	2.39%	59.44%
sous 05percent 03	29.01%	6.01%	61.96%	21.08%	12.70%	6.67%	2.17%	5.09%	2.42%	59.52%
supérieur 10percent	29.40%	6.14%	61.08%	21.51%	12.59%	7.04%	2.28%	5.05%	2.44%	58.60%

TABLEAU 9: Fréquences des populations du panel 1. Fréquences, en pourcentage de population totale (avec exclusion des doublets et autres débris), des différents passages pour le panel 1 (3 fois avec le jeu de voltage inférieur à 5% et 1 fois avec le jeu supérieur à 10%).

Lorsque sont comparées, les fréquences des différentes populations cellulaires, que cela soit avec des voltages constants ou non, sont très proches les unes des autres et elles ne semblent pas être affectées par la modification des voltages (*Tableau 9*).

Pour la comparaison des IMF, il a été choisi d'utiliser les résultats du premier passage du panel 1 à moins de 5% comme référence pour la normalisation du panel 1 supérieur à 10%.

Canal	Moins de 5%	Supérieur à 10%	Variation(%)	Supérieur à 10% normalisé	Variation(%)
FITC	83.31	84.18	1.03	73.61	11.64
PE	67.75	75.58	10.36	65.68	3.06
PC5.5	8.75	10.07	13.11	8.86	-1.26
PC7	209.53	250.76	16.44	221.46	-5.69
APC	29.23	32.78	10.83	28.51	2.46
APC-AF750	104.84	121.68	13.84	107.98	-3.00
PB	14.76	16.54	10.76	14.23	3.59
KO	21.19	22.27	4.85	19.14	9.67

TABLEAU 10: **Normalisation de l'IMF du panel 1.** Valeurs des IMF et des variations pour la population pour chacun des 8 canaux pour le premier passage sous les 5 %, le premier passage supérieur à 10% et le premier passage supérieur à 10% normalisé sur la référence.

L'application de la normalisation, sur les IMF du panel 1 avec les voltages supérieur à 10%, réduit ces variations à des niveaux acceptables (inférieur à 5%) pour la majorité des populations cellulaires (*Tableau 10*).

Il y a des populations dont les variations entre les différents passages, donc avec les mêmes valeurs de voltages, sont supérieures aux variations observées pour les billes 8 pics. Par exemple, pour le marqueur CD16 FITC sur les neutrophiles dont on peut observer une diminution de l'IMF au cour des passages avec les mêmes voltages (passage 1 : 83.31, Passage 2 : 79.9, Passage 3 : 78.06). Alors qu'une autre population, les éosinophiles, sont plus stables en CD16 FITC (2.03, 2.05, 2.05). Ces variations semblent être dues à la population cellulaire en elle-même et non à des problèmes techniques [46].

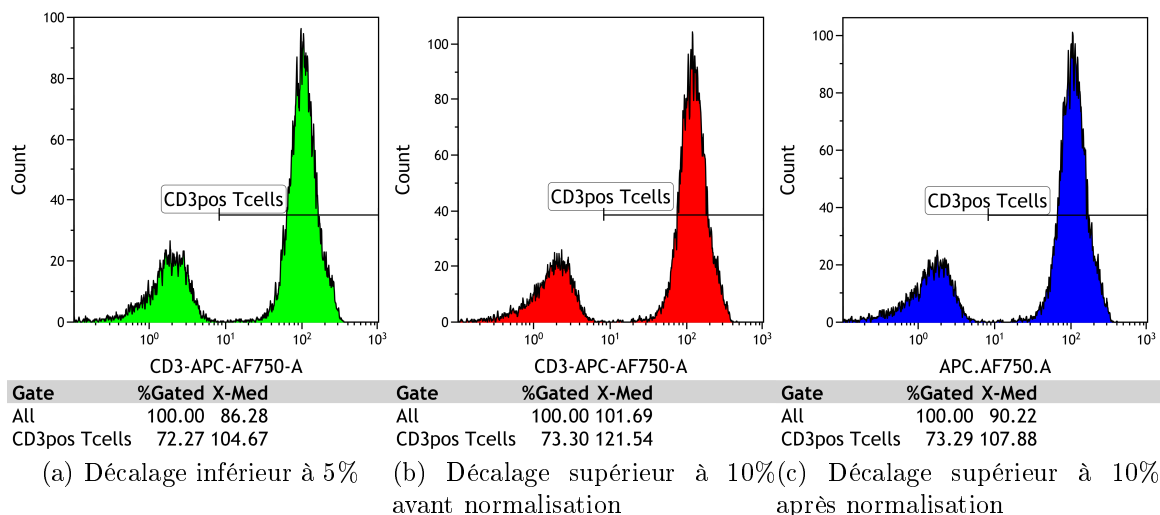


FIGURE 32: **Comparaison des populations cellulaires.** Comparaison entre les différents décalages dans les billes 8 pics, décalage inférieur à 5%(a), décalage supérieur à 10% (b), et leur impact sur l'IMF de la population des lymphocytes T CD3+, et l'impact de la normalisation (c)

Pour conclure, la normalisation avec le script R permet de corriger les mauvais alignements engendrés par des erreurs humaines et même les variations tolérées. En effet, même les variations inférieures à 5% diminuent avec le script de normalisation sur les billes 8 pics. L'application sur le panel (*Figure 32*) bien que ne permettant pas d'atteindre forcément le seuil des moins d'1% de variation, permet de réduire cette dernière à un niveau acceptable.

Bien que fonctionnelle, la normalisation nécessite l'extraction des différentes IMF des pics des billes 8 pics et leurs références, le calcul des paramètres et l'application de ces derniers sur les fichiers de cytométrie du patient. Ces opérations devront être faites pour les 3210 inclusions de PRECISESADS. Pour gagner du temps, une forme d'automatisation de la normalisation a été mise en place.

2.4 automatisation de la normalisation

L'automatisation de la normalisation se fait grâce à un mélange de scripts impliquant python et R. Le premier permet de récupérer tous les fichiers et de déterminer quelle référence est à utiliser avec les billes 8 pics. Le second (*Annexe A.7.1*, page 258), quant à lui, permet de manipuler les fichiers de cytométrie, d'extraire l'IMF des pics, de calculer les paramètres du modèle linéaire, de l'appliquer et de créer un fichier avec les données normalisées.

2.4.1 clustering 2D et 1D

La première étape de l'automatisation de la normalisation est la suppression des débris et des doublets. Pour cela on utilise la fonction `lymphGate` du package `flowStats` [14]. Cette fonction permet de définir une gate elliptique sur la zone de plus forte densité en taille/structure.

Une fois la population de billes sélectionnée, il est nécessaire d'appliquer une transformation car les données sont stockées sur une échelle linéaire alors qu'une échelle logarithmique est plus appropriée pour la visualisation des différentes populations. Il existe plusieurs transformations possibles mais celle que l'on a choisi ici est l'`arcsinh`, qui est appliquée aux différents canaux.

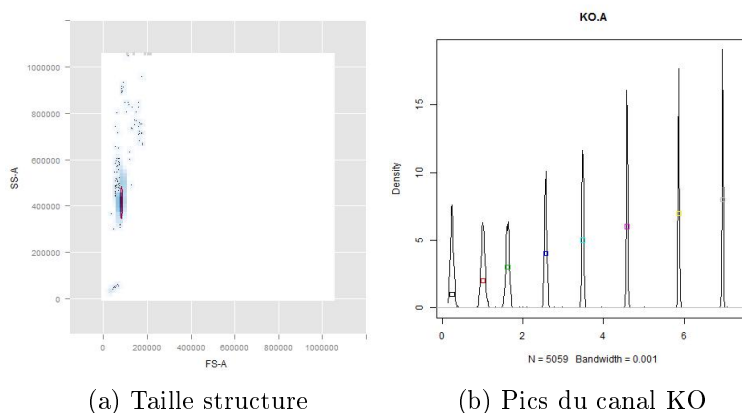


FIGURE 33: **Automatisation de l'extraction de l'IMF.** Exemple des 2 méthodes de clustering utilisées pour extraire l'IMF des billes 8 pics. Cela commence avec un clustering en fonction de la densité en taille / structure (a) suivi, pour chaque fluorochrome, d'un clustering à l'aide de l'algorithme `CKmedian.1D` (b).

Suite à cette transformation, les différents pics sont clusterisés par `CKmedian.1D` [47]. Ce package R permet la définition de groupes en 1 dimension. Le nombre de groupes/pics varie en fonction du canal et de la résolution de la machine mais reste autour de 8. Les médianes des différents pics sont utilisées pour la normalisation (*Figure 33*). Les différentes médianes (quotidiennes et références) sont envoyées à la partie normalisation du script.

2.4.2 Script de Normalisation

A partir des 2 jeux de médianes, le modèle de normalisation est construit en utilisant la régression linéaire disponible sous R. Ce modèle nous permet de calculer les paramètres de normalisation, le coefficient (α) et l'intercepte (β), de la droite formée entre les médianes des billes quotidiennes et celles de leur référence.

La définition de la régression linéaire est réalisée avec R, suivant la formule suivante :

$$Y = \alpha X + \beta$$

Cela permet de définir des paramètres α et β différents pour chacun des 8 canaux. Cette formule est ensuite appliquée à l'intensité de fluorescence des différents panels de cytométrie avec X l'intensité brute et Y la nouvelle intensité normalisée. Une fois tous les canaux normalisés, un nouveau fichier fcs est enregistré.

L'avantage de cette méthode est qu'elle est applicable sur tous les fichiers de cytométrie. En effet, plus les fichiers seront proches de la référence, plus α sera proche de 1 et β de 0.

De plus, un contrôle a été mis en place et les paramètres de normalisation sont aussi appliqués aux billes 8 pics. Leur IMF est extraite à nouveau pour être comparée à celles de la référence. Si la variation observée est supérieure à 5%, le fichier est mis de côté pour connaître les raisons de cet échec de normalisation et l'analyser manuellement si besoin.

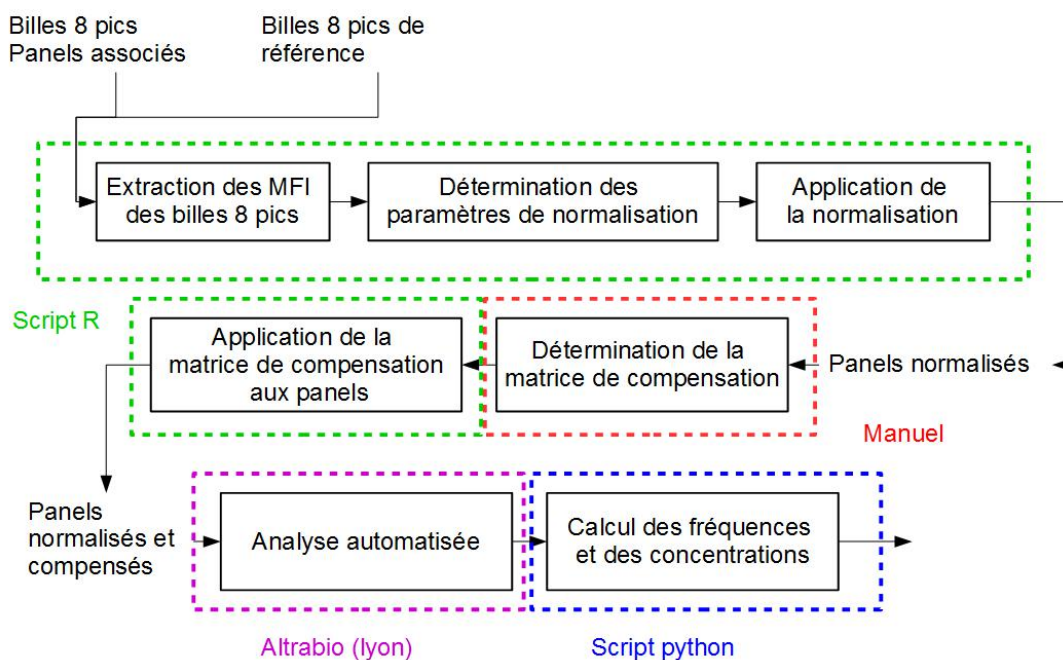


FIGURE 34: **Workflow**. Schéma représentant l'analyse des données de cytométrie.

L'automatisation de la normalisation entre dans le cadre d'une automatisation globale de l'analyse des données de cytométrie en flux. Cette automatisation globale fait intervenir des scripts réalisés sous R et sous Python. De plus, il fait intervenir

une entreprise extérieure pour la partie analyse des données de cytométrie (*Figure 34*).

L'automatisation de la normalisation permet de faire passer les milliers de fichiers de billes 8 pics plus rapidement et de générer les milliers de fichiers de cytométrie normalisés. Une fois les différents panels normalisés, l'analyse des données de cytométrie peut commencer.

2.5 Analyse des données

2.5.1 Compensation

Avant l'analyse des données de cytométrie, il est nécessaire de compenser tous les fichiers comme présenté dans l'introduction. Dans le cadre du projet PRCISESADS, le choix a été fait de réaliser la compensation manuellement sur chacun des fichiers après leur normalisation. La matrice de compensation est dans un premier temps déterminée par l'intermédiaire de Kaluza. Puis, elle est appliquée directement sur les fichiers de cytométrie en utilisant les outils disponibles dans le package flowCore sous R.

Pour évaluer la variation induite par la détermination manuelle de la compensations, 110 fichiers du panel 1 ont été compensés une deuxième fois et leurs matrices ont été comparées à celles réalisées la première fois plusieurs mois auparavant (*Figure 35*).

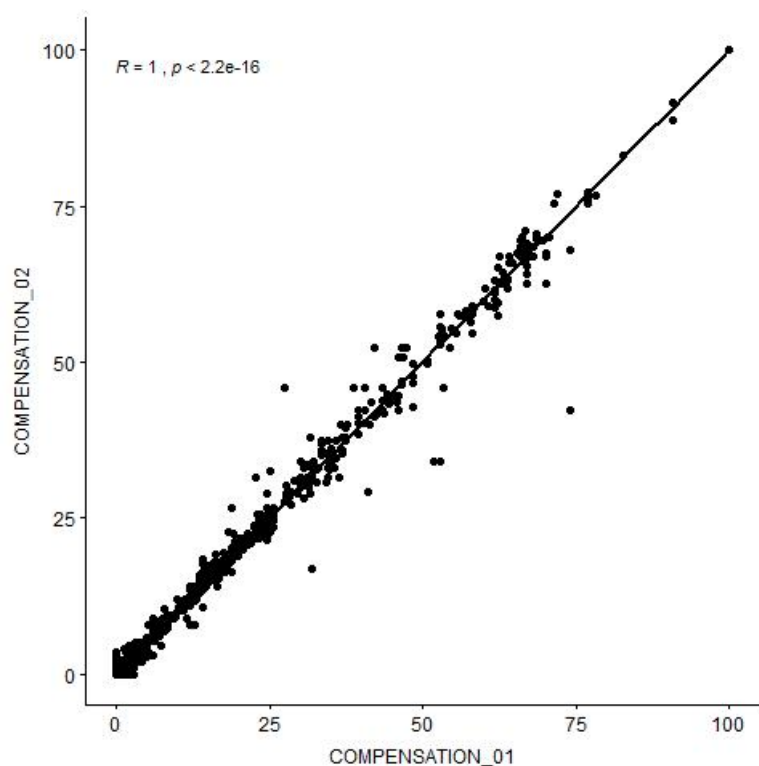


FIGURE 35: **Comparaison de la compensation.** Corrélacion entre la compensation faite pour le panel 1 au début du projet PRECISESADS (COMPENSATION_01) et celle faite à la fin du projet (COMPENSATION_02) pour 110 fichiers de cytométrie (10 fichiers sélectionnés aléatoirement dans chacun des centres).

Le score de corrélation est proche de 1, indiquant que la détermination de la compensation, par un opérateur unique, est stable dans le temps.

Pour ce projet, environ 12000 matrices ont été déterminées manuellement et appliquées sur les fichiers de cytométrie permettant ainsi l'analyse des données.

2.5.2 Analyse manuelle

L'analyse manuelle s'est concentrée sur 300 individus de la phase I avec 9 panels par individus. Ces 2700 fichiers de cytométrie ont été normalisés puis compensés comme décrit précédemment. Cette analyse manuelle a deux objectifs. Elle sert pour la préparation de l'automatisation par la création d'un jeu d'entraînement et d'un jeu de validation sur lesquels l'automate pourra apprendre. Le deuxième objectif est de permettre la réalisation d'une première exploration des données regroupant des témoins sains et des patients de chaque SAD.

Cette analyse a été faite avec Kaluza par un opérateur unique. La stratégie d'analyse de chaque panel se base sur le SOP développé pour PRECISESADS. Les données extraites de ces fichiers sont l'IMF et le nombre d'événements des populations. C'est à partir de ce nombre d'événements que les fréquences sont calculées. Elles correspondent au ratio de la population sœur sur la population parent (par

exemple la proportion de cellules CD4+ dans la population CD3+). La deuxième donnée, obtenue à l'aide du nombre d'évènements, est la concentration cellulaire.

La concentration cellulaire dans le sang est calculée à l'aide de billes FlowCount de comptages (Beckman Coulter) qui sont rajoutées dans les panels 1 et 2. Ce nouveau type de billes possède une concentration (en μL) connue.

Dans le SOP de la réalisation d'une acquisition pour PRECISESADS, il est indiqué que le même volume de billes que de sang est ajouté au tube de duracclone. Cela permet d'obtenir la concentration des populations en suivant l'équation suivante :

$$\text{Concentration cellulaire} = \frac{\text{Nombre d'évènements de la population}}{\text{Nombre d'évènements des billes}} \times \text{Concentration des billes} \quad (3)$$

avec le nombre d'évènements de la population, le nombre d'évènements des billes et la concentration des billes.

L'analyse manuelle, bien que coûteuse en temps, a permis de générer des milliers de données de fréquences, concentrations et IMF qui peuvent être exploitées.

2.5.3 Observation des résultats de 300 patients

L'exploitation des résultats des 300 patients a commencé par le panel 1 et 2 car ces panels concernent le plus grand nombre de patients et des populations bien définies. Le but de l'observation est de rechercher la présence éventuelle d'un groupement des patients qui se ferait principalement en fonction de leur centre d'origine, ou appelé effet centre.

Les centres incluent des patients appartenant aux différentes classes de maladies mais aussi des témoins sains (*Tableau 11*). Par cette hétérogénéité d'inclusion dans chaque centre nous avons considéré que les variations observées dans un centre, entre les patients (et les témoins sains), doivent être similaires et qu'il ne devrait pas y avoir d'effet centre dû à un déséquilibre dans l'inclusion.

CENTRE	LED	ScS	PR	SGS	SALP primaire, CM et CI	Témoins	TOTAL
UBO	23	13	46	66	38	41	227
IRCCS	74	102	41	46	73	49	385
DRFZ	96	113	81	58	68	102	518
UNIGE	26	21	25	30	22	41	165
KUL	18	28	41	16	14	24	141
UCL	46	32	21	6	17	0	122
IDIBELL	52	20	35	20	41	64	232
SAS	29	37	33	43	33	51	226
FPS	82	59	43	94	74	209	561
CHP	41	11	25	21	24	36	158
MHH	26	8	27	29	13	32	135

TABLEAU 11: **Synthèse des inclusions.** Effectifs des patients et des témoins pour chaque centre, ayant servi à la réalisation d'une analyse manuelle des fréquences et des IMF souhaitées. Les patients sont atteints de lupus érythémateux disséminé (LED), de Sclérodémie systémique (SSC), de Polyarthrite rhumatoïde (PR), Syndrome de Gougerot Sjögren (SGS) ou bien appartiennent au groupe des connectivites autres comprenant le syndrome antiphospholipide primaire (SAPL primaire), la connectivite mixte (CM) et la connectivite indifférenciée (CI).

Pour tester la présence d'un effet centre, une APC a été réalisée sur les fréquences et une autre sur l'IMF des patients de la phase 1 plus particulièrement sur les données du panel 1 qui servent d'exemple.

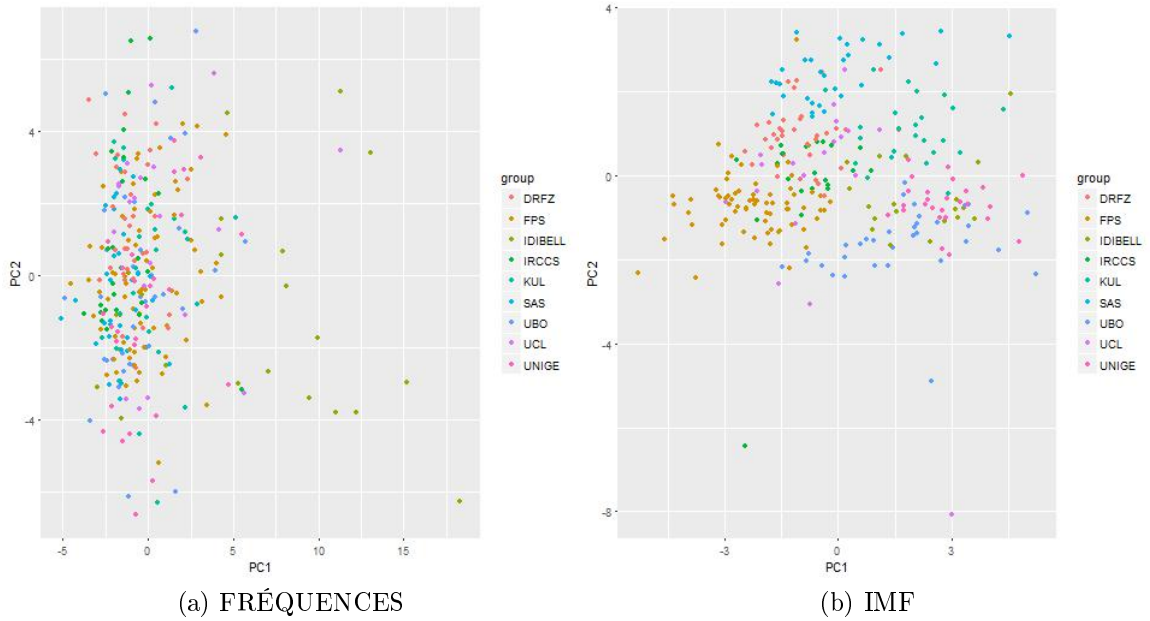


FIGURE 36: **Résultat de l'analyse manuelle.** ACP sur les fréquences du panel 1 des 300 patients de la Phase I (a) et sur leur IMF (b).

Lorsque que l'on explore, visuellement, les résultats par ACP sur les fréquences des différentes populations, il n'y a pas de cluster qui se démarque en fonction d'un centre. En revanche l'exploration des IMF par ACP met en évidence des clusters qui semblent être définis en fonction du centre d'origine (*Figure 36*).

Pour aller plus en avant dans la compréhension de cet effet centre, chaque marqueur a été regardé individuellement. Pour déterminer la présence d'un effet centre, le test de Kruskal-Wallis a été choisi pour tester si les populations dans les différents centres sont issues de la même distribution ou s'il fallait rejeter cette hypothèse. Ce test a été réalisé sur différentes populations dans chacun des 8 canaux.

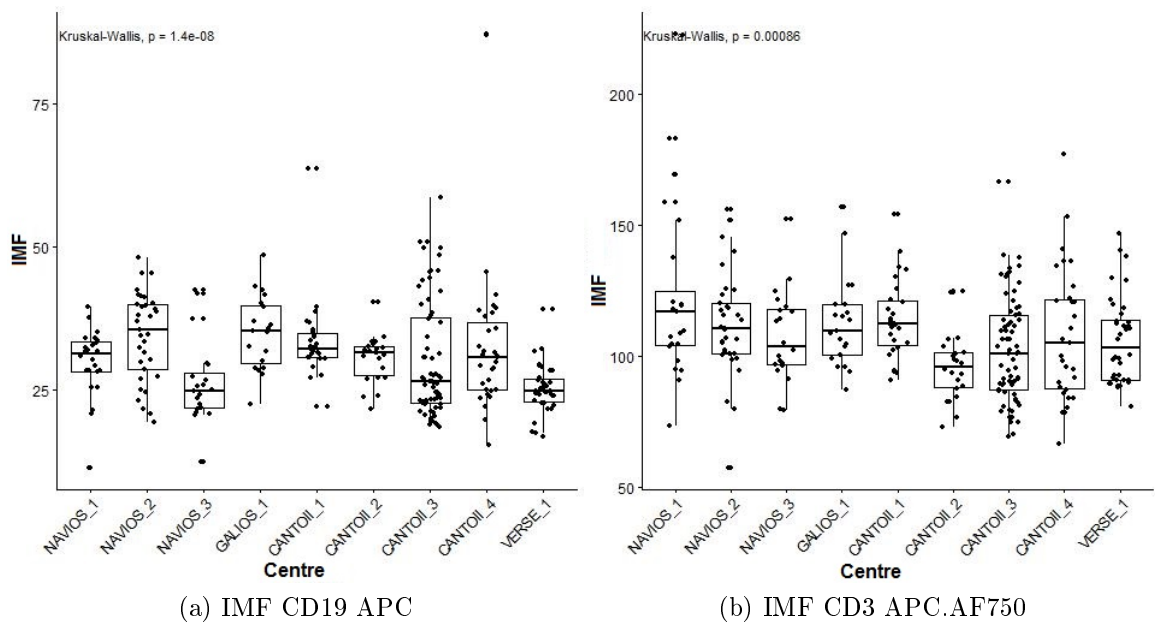


FIGURE 37: **Détail de l'analyse manuelle.** Boxplot sur les IMF du panel 1 des 300 patients de la Phase I pour CD19 APC(a) et CD3 APC.AF750 (b). Test de Kruskal-Wallis.

La comparaison entre les centres pour l'IMF est dépendante des canaux et des centres. Cette situation peut être illustrée par les résultats sur le CD19 APC et ceux sur le CD3 APC-AF750 qui valident la présence d'une différence significative entre les centres en particulier pour le canal APC ($p < 0.05$) (*Figure 37*).

Malgré la comparabilité des résultats concernant les fréquences, la présence d'un effet centre pour l'IMF est un sérieux inconvénient pour l'utilisation des données de cytométrie.

2.5.4 Comprendre l'effet centre

Pour mieux comprendre cet effet centre, une expérience a été réalisée entre les différents centres du projet. Le principe a consisté à réaliser la calibration et l'acquisition du panel 1 en réduisant au maximum les variations techniques par l'utilisation des mêmes lots de réactifs dans les 11 centres.

Les réactifs concernés sont les billes VersaComp (BC), le kit de compensation (BC), les billes 8 pics (BC) et des immunotrol (BC). L'immunotrol est du sang dit standardisé qui est utilisé pour reproduire les populations du système immunitaire (lymphocytes, granulocytes et monocytes) en taille/structure et dans l'expression de certains marqueurs de surface.

En plus de ces réactifs standardisés, du sang frais a été prélevé sur un témoin à Brest (UBO) et envoyé aux différents centres, avant d'être analysé dans un délai de

24 à 48h.

Le début de l'expérience consiste en la réalisation d'une calibration classique des 11 cytomètres avec les kits de compensation, les Versacomp et les billes 8 pics.

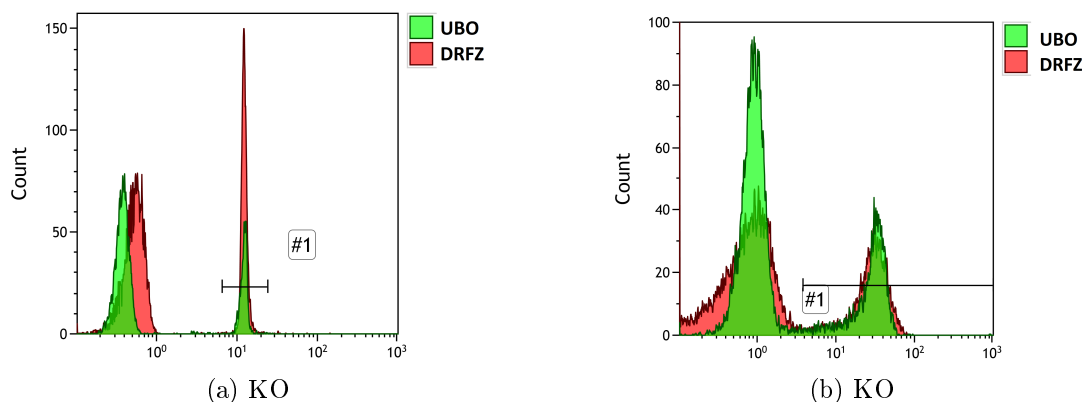
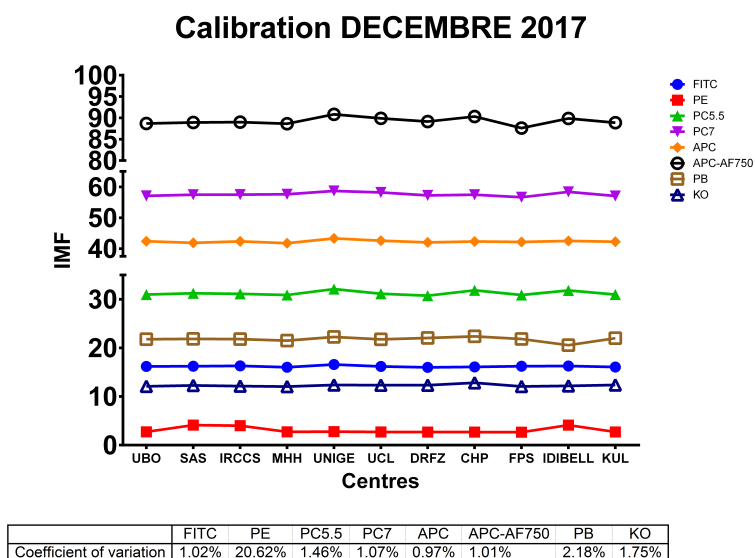


FIGURE 38: **Exemple de la calibration.** Exemple d'analyse mono-paramétriques sur le canal KO entre les centre UBO et DRFZ pour la calibration(a) et pour le sang(b).

FIGURE 39: **Calibration inter-centres.** Graphique montrant les IMF de la calibration lors de l'expérience incluant les mêmes lots de réactifs entre les 11 centres de PRECISESADS.



Les résultats de cette calibration indiquent une bonne harmonisation entre les centres avec des variations inférieures à 5% (Figure 38). Sauf pour le canal PE, où 3 centres ont utilisé la mauvaise référence (Figure 39).

Suite à cette harmonisation des centres, le même lot de billes 8 pics a été acquis dans les différents centres avec les voltages déterminés par l'harmonisation.

Ensuite, les échantillons (immunotrol et sang de témoin) sont passés dans les cytomètres des différents centres. L'acquisition des immunotrol est réalisée en suivant

le même SOP que celui d'un échantillon de sang classique.

Les fichiers de cytométrie générés ont été compensés et analysés manuellement par un opérateur unique sur Kaluza (*Figure 38*).

Dans un premier temps, l'analyse s'est concentrée sur la fréquence entre les différents centres de l'étude à l'aide des Immunotrol.

La fréquence des différentes populations cellulaires se révèle très similaire entre les centres avec des variations inférieures à 5% lorsqu'elles sont comparées à l'UBO, validant le fait que les fréquences sont comparables entre les différents centres (*Figure 40*).

Après la validation des fréquences, nous avons analysé les IMF des différentes populations. Pour cela, les résultats obtenus sur le sang ont été exploités, car les immunotrol n'ont pas été validés par BC pour la stabilité de l'expression entre les lots (contrairement aux fréquences). Sur le sang frais, les fréquences et les pourcentages ont été analysés.

Les résultats obtenus sur le sang frais sont plus variables entre les différents centres (*Figure 41*). Les plus fortes variations dans certaines fréquences de populations, comme les CD14+, peuvent être imputées au délais de transport. Malgré cela, les fréquences restent assez stables entre les différents centres, dans le seuil d'acceptabilité de 5%.

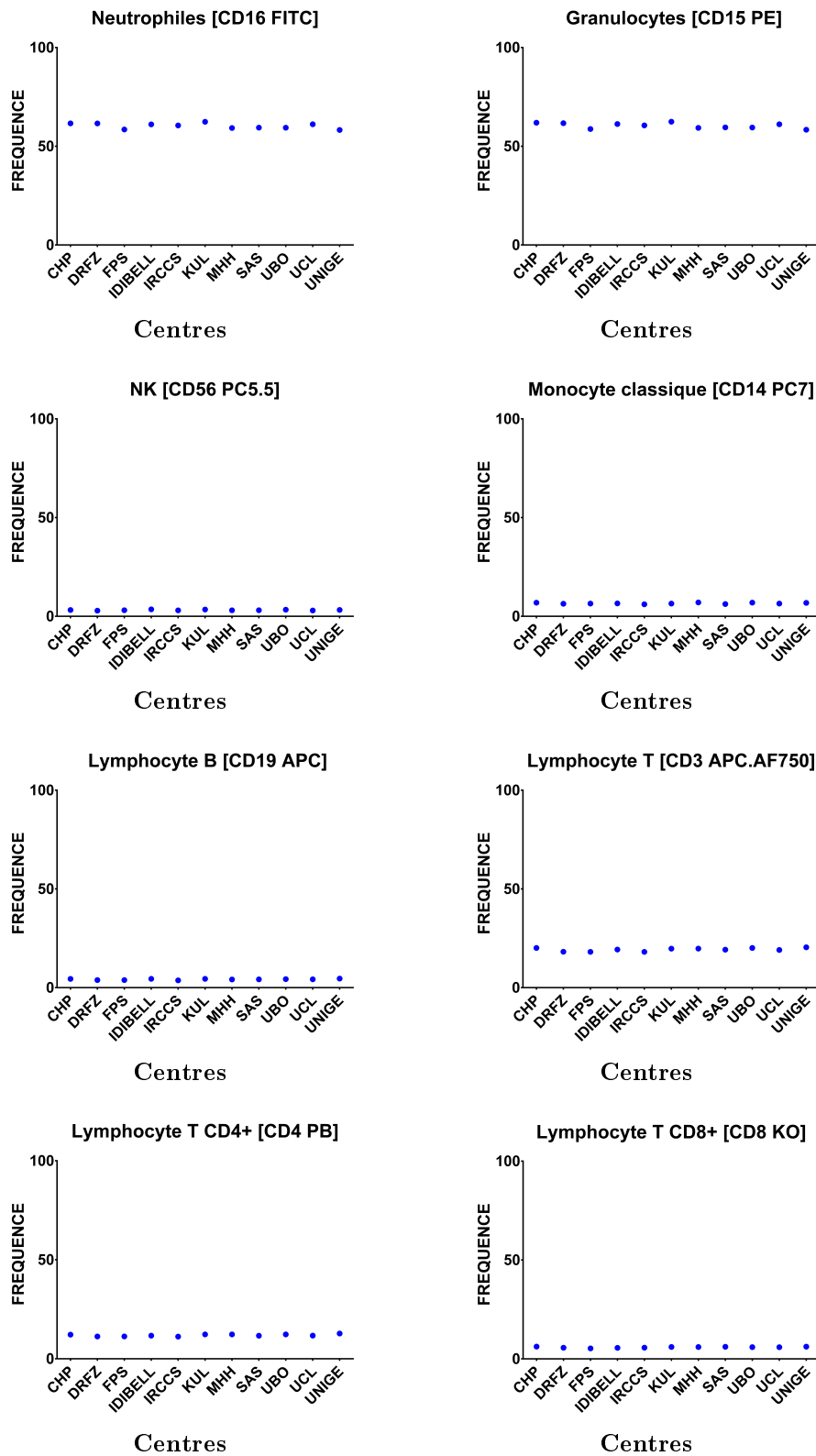
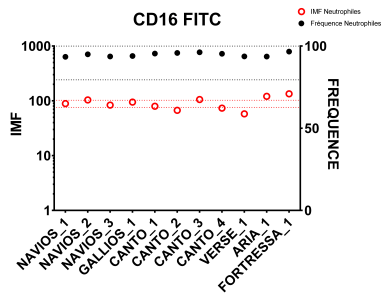
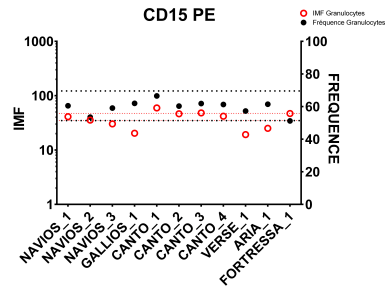


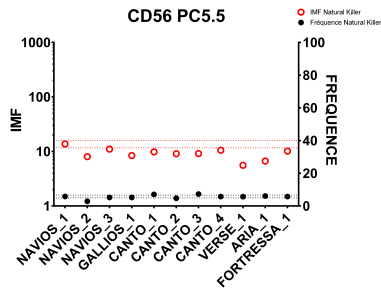
FIGURE 40: **Fréquences immunotrol.** Fréquences des différentes populations issues des Immunotrol lors de l'expérience incluant les même lots entre les 11 centres de PRECISESADS.



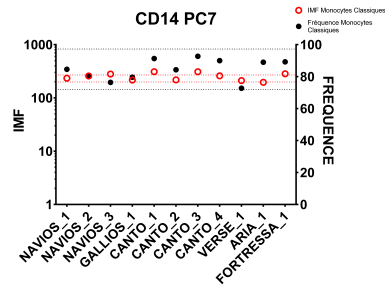
Centres



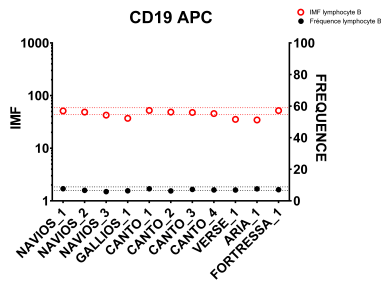
Centres



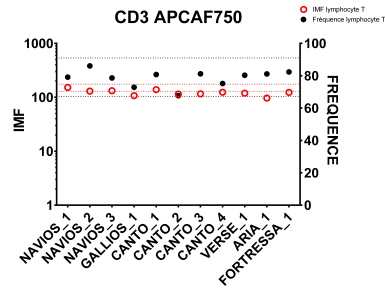
Centres



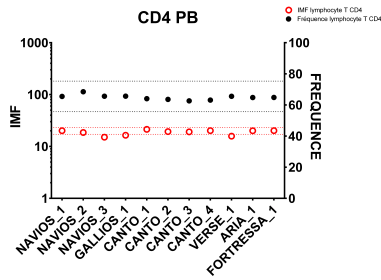
Centres



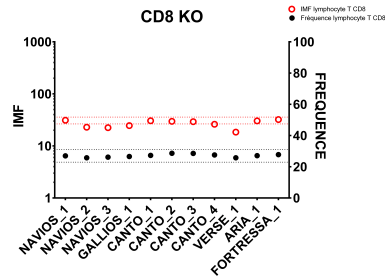
Centres



Centres



Centres



Centres

FIGURE 41: **Fréquences et IMF pour le sang frais.** Fréquences et IMF de l'échantillon de sang envoyé dans tous les centres lors de l'expérience incluant les mêmes lots de réactifs entre les 11 centres de PRECISESADS pour les 8 canaux de cytométrie. Les lignes en pointillé indiquent les limites des $\pm 5\%$.

Concernant les IMF, elles sortent régulièrement de la fourchette des $\pm 5\%$ tolérée et ce même pour des canaux avec des populations stables en fréquence entre les

centres comme les CD4+ ou bien les CD8+.

Une première tentative de correction a été effectuée à l'aide d'une normalisation inter-centre. Cette dernière est très similaire à la normalisation intra-centre en se basant aussi sur les billes 8 pics car elles proviennent du même lot. La différence vient de la référence utilisée pour cette normalisation, elle correspond aux billes du centre UBO. Une difficulté de l'utilisation de la normalisation intra-centre pour la normalisation inter-centres est que le décalage entre les deux jeux de billes n'est pas toujours linéaire.

Pour contrer cet effet, un test a été réalisé en augmentant le degré de l'équation de régression, afin de modifier différemment les fortes intensités et les faibles intensités :

$$Y = c + dX + eX^2 + fX^3$$

Cette formule a ensuite été appliquée aux panels de cytométrie avec X correspondant aux données brutes, Y à la nouvelle valeur normalisée, c, d, e et f aux paramètres de normalisation inter-centres.

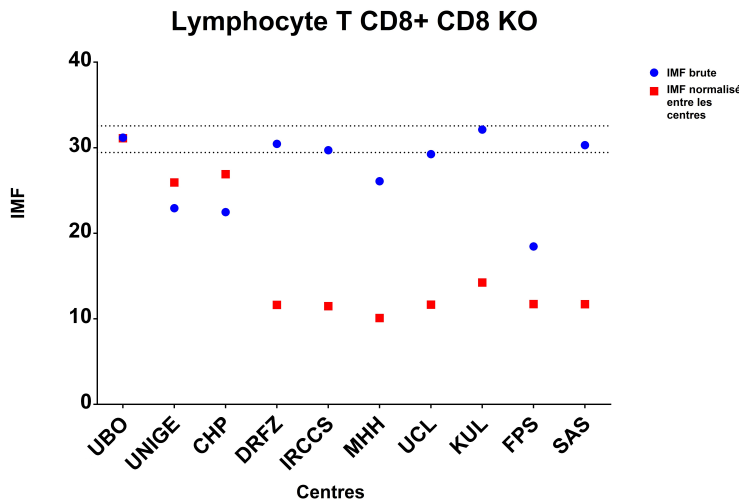


FIGURE 42: **Normalisation inter-centres des immunotrol.** IMF de l'échantillon de sang lors de l'expérience incluant les mêmes lots entre les 11 centres de PRECISE-SADS.

Malgré l'application de la correction, l'IMF de certains canaux sort des 5% acceptable (*Figure 42*). Plus particulièrement lorsque l'on regarde le canal KO, la normalisation a créé une démarcation nette entre les cytomètres BC (UBO, UNIGE, CHP) et BD.

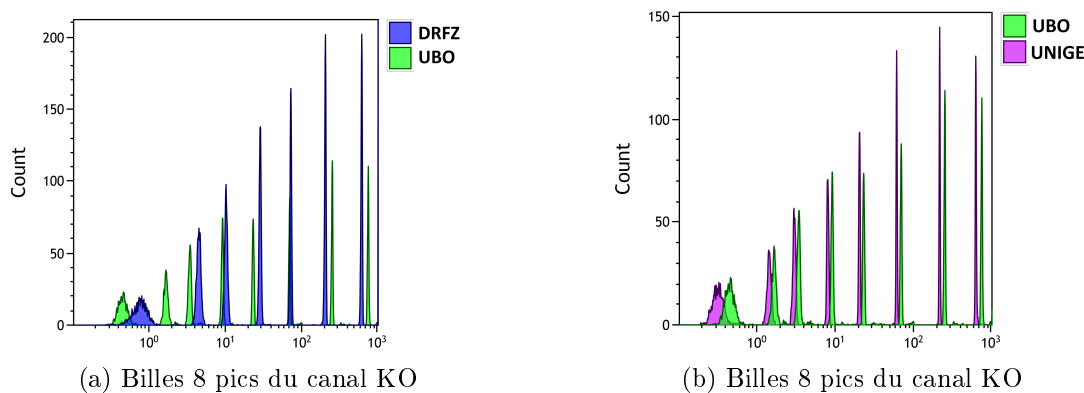


FIGURE 43: **Exemple des différences dans les billes 8 pics.** Différence selon les constructeurs sur le canal KO entre les centres UBO (appareil Beckman Coulter) et DRFZ (appareil BD)(a) et UNIGE (appareil Beckman Coulter)(b).

Ainsi, par exemple, comparés à l'UBO (BC) les différents pics des billes 8 pics, pour un centre BD (DRFZ) et un centre BC (UNIGE), il existe un décalage plus important chez le BD alors que le même lot de billes 8 pics est utilisé et que la calibration avec les billes VersaComp est correcte (*Figure 43*).

Ce décalage observé entre les centres, malgré la bonne calibration, et surtout entre les machines de constructeurs différents peut être expliqué par le fait que la fluorescence des billes 8 pics provient de différents fluorochromes que ceux utilisés dans les panels. Cette différence dans les fluorochromes impacte leur comportement face au laser et leur amplification. De ce fait, cette normalisation n'est pas efficace pour tous les centres et l'utilisation de billes 8 pics ne permet pas de corriger les différences inter-centres.

Afin de corriger l'effet centre, nous avons élaboré une nouvelle solution. Elle consiste à calculer un coefficient correcteur qui se base sur la médiane des IMF de l'ensemble des patients d'un centre. Le but est de ramener cette médiane à celle du centre de référence en partant du principe déjà évoqué que les populations des patients et des contrôles sont similaires entre les centres. Pour réaliser cette correction, un nombre important de fichiers est nécessaire pour avoir plus de poids dans le coefficient calculé, c'est pourquoi l'automatisation de l'analyse des fichiers de cytométrie s'avère plus que nécessaire.

2.6 Automatisation de l'analyse par Altrabio

L'automatisation de l'analyse de cytométrie en flux est réalisée par Altrabio (Lyon), une société de R&D experte en analyse et interprétation de données biomédicales. L'automate proposé par Altrabio utilise une méthode de machine-learning pour répliquer l'analyse manuelle faite sous Kaluza. Pour cela, l'automate utilise comme jeu d'apprentissage, pour chaque panel, 10 individus dans chacun des 11 centres.

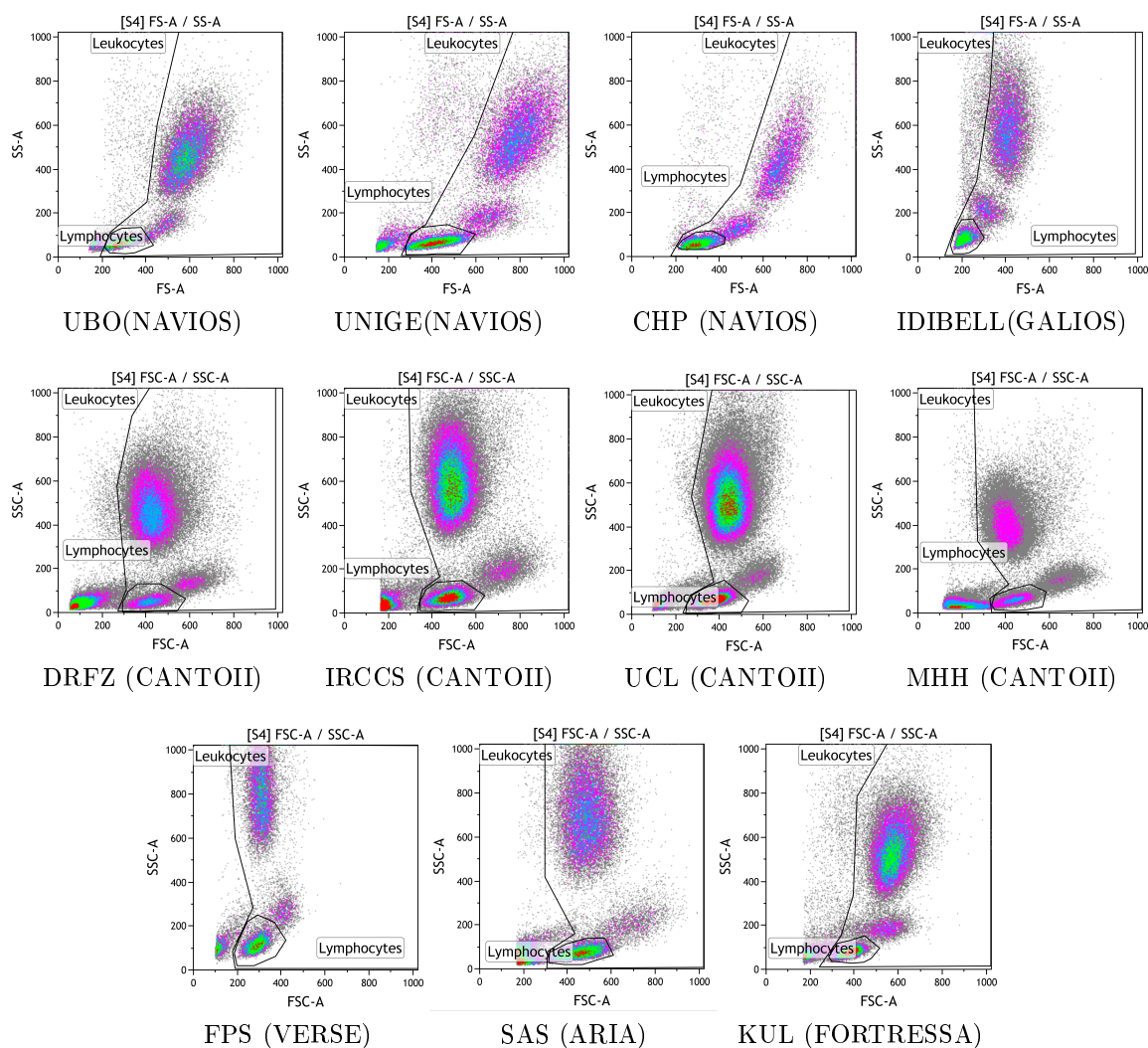


FIGURE 44: **Différences morphologiques entre les centres.** Comparaison de la taille/structure du panel 1 dans les 11 centres du projet PRECISESADS avec les représentations bi-parametriques représentatives générées par chacun des 11 instruments.

Le premier défi que doit relever l'automate est de prendre en compte la variation dans la taille/structure des différents centres. Des variations sont visibles entre les cytomètres des différents constructeurs (par exemple BC pour UBO vs BD pour FPS) mais aussi entre les cytomètres de même type (par exemple BC pour UBO vs UNIGE) (*Figure 44*).

C'est en prenant en compte ces variations que l'automate a été généré sur 10 individus de la phase I et pour chaque centre, puis validé sur un jeu de test correspondant aux individus restants de la phase I, soit 190 individus.

2.6.1 Validation des automates

Les résultats générés par l'automate d'Altrabio sur le jeu de test ont été comparés aux résultats de l'analyse manuelle. Cette comparaison est divisée en 3 catégories d'analyses : Fréquences, Concentration cellulaire et IMF. Les coefficients de corrélation sont générés sur les 3 comparaisons afin d'évaluer la performance globale de l'automate d'Altrabio.

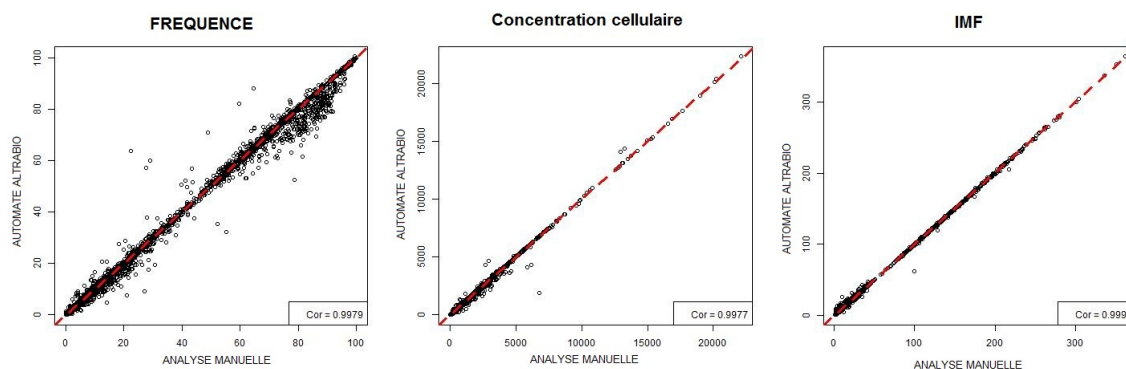


FIGURE 45: **Validation des automates Altrabio Panel 1.** Comparaison de l'analyse manuelle et de l'analyse automatisée par Altrabio sur les données du Panel 1 en fréquence (a), concentration cellulaire (b) et IMF(c)

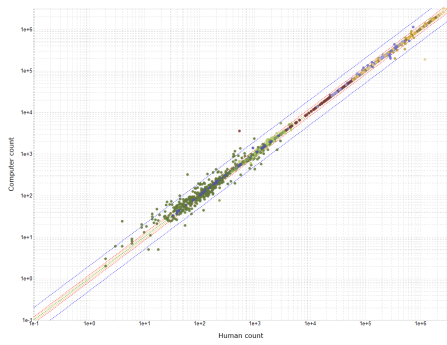
La première comparaison est faite sur les résultats du panel 1 car ce dernier implique tous les patients de l'étude. De bonnes corrélations ont été retrouvées pour les 3 types de données ($r > 0.99$) (Figure 45).

Malgré la bonne corrélation entre l'analyse manuelle et l'analyse automatisée, certaines valeurs (points) divergent fortement. Pour mieux comprendre l'origine de ces divergences, le gating de chaque point a été comparé visuellement entre l'analyse automatisée et manuelle.

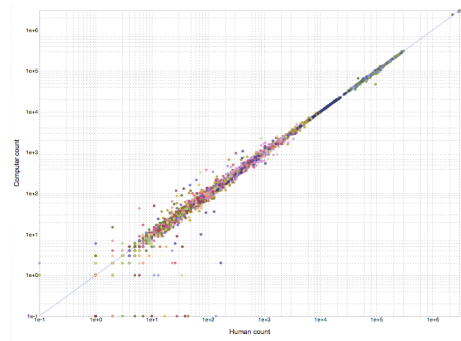
Les premiers points correspondent majoritairement à des gates servant à l'exclusion des débris et des cellules mortes. L'inclusion ou non de plus d'évènements sur ces gates est souvent contrebalancé par la dernière gate d'exclusion sur la taille/structure et n'a que peu d'impact sur la suite de l'analyse.

Pour les derniers points de divergence, ils correspondent à des gates où l'analyse manuelle a été moins constante entre les patients, faisant de l'analyse automatisée, une meilleure analyse que celle faite manuellement car plus stable entre les individus.

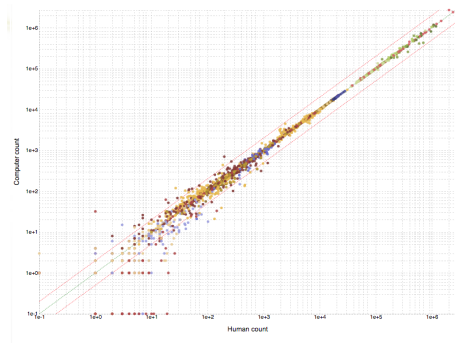
Suite à ces premiers résultats, la comparaison pour les panels restant s'est faite sur le nombre d'évènements (qui sont à l'origine des fréquences et des concentrations cellulaires).



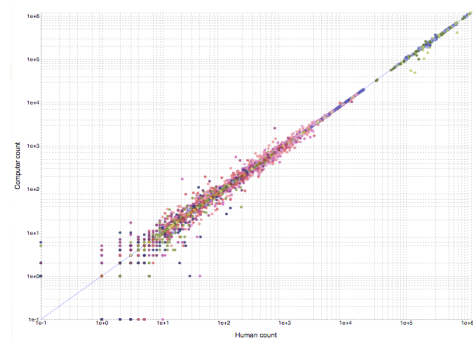
Panel 2



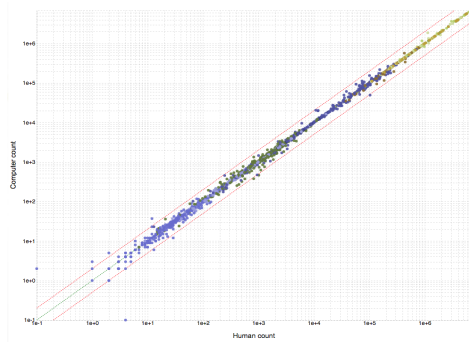
Panel 3



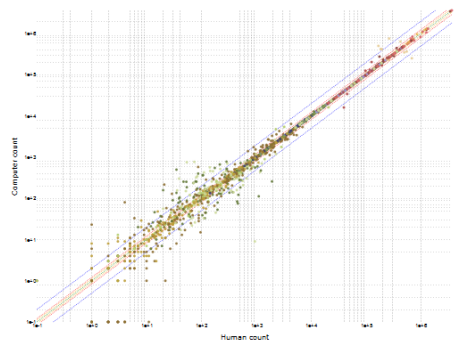
Panel 4



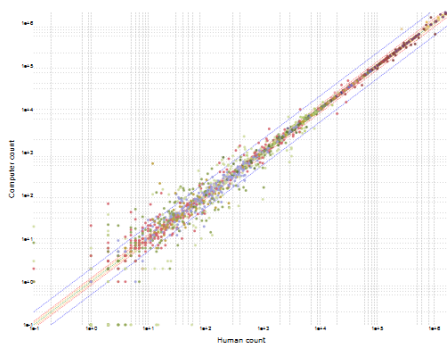
Panel 5



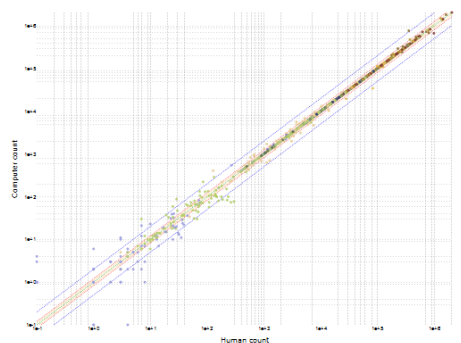
Panel 6



Panel 7



Panel 8



Panel 9

FIGURE 46: **Validation automates Altrabio.** Comparaison de l'analyse manuelle et de l'analyse automatisée par Altrabio sur les nombre d'évènements de chacune des gate des Panels 2 à 9.(a - c) Source : Altrabio (Lyon)

Les résultats des 9 automates sur l'ensemble des 9 panels montrent une très bonne corrélation avec ceux de l'analyse manuelle ($R > 0.99$) (*Figure 46*). Ces bonnes corrélations entre l'analyse manuelle et automatisée nous permettent de valider l'utilisation de l'automate.

Suite à la validation de cette analyse, les fichiers de cytométrie de la phase I, II et INCEPTION ont été envoyés à Altrabio pour être traités par les différents automates.

2.6.2 Fichiers aberrants

Malgré le nombre important de fichiers, seuls 1051 (8% des 12722 fichiers totaux attendu) n'ont pas pu être inclus dans l'analyse. Les raisons de leur absence sont diverses.

La première raison (environ 16,3% des 1051) provient de résultats manquants à cause de l'absence de fichiers de cytométrie car ils n'ont pas été envoyés et ont été perdus ou bien dû au fait que le cytomètre n'était pas opérationnel au moment de l'arrivée du prélèvement. Lorsque cela concerne les fichiers des billes 8 pics, il est impossible de normaliser les données associés à ces billes.

Une autre raison identifiée (environ 4,0%) est l'arrivée de problèmes techniques comme la corruption du fichier. Cette corruption peut prendre plusieurs aspects comme un fichier vide ou un nombre d'évènement différent entre l'entête et les données. Ces fichiers sont inutilisables pour l'analyse.

Un autre problème (1,4%) est la taille de certains fichiers qui est supérieure à 2Go. Cette taille est due à un centre qui n'a pas paramétré les bons filtres lors de l'acquisition dans le cytomètre. En effet, il est possible de mettre en place un filtre (seuil de détection) pour ne pas enregistrer les éléments inférieurs à une certaine taille ou structure. A cause de ce filtre trop faible, des débris supplémentaires ont été enregistrés et ont augmenté la taille du fichier de cytométrie. Cette taille excessive ne permet pas d'être ouvert par des logiciels comme Kaluza.

Un problème récurrent (4,8%) est l'apparition de différences dans les noms des paramètres, de leur ordre dans les fichiers ou de leurs absence dans certains fichiers. Cette différence a un impact sur le bon déroulement des différents scripts de normalisation, de compensation, et d'analyse par les automates d'Altrabio.

Après l'analyse des fichiers de cytométrie par Altrabio, une liste d'outliers a été générée en plus des résultats comme les fréquences ou les IMF.

A partir de cette liste, il est apparu que certains fichiers possédaient des profils anormaux. L'expression anormale de certains marqueurs est attendue dans l'observation des maladies (augmentation du nombre de certains types cellulaires) ou résul-

tant de traitements (déplétion des lymphocytes B [48]). Mais pour certains fichiers, la piste d'un problème technique a été préférée comme la présence de populations cellulaires positives dans l'expression de 2 marqueurs exclusifs (CD3 avec CD19 par exemple).

Ces fichiers problématiques, souvent détectés à cause de valeurs aberrantes, peuvent avoir des origines techniques ou biologiques. Ces fichiers, une fois identifiés, ont été évalués manuellement. Les résultats de cette analyse permettent de déterminer si le fichier est conservé, analysé manuellement par l'intermédiaire de Kaluza ou bien retiré définitivement de l'analyse.

Environ 10,4% de ces fichiers possèdent des profils hors-normes. En effet, certaines populations que l'on retrouve usuellement dans de nombreux individus sont absentes pour certaines inclusions. Par exemple la population des Neutrophiles a été retrouvée sans l'expression (ou très faible) du CD16, qui la caractérise pourtant.

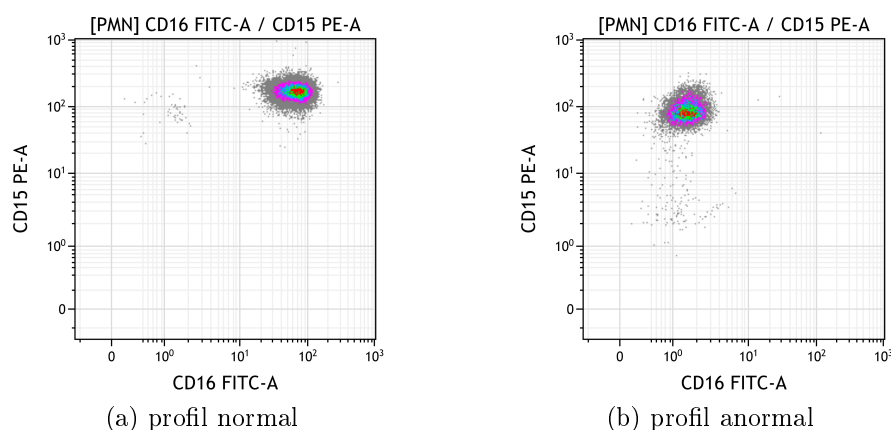


FIGURE 47: **Exemple de profil anormal des Neutrophiles.** Exemple d'un profil normal (a) et d'un profil anormal (b) pour l'expression du CD16 sur les Neutrophiles.

En effet, on observe que l'absence de granulocytes CD16+ est combinée à une augmentation du nombre d'évènements dans la population éosinophiles. Dans ce cas là, les neutrophiles semblent avoir perdu le CD16 ou bien expriment un CD16 différent ne pouvant pas être pleinement identifié par l'anticorps (*Figure 47*). Les profils d'expression de certains fichiers étaient hors normes sans pour autant avoir pu déterminer de cause technique. Des cas comme la présence d'une deuxième population positive pour l'expression du CD3.

Lorsque certains outliers ont été évalués (environ 2,2%), il est apparu que le panel avait un très grand nombre d'évènements qualifié de débris qui masquaient les lymphocytes et empêchaient l'exploitation fiable des résultats de ce panel. L'origine de cette augmentation des débris peut être due à un problème durant la préparation de l'échantillon.

Il est survenu, pour certaines inclusions (3,7%), que des panels aient été inversés. Par exemple, un panel 7 a été acquis avec le protocole d'acquisition du panel 9 et vice versa. Cette inversion a été identifiée par la présence d'un signal positif dans un canal sensé être vide et par l'aspect de certains profils bi-paramétriques qui ont permis de retrouver le panel original (*Figure 48*).

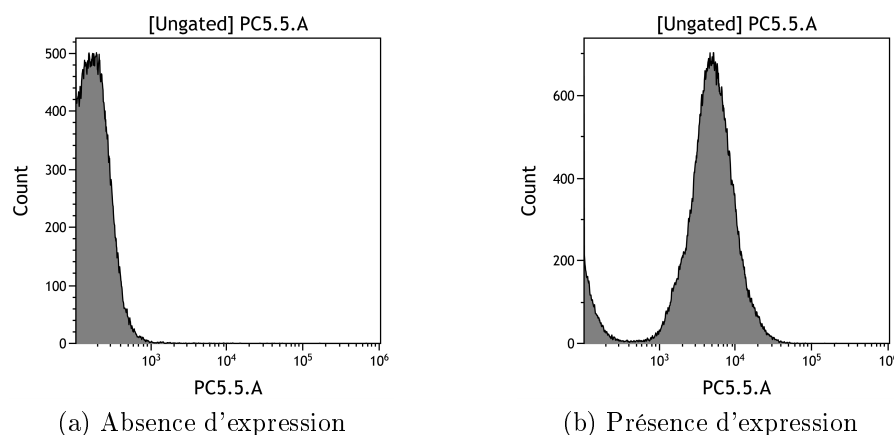


FIGURE 48: **Exemple de profil d'inversion de panel.** Profil d'expression du panel 7 (a) et du panel 9 (b) du même échantillon de sang. L'inversion entre ces deux panels est mise en évidence par l'absence d'expression dans le canal PC5.5 pour le panel 7 couplé à la présence d'expression dans le panel 9. Sachant que le canal PC5.5 du panel 9 est vide (*Figure 20*).

Malgré la puissance de l'automatisation, des erreurs dans le gating peuvent survenir. 8,3% des fichiers outliers ont été identifiés comme ayant des erreurs dans le gating automatisé. La solution choisie est d'analyser manuellement ces fichiers.

Des problèmes sont également apparus sur un nombre de billes de comptage (1,7%). Ces problèmes peuvent être divisés en deux catégories. La première est celle d'erreurs de gatings, corrigible par une analyse manuelle. La deuxième catégorie est l'absence des billes de comptage. En effet, dans certains cas, les billes de comptage n'ont pas été ajoutées au panel 1 et 2. Cette absence a un impact sur les données de concentrations cellulaires qui n'ont pas pu être rendues.

Enfin, le problème le plus courant (47,2%) est celui du découplage de certains fluorochromes en tandem qui impacte l'IMF mais aussi la fréquence des populations étudiées et par conséquent, rend les fichiers inexploitable.

2.6.3 Découplage de fluorochromes

Certains panels de cytométrie, comme le panel 9, ont dans leur cocktail de fluorochromes le CD19 APC et le CD3 APC-AF750. Ces marqueurs permettant de définir les populations lymphocytaires B et T respectivement. De ce fait, l'expression de ces

marqueurs est dite exclusive. Mais, pour certains patients, il est apparu une population positive en CD3 et en CD19. Dans le même temps la population CD3+ CD19-habituelle n'était pas retrouvée (*Figure 49*).

Pour finir, il y avait une discordance entre les différents panels car cette population CD3+ CD19+ n'était pas retrouvée dans le panel 1 alors que ce dernier possède les mêmes marqueurs CD19 APC et CD3 APC AF750.

Cet effet a été observé sur plusieurs inclusions, il va d'un faible signal positif en CD19 à l'absence total de signal CD3+. Dans ces cas là, la séparation des sous-populations lymphocytaires est impossible.

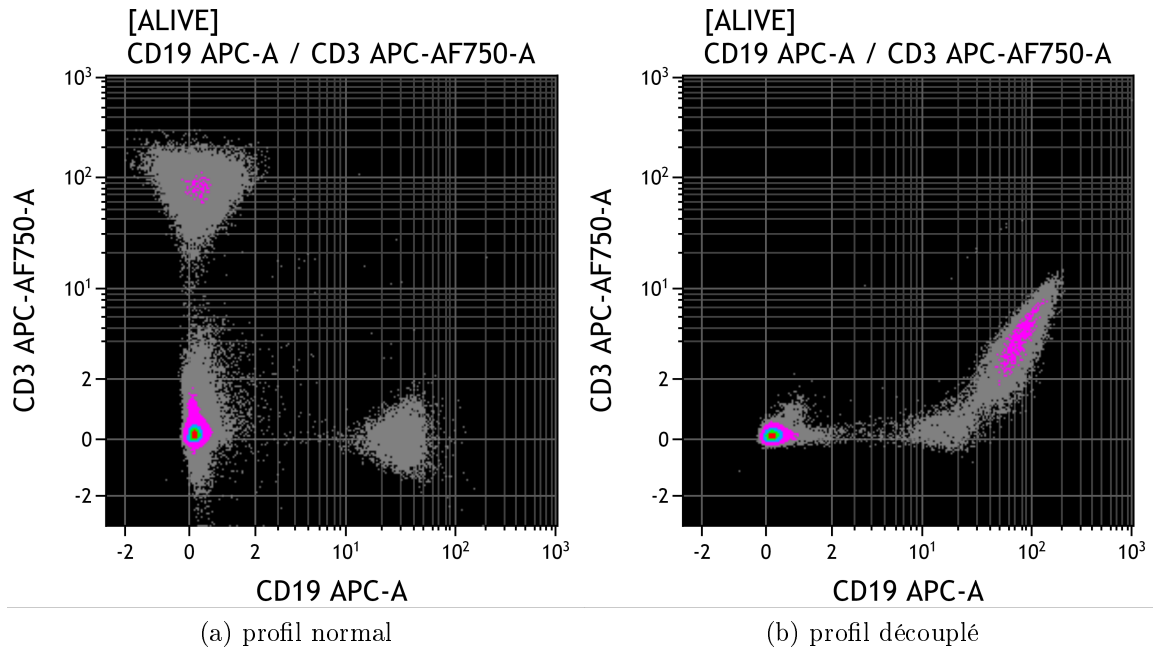


FIGURE 49: **Exemple de découplage de l'APC.** Exemple d'un profil normal (a) et d'un profil découplé (b) de l'APC.AF750 vers l'APC.

Ce problème a été observé sur d'autres canaux possédant des fluorochromes en tandem, c'est à dire le PC7 et le PC5.5. Ces derniers pouvant se découpler et renvoyer un signal positif en PE. Plus particulièrement sur les Monocytes CD14+ PC7 où, similairement au CD3 APC AF750, la population est négative pour le marqueur CD14. Cette population étant positive dans les panels précédents (*Figure 50*).

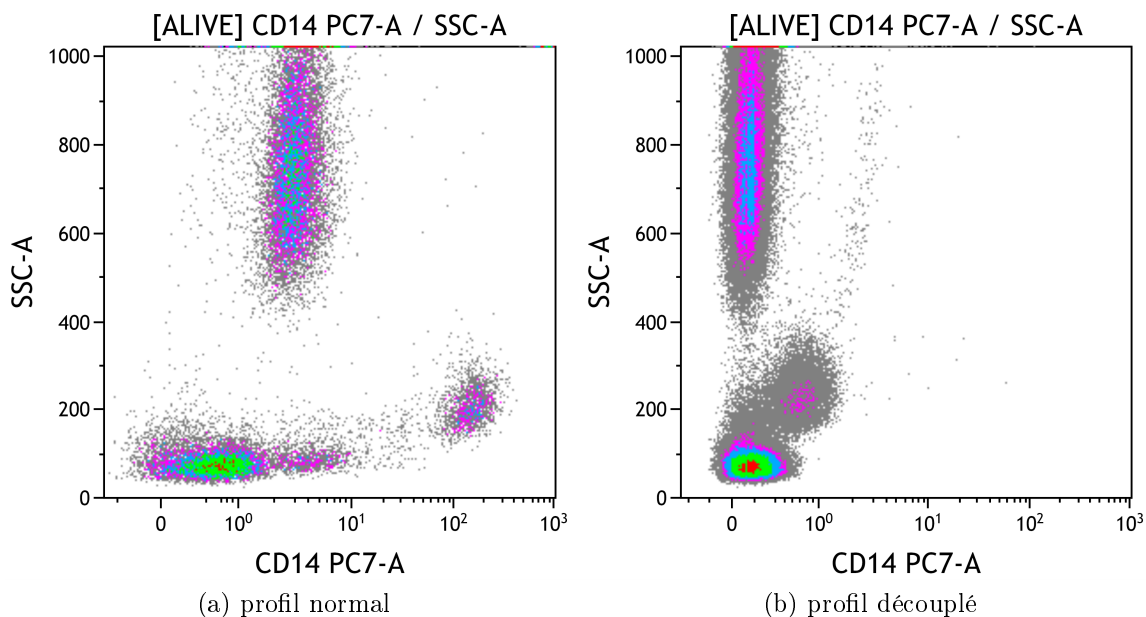


FIGURE 50: **Exemple de découplage PC7.** Comparaison des panels 1 (a) et 9 (b) du CD14-PC7 du même patient.

Ce phénomène illustre parfaitement les difficultés rencontrées lors du découplage de fluorochromes en tandem. Plus particulièrement pour les couples de fluorochromes comme APC-AF750, PE-Cy7 (PC7) ou PE-Cy5.5(PC5.5) qui peuvent se découpler. Ce qui pose problème car on se retrouve avec une diminution, voir une absence, du signal en AF750 et une augmentation du signal en APC.

Un découplage de fluorochromes peut être dû à une exposition trop longue à la lumière, ou à des températures trop forte. La cause de ce découplage n'a pas encore été bien établie, mais il semblerait que cela concerne les conditions de stockage ou de manipulation des tubes duracloes. En effet, bien qu'un découplage ait été observé dans plusieurs des centres de l'étude, la majorité des cas sont retrouvés dans un même centre.

Dans tous les cas, les fichiers où le découplage a été observé ont été retirés de l'analyse car impactant les résultats (IMF, fréquences, nombres absolus).

2.6.4 Observation de l'effet lot et de l'effet centre

Les résultats de l'automatisation ont permis d'observer l'évolution, au cours du temps, de l'IMF des populations et de la comparer à l'évolution des IMF des billes 8 pics sur un grand nombre de patients.

Nous avons analysé les pics "low" et "high" des billes 8 pics pour le canal PE. L'évolution de ces pics se fait par des changements brutaux. Ces changements correspondent le plus souvent à des changements dans les numéros des lots de billes 8 pics, nécessitant l'établissement de nouvelles références (*Figure 51*).

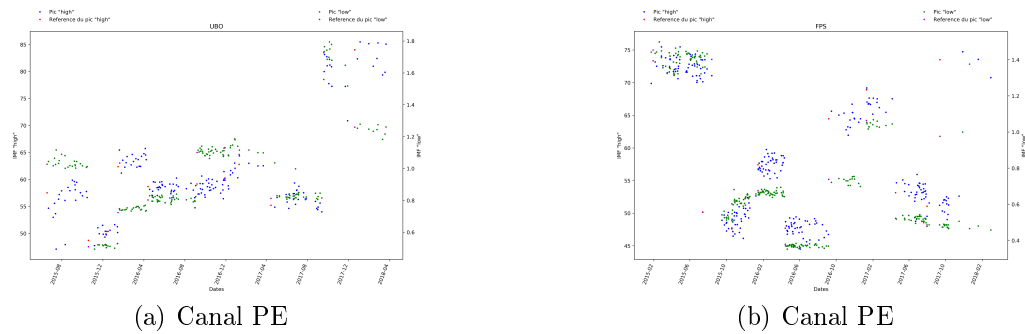


FIGURE 51: **Évolution des IMF des billes 8 pics.** Évolution des pics faibles (en vert) avec leurs références (vert clair) et forts (en bleu) avec leurs références (en rouge) pour la PE au cours de l'étude pour les centres UBO (a) et FPS (b).

En regardant l'évolution au cours du temps de l'IMF de certaines populations (représenté ici par l'expression du CD15 dans le canal PE sur les granulocytes) (*Figure 52*), on pourrait s'attendre à une certaine stabilité au vu du nombre d'individus inclus. Mais des décalages nets sont apparus à certaines dates qui sont partagées entre les centres. Après recoupement, ces dates de décalages dans l'IMF correspondent aussi aux 2 dates de changement de lots dans les kits de compensation duraclohes.

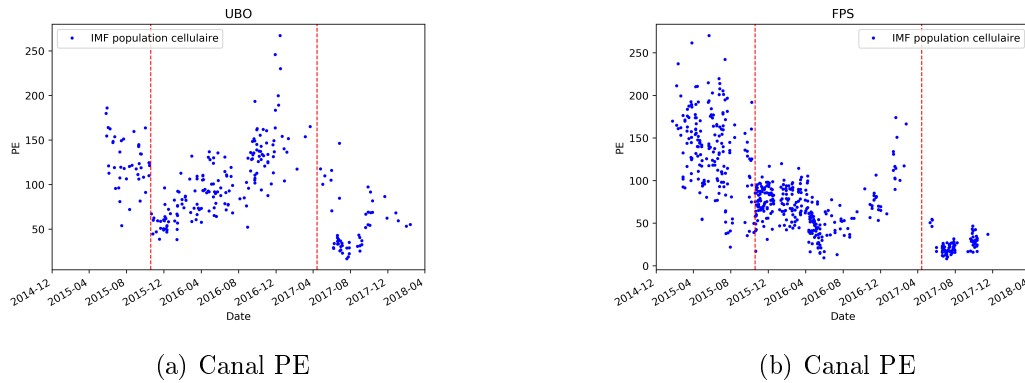


FIGURE 52: **Évolution de l'IMF du panel 1.** Evolution de l'IMF des populations cellulaires CD15 PE positives dans le panel 1 au cours de l'étude pour les centres UBO (a) et FPS (b). En pointillé rouge, sont indiquées les dates de changement de lot de kit de compensation duraclone.

En effet, 3 différents lots de duraclones ont été utilisés au cours de l'étude. L'introduction d'un nouveau lot a été précédée d'une phase de test où les résultats du nouveau lot sont comparés au précédent. Lorsque la variation entre les lots a été jugée acceptable, la production de ce nouveau lot de duraclones a été validée et les différents centres ont commencé à l'utiliser.

Mais cette utilisation d'un nouveau lot ne s'est pas faite au même moment dans tous les centres. En effet, pour certains centres, l'utilisation d'un nouveau lot de tube duraclone n'est pas corrélée avec une calibration, alors qu'ils possèdent aussi des décalages de l'IMF au date de calibration.

C'est ce qui nous dirige vers le changement de lot du kit de compensation lors de la calibration comme responsable de cet effet lot. En effet, comme le nouveau lot était considéré comme similaire à l'ancien, les cibles n'ont pas été changé pour le nouveau lot.

2.7 Effet lot

Ce nouvel effet a besoin d'être corrigé pour réduire la variation technique introduite. Nous sommes parti du principe que la médiane des différentes IMF est constante tout au long de l'étude. L'idée est alors de "normaliser" les lots en calculant un coefficient qui prendra pour référence le premier lot.

2.7.1 Calcul du coefficient lot

Le coefficient a été calculé pour chacun des lots, pour chaque canal de chaque centre, soit 176 coefficients pour chaque panel. Seulement les exemples de l'UBO et

FPS sur le canal PE du panel 1 sont représentés car ce sont ces derniers qui présentent le mieux la variation entre les lots pour le panel 1.

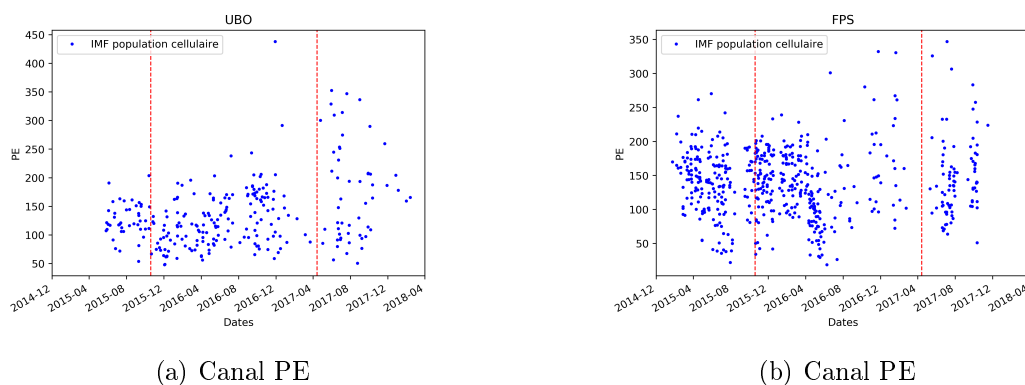


FIGURE 53: **Évolution de l'IMF du panel 1 corrigé.** Évolution de l'IMF corrigée des populations cellulaires CD15 PE positives dans le panel 1 au cours de l'étude pour les centres UBO (a) et FPS (b). En pointillé rouge, sont indiquées les dates de changement de lot de kit de compensation duraclone.

L'application du coefficient lot permet de réduire l'impact du changement de lot en obtenant une IMF stable entre les différents lots. Le coefficient a été calculé et appliqué sur l'IMF des différents panels dans les différents centres (*Figure 53*).

Le calcul et l'application des coefficients sont réalisés à l'aide d'un script écrit en Python (*Annexe A.7.2*, page 261).

2.7.2 Observation de l'effet centre après correction de l'effet lot

Après la correction de l'effet lot, de nouvelles ACP sont réalisées sur les résultats corrigés pour les panel 1 et 2 afin de voir si cela permet de corriger l'effet centre.

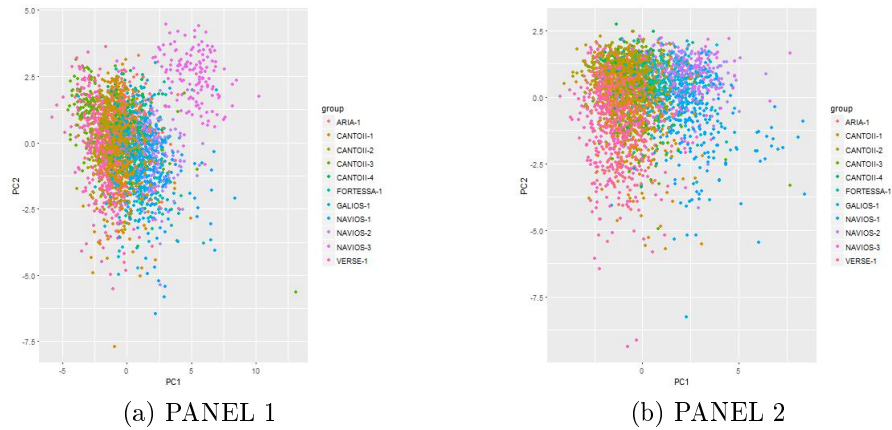


FIGURE 54: **ACP sur l'IMF après correction effet lot.** ACP sur l'IMF après correction de l'effet lot sur le panel 1 (a) et le panel 2 (b).

Les résultats de l'ACP montrent qu'il existe toujours des clusters définis par leur centre d'origine. En particulier pour le panel 1 où un des centres se détache très nettement des autres (*Figure 54*).

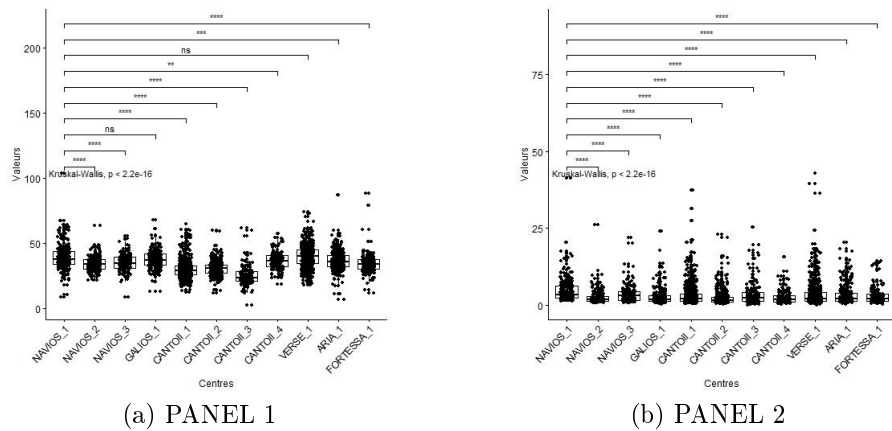


FIGURE 55: **Évolution de l'IMF du panel 1.** Boxplots représentant l'IMF après correction de l'effet lot sur le CD19 APC du panel 1(a) et le CD123 APC du panel 2(b). Test statistique de Kruskal-Wallis.

Le maintien d'un effet centre est validé par le test de Kruskal-Wallis en comparant les IMF des centres pour chacun des 8 fluorochromes. En reprenant l'exemple sur le canal APC, que cela soit pour le panel 1 ou 2, la différence entre les centres est significative ($p < 0.05$). Pour pouvoir exploiter les données des IMF, il est donc nécessaire de trouver une méthode de correction de cet effet centre (*Figure 55*).

2.8 Effet centre

L'objectif principal de cette correction de l'effet centre est de pouvoir rendre les IMF comparables entre les centres dans la mesure où les recrutements des patients

sont homogène entre les centres. Le deuxième objectif est de pouvoir appliquer cette correction directement sur les fichiers de cytométrie, permettant ainsi l'analyse non-supervisée qui utilise les intensités de fluorescence. Pour cela, la même correction doit être applicable que la valeur de l'intensité de fluorescence soit faible ou forte.

2.8.1 Calcul du coefficient de l'effet centre

Ce calcul de correction de l'effet centre est très similaire à la normalisation intra-centre. En effet, le principe est de calculer la médiane des IMF pour chaque population cellulaire et à partir de ces médianes de réaliser une "normalisation" avec comme valeurs de références, celles de l'UBO.

Par exemple, pour le canal FITC du panel 1, l'information de l'IMF est disponible pour 5 populations cellulaires (Neutrophiles, Éosinophiles, Monocytes intermédiaires, Monocytes classiques, Monocytes non-classiques). La médiane des IMF des différentes inclusions du centre est réalisée pour chacune des 5 populations. Ces médianes pour chaque groupe d'IMF sont utilisées pour définir une régression linéaire similairement à la normalisation des billes 8 pics. Cette régression va aligner les différentes médianes d'un centre à celles de l'UBO pour déterminer le coefficient et l'intercepte (*Figure 56*)

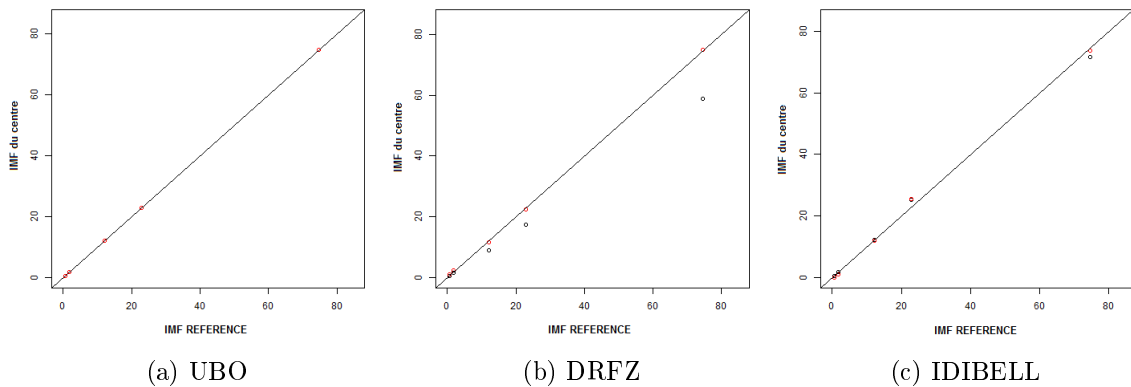


FIGURE 56: **Correction effet centre.** Correction de l'effet centre sur la moyenne de l'IMF de 4 populations cellulaires pour les centres UBO (a), DRFZ (b) et IDIBELL (c).

L'avantage de prendre en compte plusieurs populations cellulaires est que ceci permet de définir les paramètres de la régression linéaire avec plusieurs points de comparaison dont des populations dites négatives. L'autre avantage est la possibilité d'appliquer les paramètres de corrections aux fichiers fcs pour corriger l'IMF des panels et de pouvoir, par la suite, utiliser des méthodes non-supervisées.

Après avoir calculé les paramètres de correction pour chacun des canaux du panel 1 et 2, une ACP a été réalisée sur les IMF qui ont subi cette correction.

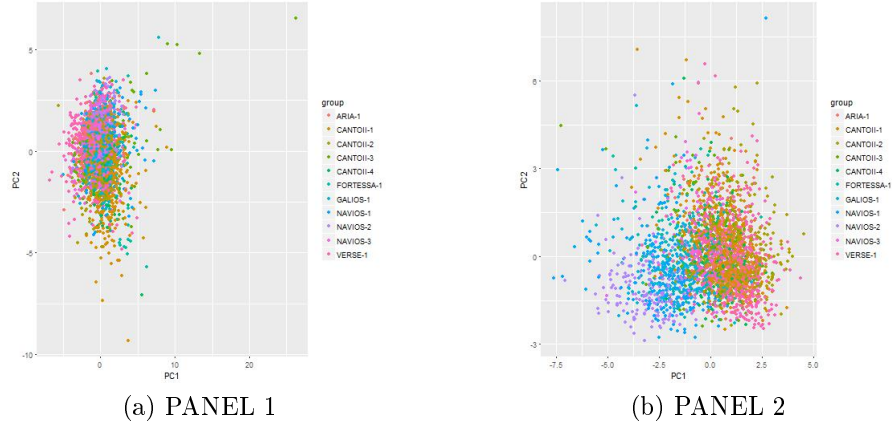


FIGURE 57: **ACP sur l'IMF après correction de l'effet centre.** ACP sur l'IMF après correction de l'effet centre sur le panel 1 (a) et le panel 2 (b).

Les ACP semblent meilleures que précédemment mais l'effet centre reste présent, en particulier pour le panel 2 (*Figure 57*). La validation utilisant le test de Kruskal-Wallis avec un seuil significatif fixé à 0,05 est réalisée. Les résultats indiquent que la correction a été efficace pour certaines populations car ne possédant plus de différences significatives entre les centres. Mais il reste quelques populations significativement différentes (*Figure 58*).

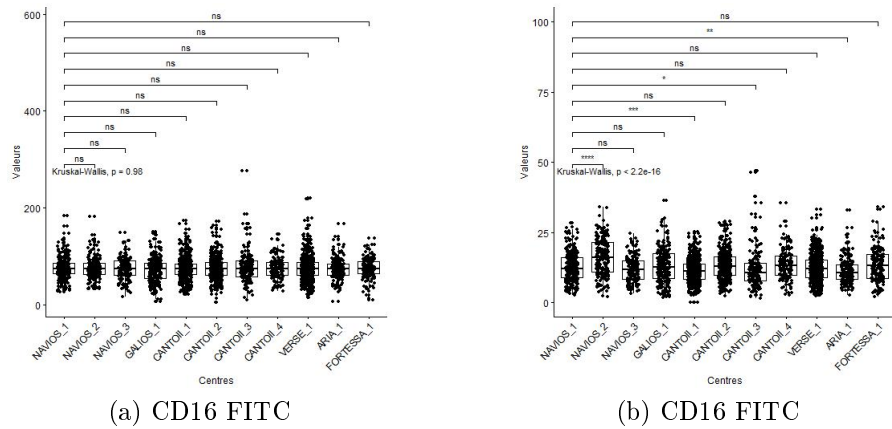
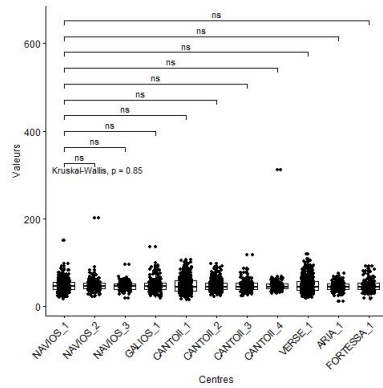
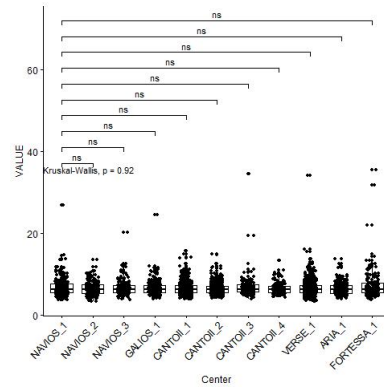


FIGURE 58: **Application correction effet centre PANEL 1.** Boxplot sur l'IMF des Neutrophiles (a) et des Monocytes intermédiaires (b) après application de la correction de l'effet centre pour les 11 centres.



(a) CD11c PC7



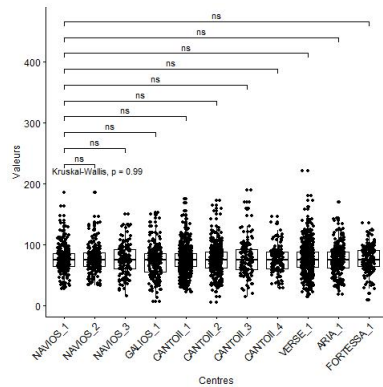
(b) CD56 PC5

FIGURE 59: **Application correction effet centre PANEL 2.** Boxplot sur l'IMF des mDC (a) et des mDC2 (b) après application de la correction de l'effet centre pour les 11 centres.

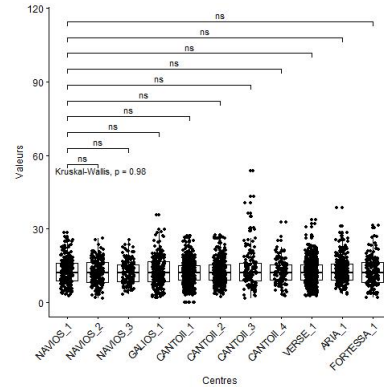
La correction semble incomplète pour certaines populations de certains centres (*Figure 58 et 59*). Face à ces différences inter-centres qui persistent, une deuxième proposition a été faite pour obtenir une correction efficace de l'effet centre.

2.8.2 Calcul du coefficient inter-center version 2

Pour pouvoir exploiter les IMF et les comparer entre les centres, une deuxième méthode de correction de l'effet centre a été réalisée. Le but est que chaque population ait un coefficient correcteur. Pour cela, un ratio des médianes est fait entre celle du centre analysé et celle du centre de référence (UBO). Dans un deuxième temps, le coefficient centre a été appliqué aux IMF de tous les centres pour valider la correction.



(a) CD16 FITC



(b) CD16 FITC

FIGURE 60: **Application correction effet centre PANEL 1.** Boxplot sur l'IMF du CD16 FITC des Neutrophiles (a) et des Monocytes intermédiaires (b) après application de la correction de l'effet centre pour les 11 centres.

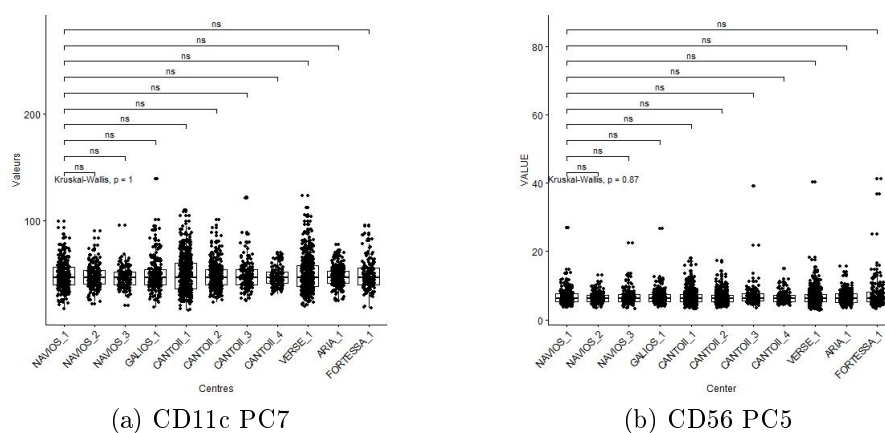


FIGURE 61: **Application correction effet centre PANEL 2.** Boxplot sur l'IMF des mDC (a) et des mDC2 (b) après application de la correction de l'effet centre pour les 11 centres.

Une fois appliqué, ce coefficient permet de réduire les différences entre les centres (*Figure 60 et 61*). Mais, pour un même canal, les coefficients ne sont pas identiques entre les populations. Par exemple, le coefficient pour les intensité de fluorescence du canal FITC des neutrophiles est de 1,26651403375. Alors que pour le même canal il est de 1,35538123243 pour les monocytes intermédiaires. C'est pour cette raison que, finalement, ces coefficients ne peuvent pas être utilisés directement sur les intensité de fluorescence du fichier de cytométrie.

Pour finir, une comparaison des différentes ACP a été réalisée. Cela permet d'observer la diminution des variations spécifiques aux centres (*Figure 62*). En particulier sur le panel 2, où la comparaison entre la correction de centre, première et deuxième version, semble confirmer une meilleure correction en utilisant un coefficient correcteur pour chaque population (*Figure 63*).

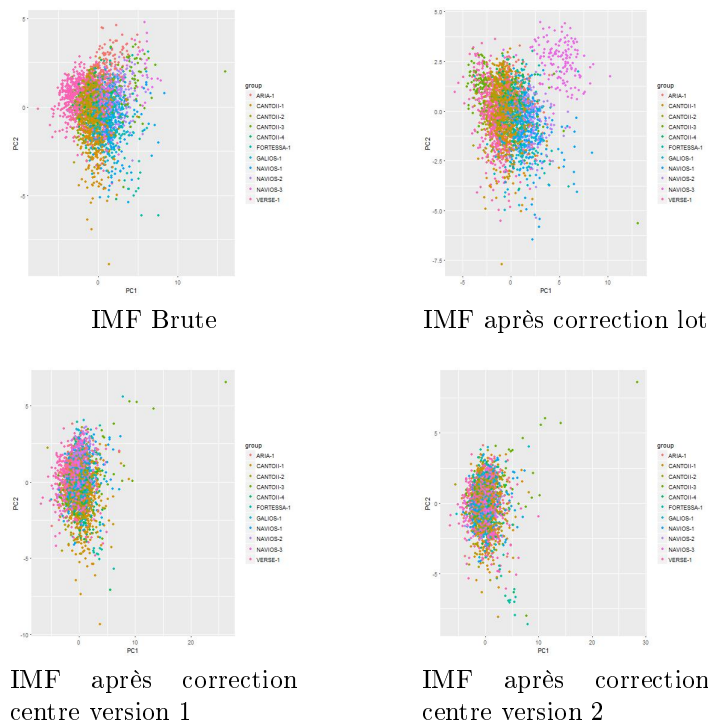


FIGURE 62: **Récapitulatif des corrections appliquées au panel 1.** Résumé des différentes ACP montrant l'effet des différentes corrections appliquées aux IMF du panel 1.

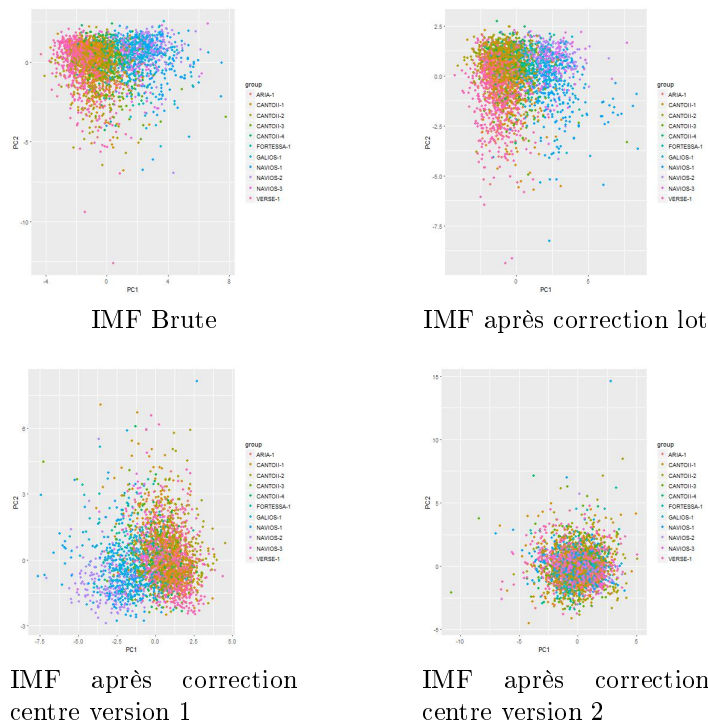


FIGURE 63: **Récapitulatif des corrections appliquées au panel 2.** Résumé des différentes ACP montrant l'effet des différentes corrections appliquées aux IMF du panel 2.

Les variations techniques à l'origine de l'effet centre semblent avoir été résolues lorsque l'on regarde les résultats de l'ACP. Les différents coefficients sont efficaces aussi bien pour le panel 1 (*Figure 62*) que pour le panel 2 (*Figure 63*).

Le calcul et l'application des coefficients sont réalisés à l'aide d'un script écrit en Python (*Annexe A.7.3, page 271*).

Grâce à ces coefficients, les données d'IMF peuvent être utilisées dans des analyses bioinformatiques avec les fréquences et les concentrations cellulaires.

L'ensemble de cette procédure fait l'objet d'un article en cours de soumission à la revue *Nature Methods* (*Annexe A.3, page 111*).

3 DISCUSSION

L'objectif de comparabilité des données de cytométrie fait face à de multiples obstacles. En effet, de nombreux points peuvent être sources de variabilités. Plusieurs études précédentes ont essayé de mettre en place une standardisation pour la cytométrie en flux s'attaquant à limiter les différentes sources de variations [7, 8, 20, 21, 22, 23, 24, 25, 26].

3.1 Principaux résultats de mon travail

Mes travaux ont abouti à plusieurs articles, publié ou en cours de soumission, traitant sur différents aspects de la cytométrie en flux, allant de la préparation des acquisitions par l'harmonisation des cytomètres (*Annexe A.4*) à l'utilisation des données de cytométrie dans le cadre de l'étude PRECISESADS (*Annexe A.5*) en passant par le maintien de la comparabilité par différentes méthodes (*Annexe A.3*).

En effet, pour l'étude PRECISESADS, un travail important a été effectué afin d'assurer la comparabilité des données par la mise en place de l'harmonisation inter-centre impliquant des billes de capture. Le maintien de cette calibration se fait par l'intermédiaire des billes 8 pics et de leur utilisation lors de chaque inclusion.

Lorsque les billes 8 pics ont échoué dans l'alignement avec leur référence, le script de normalisation permet de réduire ces variations et il est appliqué sur les fichiers du panel. Comme présenté dans le test de la normalisation, les CV des billes 8 pics sont passés de 14.53% à -0.01% en moyenne. Pour le panel, les valeurs sont inférieures à 5% par rapport à la référence .

Un des résultats de cette étude est la détermination manuelle de la compensation suivis par leur analyse automatisée sur plus de 12000 fichiers de cytométrie. Cette accomplissement est permis, en partie, par l'automatisation de l'analyse par Altrabio. Les résultats des différents automates développés pour PRECISESADS atteignent des corrélations supérieurs à 0,99 lorsqu'ils sont comparés à l'analyse manuelle.

A partir de cette analyse, des effets lot et centre ont pu être identifiés sur les IMF. Remettant en cause l'utilisation des IMF dans l'étude globale mais pas celle des fréquences et nombres absolus. Ces deux métriques semblent beaucoup moins sensibles aux variations entre différents cytomètres.

Concernant l'IMF, la solution choisie pour rendre utilisable cette métrique fut d'utiliser des coefficients spécifiques. Ces coefficients ont permis de réduire fortement les différences entre les lots et surtout les différences entre les centres. La détermination et l'application des coefficients est faite à l'aide de scripts python.

3.2 Les méthodes de calibration

Une source de variations peut être liée au cytomètre en flux lui-même. Particulièrement lorsque l'on compare plusieurs cytomètres, des différences dans les bancs optiques ou dans l'électronique propre à chaque machine sont observées entre les constructeurs.

Plusieurs études comparant les résultats entre les centres n'ont pas réalisé de calibration entre les cytomètres [7, 8, 21, 23, 26].

Pour contrer cette variation technique, la solution proposée par certaines études est une phase de calibration des machines entre elles à l'aide d'une IMF cible pour les différents canaux [20, 22, 24, 25]. La calibration des différentes machines se fait à l'aide de billes standardisées, tel que des billes 8 pics ou bien des billes de captures.

Face au même problème, le choix pour PRECISESADS a été de mettre en place une phase d'harmonisation des différentes machines, similaire à la solution proposée par Euroflow [20].

Cette calibration régulière des billes 8 pics permet d'assurer la stabilité dans le temps de l'IMF en intra-centre lorsqu'un prélèvement est analysé. En effet, des variations peuvent être engendrées à cause du vieillissement des différents composants du cytomètre et c'est le rôle des billes 8 pics de compenser ces variations techniques. Là où se démarque PRECISEADS est dans le degré de tolérance des variations. En effet, l'étude la plus stricte [20] tolère des variations jusqu'à 15% de l'IMF cible alors que pour PRECISESADS cette limite a été fixée à 5% de l'IMF cible afin de restreindre l'impact des variations.

L'ajout décisif dans le cadre de PRECISESADS est le maintien de l'IMF par la normalisation en intra-centre. Cette transformation des données permet de corriger les variations causées par un décalage dans l'alignement des billes 8 pics avec leur référence, qu'il soient inférieur à 5% ou non. La méthode de normalisation existait déjà mais n'avait jamais été appliquée à des études multi-centriques.

3.3 Mise en place d'un essai multi-centrique prospectif

3.3.1 Les points à considérer

— Réactif / Anticorps

Une première source de variabilité est induite par les réactifs, et plus particulièrement les couples fluorochromes/anticorps. Leur influence peut être sous deux formes :

En premier, il y a le réactif qui peut changer entre les lots ou en fonction de sa conservation. En effet, le rapport du nombre de molécules de fluorochromes

et du nombre de molécules d'anticorps peut être différents d'un lot à l'autre et entraîner des variations dans les fluorescences obtenues. De plus, l'utilisation classique des couples fluorochromes/anticorps se fait sous forme liquide où chaque composant d'un panel est conservé individuellement à des températures proches de 4 degrés Celcius et à l'abri de la lumière.

En deuxième, il y a la manipulation durant l'inclusion, chaque pipetage pouvant être une source d'erreur. Ce risque devient de plus en plus important avec l'augmentation du nombre d'inclusions et du nombre de panels. Seulement une étude [21] a adressé la manipulation des réactifs en utilisant un automate pour réaliser le pipetage.

Pour PRECISESADS, le choix a été fait d'utiliser les formulations duraclones. Ces tubes proposent des formulations fluorochromes/anticorps pré-mélangés sous forme sèche. Le premier avantage de ce type de tube est l'absence d'erreur dans le pipetage des volumes pour ces réactifs. L'autre avantage est la facilité de conservation de ces tubes sur de longues durées. En effet, ces tubes se conservent à température ambiante à l'abri de la lumière et ainsi permettent de faciliter le stockage.

— Billes de calibrations

Le paramétrage des machines est une étape importante dans les essais multicentrique. En effet, c'est par ce biais que les machines de différents constructeurs peuvent être comparables. Pour cela des études ont proposé d'utiliser des billes de calibration.

Les billes de calibration, ou billes de captures, permettent d'obtenir une IMF standardisée. L'autre avantage de ces billes vient de leur capacité à capturer des anticorps en suspension, couplé avec des tubes contenant un seul fluorochrome/anticorps du panel d'intérêt. Grâce à cela, le spectre de fluorescence correspond à celui utilisé dans les panels (contrairement aux billes 8 pics).

— SOP

Une seconde source de variabilité peut être trouvée dans la pratique de chaque laboratoire. La mise en place de SOPs permet une standardisation des procédures pour avoir une meilleure reproductibilité et traçabilité entre les laboratoires. Ces protocoles doivent couvrir l'acquisition d'un patient, la stratégie d'analyse et la manière de centraliser les données sur un serveur.

Cette méthode de standardisation semble être le premier pas vers la comparabilité des données [20, 21, 8, 23]. Similairement pour PRECISESADS, des SOP ont été implémentés pour décrire différentes parties, allant de l'acquisition des données à l'envoi sur un serveur distant *Annexe A.6*.

— Analyse des données

L'analyse des données de cytométrie est une étape très consommatrice en temps, que cela soit pour la détermination de la matrice de compensation ou l'exécution du gating. Pour ces deux parties, il est préférable de centraliser l'analyse à un opérateur unique tout au long de l'étude. En effet, l'utilisation de plusieurs opérateurs peut induire des variations pour la compensation ou pour l'analyse [7].

L'automatisation de la compensation n'est pas encore opérationnelle. La compensation a été réalisée manuellement pour environ 12000 fichiers. Cette compensation manuelle, bien que réalisée par un opérateur unique, reste cependant une source de variation possible. De plus, cette étape manuelle est extrêmement chronophage. A elle seule, elle a nécessité plus de 5 mois de travail pour générer la totalité des matrices.

Une analyse de répétabilité a été réalisée afin d'évaluer l'impact de l'action d'un opérateur unique. Plus de 100 fichiers ont été réanalysés à plusieurs mois d'intervalle et les valeurs des matrices de compensation ont été comparées (*Figure 35*). Les résultats montrent la très bonne répétabilité et excluent donc l'introduction de variabilité induite par l'opérateur.

L'automatisation de l'analyse est devenue une nécessité face à la quantité de données produites. Parmi la liste des publications des études multicentrique de cytométrie en flux, une seule essaie de mettre en place une automatisation de l'analyse en comparant 2 algorithmes : flowDensity et Opencyto [23].

Bien qu'il existe plusieurs autres algorithmes pour réaliser cette automatisation [42, 19, 9], c'est l'automate d'Altrabio qui nous permet d'obtenir des résultats très similaires à ceux de l'analyse manuelle (*Figure 45*). Outre le gain de temps, ces méthodes assurent une stabilité dans l'analyse du début à la fin du projet.

Cette automatisation de l'analyse nécessite de mettre en place des contrôles de qualité impliquant des connaissances biologiques pour reconnaître l'origine possible de fichiers aberrants.

Dans le cadre de PRECISESADS, les données d'environ 3000 patients ont été compensées et analysées. Les données ainsi produites ont été incluses dans l'analyse "OMIQUE" globale réalisée pour le projet et fait l'objet d'un article en cours de soumission à la revue Nature (*Annexe A.3*). La prochaine étape de ces analyses est d'utiliser des algorithmes non-supervisés pour explorer plus en avant la quantité de données. Pour cela il est nécessaire d'avoir une IMF comparable entre les différents centres.

3.3.2 Les scripts développés pour palier à chacun des aspects critiques

- Script de normalisation intra-centre.

Comme observé lors du test de normalisation, un décalage dans l'IMF des billes 8 pics, même léger, a un impact sur l'IMF des panels. Cette variation est purement d'origine technique et doit être corrigée.

Pour cela le script de normalisation intra-centre permet de calculer les paramètres de correction, efficace et applicable, pour chacun des canaux.

Cette normalisation a pour avantage de se baser sur des billes standardisées. De plus, elle permet de renforcer la stabilité en intra-centre avec les références définies avec des billes de calibrations.

- Script de correction de l'effet lot.

Les variations dans les données de cytométrie ont de multiples sources possibles. Une des sources de variations est observée en intra-centre. Appelées effet lot, ces variations correspondent à des dates de calibration entre les centres lors des changements de lots. L'hypothèse actuelle est que l'influence des variations entre les lots de duraclones a été sous estimée lors de l'utilisation du kit de compensation duraclone et que de nouvelles valeurs cibles auraient du être définies pour chaque nouveau lot.

Le calcul des coefficients de correction de l'effet lot a permis de réduire les différences des IMF pour les stabiliser au sein de chaque centre. Cependant, selon nos ACP, l'application de cette correction par canaux ne permet pas de corriger les différences observées entre les centres.

Les résultats des Immunotrol, sur les fréquences cellulaires, indiquent une bonne reproductibilité entre les différents centres. Les différences qui persistent concernent la correction des données sur les IMF.

- Script de correction de l'effet centre.

L'observation de différences significatives entre les centres ne nous permet pas de comparer les IMF et ce malgré l'étape de calibration. En effet, la calibration entre les différents centres avec les billes de capture montre qu'il est possible d'obtenir 1 pic positif avec des IMF comparables dans chaque centre. Par contre, le comportement du pic négatif présente des variations importantes entre les centres. Cela voudrait dire que la calibration a marché pour l'intensité de fluorescence du pic positif mais pas pour le reste du spectre des intensités de fluorescence. L'origine possible de ces variations peut être trouvée dans le comportement des fluorochromes qui serait influencé par l'anticorps, par des

différences dans les bancs optiques (filtres) et par la valeur de son intensité de fluorescence.

Les centres n'incluant pas forcément le même nombre de patients dans chaque classe de maladie, cela peut être une source de biais dans certains centres. L'option serait d'utiliser uniquement des témoins sains lors des comparaisons inter-centres. Malheureusement, leur relatif faible nombre et leur absence dans certains centres n'en font pas une solution pour cette étude.

La correction de cet effet centre à l'aide d'un coefficient spécifique à la population et au centre, permet d'obtenir des IMF comparables. Ainsi, après l'application du script dédié, les IMF peuvent être utilisées dans l'étude pour ajouter des informations précieuses.

Bien que la correction apportée semble fonctionner entre les IMF provenant des différents centres, l'un de mes objectifs était aussi d'avoir la possibilité de mettre en évidence de nouvelles populations cellulaires. La correction par coefficient marche pour l'IMF de la population cellulaire sélectionnée mais pour une autre population cellulaire, sur le même canal, le coefficient peut être différent. La présence de plusieurs coefficients pour le même canal empêche leur application sur les fichiers de cytométrie, et par conséquent l'analyse de ces fichiers à l'aide de méthodes de clustering non-supervisées est impossible.

La réduction des variations est possible grâce à une combinaison des méthodologies décrites pré- et post- analytiques.

Les différents scripts pour la comparabilité des données sont fonctionnels mais des améliorations peuvent être apportées.

Concernant le script de normalisation intra-centre, l'amélioration possible est de le rendre plus simple d'utilisation afin de pouvoir l'intégrer facilement dans d'autres études.

Un autre point est de revoir le choix d'appliquer uniquement le coefficient si l'intercepte est supérieur à 1. En effet, un paramètre dépassant cette valeur modifie trop les populations dites négative. L'amélioration de cette partie pourrait venir dans la modification de l'équation de normalisation pour mieux prendre en compte ces populations négatives.

Pour l'effet lot, l'amélioration se porte plus en amont de l'analyse, avec le changement de références lors du changement de lots.

Dernièrement, la correction de l'effet centre bénéficierait grandement de pouvoir être normalisé avec les billes de capture et le kit de compensation. Permettant ainsi

de normaliser les fichiers fcs et débloquer les analyse non-supervisées.

Des premiers essais ont été réalisé mais non fructueux, cette normalisation inter-centre semble nécessiter le développement de nouvelles billes de capture. Ces nouvelles billes de capture doivent posséder plusieurs pics dans le but de couvrir plus largement le spectre des intensités de fluorescence.

3.4 Conclusion générale

Comme décrit dans l'introduction, plusieurs études ont essayé de standardiser les résultats de cytométrie. Certaines se concentrent sur la partie pré-analytique avec l'établissement de protocoles, la mise en place de calibration utilisant des billes 8 pics ou des billes de calibration. D'autres s'intéressent à la partie analytique avec la centralisation de l'analyse et son automatisation. De plus, la majorité des études comparent principalement les fréquences, délaissant ainsi l'IMF qui peut être une source d'information.

Les travaux combinant toutes ces approches et s'intéressant à l'IMF le font en tolérant des variations de 15% [20] ou plus. De plus, aucune étude multi-centrique n'a proposé d'appliquer une approche informatique de normalisation des données se basant sur les billes 8 pics et des corrections post-analytiques.

Plusieurs points sont à surveiller lors d'études multi-centriques en cytométrie en flux. Premièrement, les centres doivent avoir bien paramétré leur cytomètre. C'est à dire qu'il ne doit pas y avoir de seuil de détection trop faible rendant inutilisable des fichiers de cytométrie. Deuxièmement, il faut s'assurer de la bonne conservation des tubes de cytométrie et d'une utilisation évitant le découplage des fluorochromes en tandem. Troisièmement, il faut s'assurer du bon report des voltages des billes 8 pics à ceux des panels car c'est sur cela que se base une bonne normalisation. bien évidemment, un grand nombre de ces remarques peuvent aussi être appliquées pour des études plus restreintes avec 1 seul cytomètre.

L'étude PRECISESADS nous permet de proposer un retour d'expérience et un protocole pour gérer les études de cytométrie multi-centriques prospectives en prenant en compte l'effet centre.

Un point important pour toute analyse multi-centrique est la définition de SOPs très précis envoyés aux différents membres de l'étude. Cela évite d'introduire des variations dues aux habitudes diverses de chaque centre.

RÉACTIFS Le choix de réactifs est une étape importante à prendre en compte dans l'élaboration d'une étude. Pour la cytométrie en flux, il faut prendre en compte les capacités des différentes machines en terme de lasers et de PMTs. Ce sont ces

paramètres qui vont déterminer le choix des fluorochromes utilisables dans l'étude.

Une fois la combinaison de fluorochromes définie et les anticorps sélectionnés, il est préférable de les utiliser sous forme sèche, comme avec les duraclones, et ce afin de réduire les problèmes de manipulation entre les différents centres du projet.

De plus, une limitation du nombre de lots est fortement recommandée pour faciliter toute correction que l'utilisation de plusieurs lots différents pourrait entraîner.

Le point d'amélioration principal, par rapport à PRECISESADS, pourrait venir avec le développement de nouvelles billes de standardisation. L'idée étant d'avoir un mélange entre des billes de capture et des billes 8 pics. Ces nouvelles billes de capture présenteraient plusieurs pics positifs (contrairement aux billes de capture actuellement proposée par les différents fabricants).

L'objectif de cette amélioration serait de couvrir un spectre plus large des intensités de fluorescence que les billes de captures classiques et ce en utilisant des fluorochromes des panels de cytométrie.

CALIBRATION La calibration des cytomètres est très similaire à celle décrite précédemment pour PRECISESADS, avec une harmonisation des cytomètres en utilisant 2 pics comme cibles déterminées dans un centre de référence, et les autres centres qui modifient leurs voltages pour atteindre ces cibles $\pm 5\%$. Une fois cette calibration validée, elle est suivie au quotidien de l'utilisation des billes 8 pics.

Il est conseillé de réaliser des calibrations régulièrement (tous les 3 à 4 mois) afin de définir plusieurs références pour les billes 8 pics, tout en prenant en compte les changements de lots. En effet pour les changements de lots, le centre référence doit réaliser une calibration comparée avec l'ancien lot, passer le nouveau lot avec les paramètres ainsi définis et générer les nouvelles cibles qu'il transmet alors aux autres centres de l'étude.

Avant l'acquisition d'un nouveau patient, les billes 8 pics doivent être passées pour paramétrer le cytomètre avec les références déterminées lors de la dernière calibration pour chaque canal.

ANALYSES Avant de réaliser l'analyse, il est nécessaire de normaliser les données. Il faut d'abord normaliser en intra-centre. Cette normalisation se base sur les billes 8 pics et permet de corriger les erreurs lors de l'acquisition et ainsi réduire les variations.

Grâce aux nouvelles billes de calibration, une normalisation inter-centre est possible. Cette normalisation se base sur les billes de capture au spectre large et, similairement à la normalisation intra-centre, utilise une régression linéaire pour déterminer

les paramètres de normalisation. Par cette méthode, les variations inter-centres devraient être réduites.

Une fois les données normalisées, la phase de compensation commence. Cette phase reste l'étape la plus limitante de l'analyse des données de cytométrie car elle nécessite d'être réalisée par un opérateur unique de manière manuelle. La réalisation de mono-marquages avant l'acquisition de chaque patient serait idéale mais assez peu réaliste dans la pratique quotidienne avec l'inclusion de milliers de patients.

L'exploitation des fichiers de cytométrie doit et peut être automatisée. Les algorithmes d'automatisation de l'analyse s'améliorent chaque année et permettent d'obtenir des résultats stables qui se rapprochent de l'analyse manuelle. Une première approche peut être faite lorsque l'on s'intéresse à des populations connues. Elle consiste en une analyse supervisée imitant le gating manuel, tel que celle proposé par Altrabio ou bien Opencyto. La deuxième approche est plus dans la découverte par une analyse non-supervisée des données. Cette dernière est possible grâce à une normalisation inter-centre fonctionnelle. Elle peut être réalisée par des outils comme FLOCK.

Il est important de réaliser des contrôles, par utilisation de réactifs standardisés comme des Immunotrol, sur les différents points comme la calibration ou sur les résultats produits par l'automatisation. En effet, lors d'une étude multi-centrique de nombreux paramètres peuvent être une source de problèmes (fichiers manquant, noms des canaux différents, ...).

Une option possible serait de garder les mêmes voltages que ceux déterminés lors de la première calibration tout au long de l'étude. L'idée est que la normalisation intra-centre et inter-centre devraient corriger toutes les variations en ayant l'avantage de réduire les erreurs lors des changements de voltages. De plus, il faudrait observer l'influence de ce voltage constant sur les valeurs des compensations.

Ces propositions devraient permettre à la cytométrie de faire face aux différents problèmes de comparabilité des données avec une emphase sur la standardisation et la reproductibilité des résultats. Ces nouvelles procédures sont d'autant plus pertinentes aujourd'hui avec l'entrée de la cytométrie en flux dans l'ère de la bioinformatique, de par le développement de bases de données publiques comme Flow Repository, l'augmentation des outils sous R et sous Python et l'augmentation du nombre de patients inclus dans les nouvelles études.

Références

- [1] B. H. Davis, J. Holden, M. Bene, M. Borowitz, R. Braylan, D. Cornfield, W. Gorczyca, R. Lee, R. Maiese, A. Orfao, *et al.*, “2006 bethesda international consensus recommendations on the flow cytometric immunophenotypic analysis of hematolymphoid neoplasia : medical indications,” *Cytometry Part B : Clinical Cytometry : The Journal of the International Society for Analytical Cytology*, vol. 72, no. S1, pp. S5–S13, 2007.
- [2] R. Nguyen, S. Perfetto, Y. D. Mahnke, P. Chattopadhyay, and M. Roederer, “Quantifying spillover spreading for comparing instrument performance and aiding in multicolor panel design,” *Cytometry Part A*, vol. 83, no. 3, pp. 306–315, 2013.
- [3] R. F. Murphy and T. M. Chused, “A proposal for a flow cytometric data file standard,” *Cytometry : The Journal of the International Society for Analytical Cytology*, vol. 5, no. 5, pp. 553–555, 1984.
- [4] P. N. Dean, C. B. Bagwell, T. Lindmo, R. F. Murphy, and G. C. Salzman, “Data file standard for flow cytometry,” *Cytometry*, vol. 11, no. 3, pp. 323–332, 1990.
- [5] L. Seamer, C. Bagwell, L. Barden, D. Redelman, G. Salzman, J. Wood, and R. Murphy, “Proposed new data file standard for flow cytometry, version fcs 3.0,” *Cytometry : The Journal of the International Society for Analytical Cytology*, vol. 28, no. 2, pp. 118–122, 1997.
- [6] D. Parks, “Data file standard for flow cytometry version fcs 3.1 normative reference,” 2008.
- [7] L. Westera, T. Van Viegen, J. Jeyarajah, A. Azad, J. Billsborough, G. R. Van Den Brink, J. Cremer, S. Danese, G. D’haens, L. Eckmann, *et al.*, “Centrally determined standardization of flow cytometry methods reduces interlaboratory variation in a prospective multicenter study,” *Clinical and translational gastroenterology*, vol. 8, no. 11, p. e126, 2017.
- [8] M. Streitz, T. Miloud, M. Kapinsky, M. R. Reed, R. Magari, E. K. Geissler, J. A. Hutchinson, K. Vogt, S. Schlickeiser, A. H. Kverneland, *et al.*, “Standardization of whole blood immune phenotype monitoring for clinical trials : panels and methods from the one study,” *Transplantation research*, vol. 2, no. 1, p. 17, 2013.
- [9] N. Aghaeepour, G. Finak, H. Hoos, T. R. Mosmann, R. Brinkman, R. Gottardo, R. H. Scheuermann, F. Consortium, D. Consortium, *et al.*, “Critical assessment of automated flow cytometry data analysis techniques,” *Nature methods*, vol. 10, no. 3, pp. 228–238, 2013.
- [10] K. O’Neill, N. Aghaeepour, J. Špidlen, and R. Brinkman, “Flow cytometry bioinformatics,” *PLoS computational biology*, vol. 9, no. 12, p. e1003365, 2013.
- [11] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, *et al.*, “Bioconductor : open software development for computational biology and bioinformatics,” *Genome biology*, vol. 5, no. 10, p. R80, 2004.

- [12] F. Hahne, N. LeMeur, R. R. Brinkman, B. Ellis, P. Haaland, D. Sarkar, J. Spidlen, E. Strain, and R. Gentleman, “flowcore : a bioconductor package for high throughput flow cytometry,” *BMC bioinformatics*, vol. 10, no. 1, p. 106, 2009.
- [13] D. Sarkar, N. Le Meur, and R. Gentleman, “Using flowviz to visualize flow cytometry data,” *Bioinformatics*, vol. 24, no. 6, pp. 878–879, 2008.
- [14] F. Hahne, N. Gopalakrishnan, A. H. Khodabakhshi, C.-J. Wong, and K. Lee, “flowstats : Statistical methods for the analysis of flow cytometry data,” *R Packag. version*, vol. 3, no. 1, 2009.
- [15] N. Aghaeepour, P. Chattopadhyay, M. Chikina, T. Dhaene, S. Van Gassen, M. Kursa, B. N. Lambrecht, M. Malek, G. McLachlan, Y. Qian, *et al.*, “A benchmark for evaluation of algorithms for identification of cellular correlates of clinical outcomes,” *Cytometry Part A*, vol. 89, no. 1, pp. 16–21, 2016.
- [16] P. Duchesne and J. Turgeon, “Flock : a method for quick mapping of admixture without source samples,” *Molecular ecology resources*, vol. 9, no. 5, pp. 1333–1344, 2009.
- [17] N. Aghaeepour, “flowtype : Phenotyping flow cytometry assays,” 2012.
- [18] N. Aghaeepour, A. Jalali, K. O’Neill, P. K. Chattopadhyay, M. Roederer, H. H. Hoos, and R. R. Brinkman, “Rchyoptymx : cellular hierarchy optimization for flow cytometry,” *Cytometry Part A*, vol. 81, no. 12, pp. 1022–1030, 2012.
- [19] K. Lo, F. Hahne, R. R. Brinkman, and R. Gottardo, “flowelust : a bioconductor package for automated gating of flow cytometry data,” *BMC bioinformatics*, vol. 10, no. 1, p. 145, 2009.
- [20] T. Kalina, J. Flores-Montero, V. Van Der Velden, M. Martin-Ayuso, S. Böttcher, M. Ritgen, J. Almeida, L. Lhermitte, V. Asnafi, A. Mendonca, *et al.*, “Euroflow standardization of flow cytometer instrument settings and immunophenotyping protocols,” *Leukemia*, vol. 26, no. 9, p. 1986, 2012.
- [21] M. Hasan, B. Beitz, V. Rouilly, V. Libri, A. Urrutia, D. Duffy, L. Casard, J. P. D. Santo, E. Mottez, L. Quintana-Murci, M. L. Albert, and L. Rogge, “Semi-automated and standardized cytometric procedures for multi-panel and multi-parametric whole blood immunophenotyping,” *Clinical Immunology*, vol. 157, no. 2, pp. 261 – 276, 2015.
- [22] F. Solly, L. Rigollet, L. Baseggio, J. Guy, J. Borgeot, E. Guérin, A. Debliquis, B. Drenou, L. Campos, F. Lacombe, *et al.*, “Comparable flow cytometry data can be obtained with two types of instruments, canto ii, and navios. a geil study,” *Cytometry Part A*, vol. 83, no. 12, pp. 1066–1072, 2013.
- [23] G. Finak, M. Langweiler, M. Jaimes, M. Malek, J. Taghiyar, Y. Korin, K. Radassi, L. Devine, G. Obermoser, M. L. Pekalski, *et al.*, “Standardizing flow cytometry immunophenotyping analysis from the human immunophenotyping consortium,” *Scientific reports*, vol. 6, p. 20686, 2016.
- [24] A.-E. Depince-Berger, K. Sidi-Yahya, M. Jeraiby, and C. Lambert, “Basophil activation test : Implementation and standardization between systems and between instruments,” *Cytometry Part A*, vol. 91, no. 3, pp. 261–269, 2017.

- [25] R. A. Hoffman, L. Wang, M. Bigos, and J. P. Nolan, "Nist/isac standardization study : Variability in assignment of intensity values to fluorescence standard beads and in cross calibration of standard beads to hard dyed beads," *Cytometry Part A*, vol. 81, no. 9, pp. 785–796, 2012.
- [26] M. C. Jaimes, H. T. Maecker, M. Yan, V. C. Maino, M. B. Hanley, A. Greer, J. M. Darden, and M. P. D'Souza, "Quality assurance of intracellular cytokine staining assays : analysis of multiple rounds of proficiency testing," *Journal of immunological methods*, vol. 363, no. 2, pp. 143–157, 2011.
- [27] D. M. O'Hara, Y. Xu, Z. Liang, M. P. Reddy, D. Y. Wu, and V. Litwin, "Recommendations for the validation of flow cytometric testing during drug development : Ii assays," *Journal of immunological methods*, vol. 363, no. 2, pp. 120–134, 2011.
- [28] J.-M. Berthelot, B. L. Goff, A. Neel, Y. Maugars, and M. Hamidou, "Netosis : At the crossroads of rheumatoid arthritis, lupus, and vasculitis," *Joint Bone Spine*, vol. 84, no. 3, pp. 255 – 262, 2017.
- [29] M. Howell and M. Shepherd, "The immune system," *Anaesthesia & Intensive Care Medicine*, 2018.
- [30] C. C. Goodnow, J. Sprent, B. F. de St Groth, and C. G. Vinuesa, "Cellular and genetic mechanisms of self tolerance and autoimmunity," *Nature*, vol. 435, no. 7042, p. 590, 2005.
- [31] J. H. Cho and M. Feldman, "Heterogeneity of autoimmune diseases : pathophysiologic insights from genetics and implications for new therapies," *Nature medicine*, vol. 21, no. 7, p. 730, 2015.
- [32] A. Fava and M. Petri, "Systemic lupus erythematosus : Diagnosis and clinical management," *Journal of autoimmunity*, 2018.
- [33] S. Ganapathy, V. Vedam, V. Rajeev, and R. Arunachalam, "Autoimmune disorders-immunopathogenesis and potential therapies.," *Journal of Young Pharmacists*, vol. 9, no. 1, 2017.
- [34] A. H. Talel Badri, "Scleroderma," 2018.
- [35] M. Kourilovitch, C. Galarza-Maldonado, and E. Ortiz-Prado, "Diagnosis and classification of rheumatoid arthritis," *Journal of autoimmunity*, vol. 48, pp. 26–30, 2014.
- [36] G. S. Firestein and I. B. McInnes, "Immunopathogenesis of rheumatoid arthritis," *Immunity*, vol. 46, no. 2, pp. 183–196, 2017.
- [37] R. I. Fox, "Sjögren's syndrome," *The Lancet*, vol. 366, no. 9482, pp. 321–331, 2005.
- [38] S. Miyakis, M. D. Lockshin, T. Atsumi, D. W. Branch, R. L. Brey, R. Cervera, R. H. Derksen, P. G. de Groot, T. Koike, P. L. Meroni, *et al.*, "International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (aps)," *Journal of Thrombosis and Haemostasis*, vol. 4, no. 2, pp. 295–306, 2006.
- [39] T. Ahsan, U. Erum, A. Dahani, and D. Khowaja, "Clinical and immunological profile in patients with mixed connective tissue disease.," *JPMA. The Journal of the Pakistan Medical Association*, vol. 68, no. 6, pp. 959–962, 2018.

- [40] M. Mosca, C. Tani, C. Neri, C. Baldini, and S. Bombardieri, “Undifferentiated connective tissue diseases (uctd),” *Autoimmunity reviews*, vol. 6, no. 1, pp. 1–4, 2006.
- [41] C. Jamin, L. L. Lann, D. Alvarez-Errico, N. Barbarroja, T. Cantaert, J. Duceux, A. M. Dufour, V. Gerl, K. Kniesch, E. Neves, E. Trombetta, M. Alarcón-Riquelme, C. Marañon, and J.-O. Pers, “Multi-center harmonization of flow cytometers in the context of the european “precisesads” project,” *Autoimmunity Reviews*, vol. 15, no. 11, pp. 1038 – 1045, 2016.
- [42] G. Finak, J. Frelinger, W. Jiang, E. W. Newell, J. Ramey, M. M. Davis, S. A. Kalams, S. C. De Rosa, and R. Gottardo, “Opencyto : an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis,” *PLoS computational biology*, vol. 10, no. 8, p. e1003806, 2014.
- [43] F. Hahne, A. H. Khodabakhshi, A. Bashashati, C.-J. Wong, R. D. Gascoyne, A. P. Weng, V. Seyfert-Margolis, K. Bourcier, A. Asare, T. Lumley, *et al.*, “Per-channel basis normalization methods for flow cytometry data,” *Cytometry Part A : The Journal of the International Society for Advancement of Cytometry*, vol. 77, no. 2, pp. 121–131, 2010.
- [44] A. Schwartz, E. Fernandez Repollet, R. Vogt, and J. W. Gratama, “Standardizing flow cytometry : construction of a standardized fluorescence calibration plot using matching spectral calibrators,” *Cytometry Part A*, vol. 26, no. 1, pp. 22–31, 1996.
- [45] P. N, “flowbeads : flowbeads : Analysis of flow bead data,” *R Packag. version*, 2019.
- [46] J. C. Stewart, M. L. Villasmil, and M. W. Frampton, “Changes in fluorescence intensity of selected leukocyte surface markers following fixation,” *Cytometry Part A : the journal of the International Society for Analytical Cytology*, vol. 71, no. 6, pp. 379–385, 2007.
- [47] J. Song, H. Wang, and M. J. Song, “Package ‘ckmeans. 1d. dp’,” *Optimal and Fast Univariate Clustering, July 2017*, 2018.
- [48] K. Hofmann, A.-K. Clauder, and R. A. Manz, “Targeting b cells and plasma cells in autoimmune diseases,” *Frontiers in immunology*, vol. 9, 2018.

A Annexe

A.1 Cytométrie report

PRECISESADS
Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases

FLOW CYTOMETRY REPORT
*Obligatoire

Site *
Institution

Flow Cytometer *
 Navios
 Galios
 Canto II
 Verse
 Aria
 Fortessa

Patient ID *

Cohort
 Inception
 Phase I
 Phase II

Day and Hour of sample drawing *
Jour ▼ Mois ▼ 2019 ▼ h ▼ : min ▼

Day and Hour of sample arrival *
Jour ▼ Mois ▼ 2019 ▼ h ▼ : min ▼

Day and Hour of sample preparation *
Jour ▼ Mois ▼ 2019 ▼ h ▼ : min ▼

Day and Hour of flow cytometry acquisition *
Jour ▼ Mois ▼ 2019 ▼ h ▼ : min ▼

PMT Voltages in accordance with QCB Peaks Beads *
 Yes
 No

Temperature stability *
 <22°C
 22-25°C
 >25°C

Flow Count Beads
Lot number of flow count beads

Concentration (beads/μL)

Modifier ce formulaire

FIGURE A1: Outils de suivi pour la cytométrie en flux (partie 1/2).

Duraclone tubes

Tube 1: Lot number

Tube 2: Lot number

Tube 3: Lot number
 NA if not applicable

Tube 4: Lot number
 NA if not applicable

Tube 5: Lot number
 NA if not applicable

Tube 6: Lot number
 NA if not applicable

Tube 7: Lot number
 NA if not applicable

Tube 8: Lot number
 NA if not applicable

Tube 9: Lot number
 NA if not applicable

Duraclone acquisitions

Minimum Number of cells reached : Tube 1 *
 Yes
 No

Minimum Number of cells reached : Tube 2 *
 Yes
 No

Minimum Number of cells reached : Tube 3 *
 Yes
 No
 NA

Minimum Number of cells reached : Tube 4 *
 Yes
 No
 NA

Minimum Number of cells reached : Tube 5 *
 Yes
 No
 NA

Minimum Number of cells reached : Tube 6 *
 Yes
 No
 NA

Minimum Number of cells reached : Tube 7 *
 Yes
 No
 NA

Minimum Number of cells reached : Tube 8 *
 Yes
 No
 NA

Minimum Number of cells reached : Tube 9 *
 Yes
 No
 NA

N'envoyez jamais de mots de passe via Google Forms.

100 % ; vous avez réussi.

Fourni par Google Forms

Ce contenu n'est ni rédigé, ni cautionné par Google.
 Signaler un cas d'utilisation abusive - Conditions d'utilisation - Clauses additionnelles

FIGURE A2: Outils de suivi pour la cytométrie en flux (partie 2/2).

A.2 Effet Lot et correction

Évolution de l'IMF du panel 1 au cours de PRECISESADS :

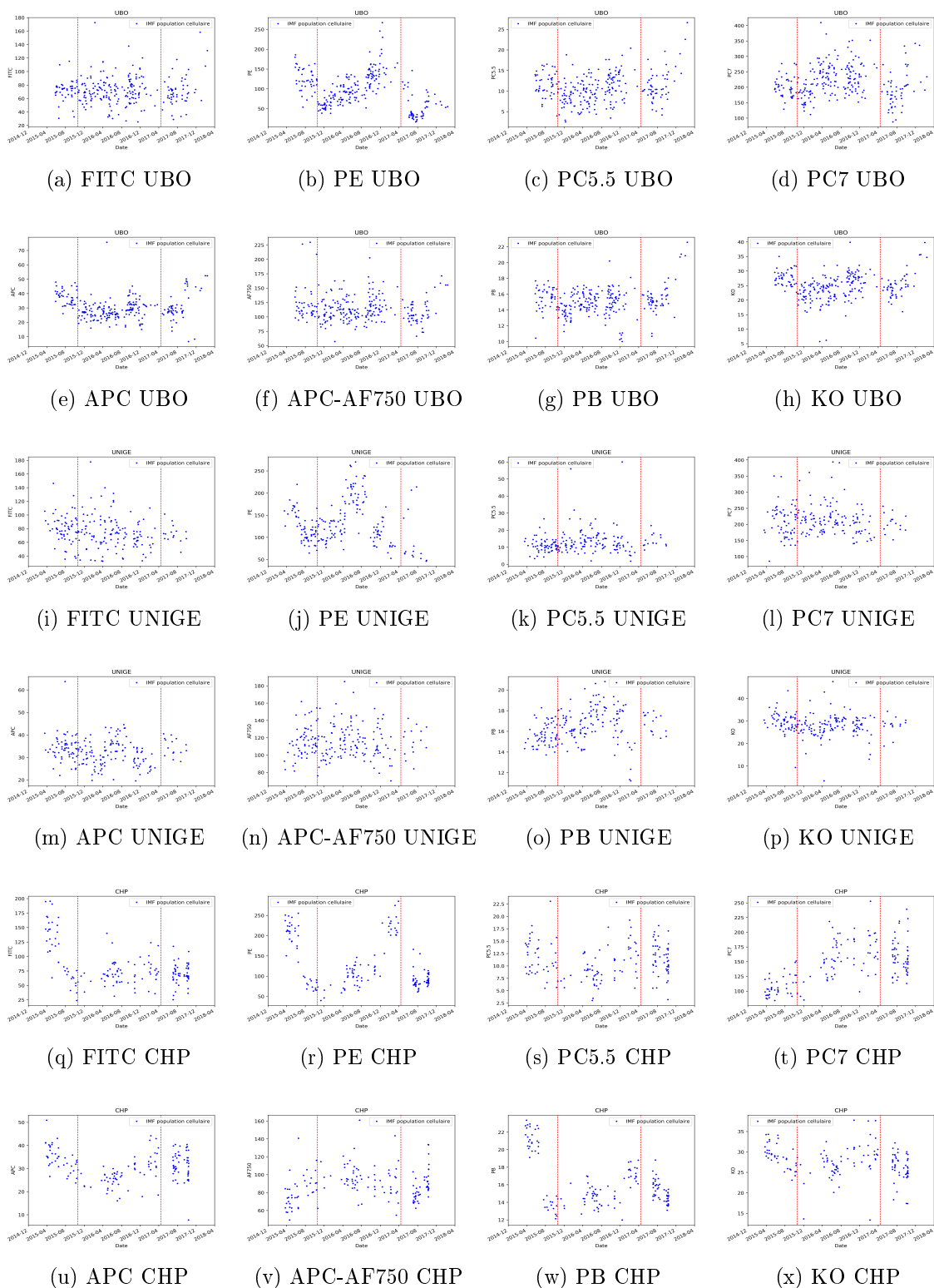


FIGURE A3: Évolution de l'IMF de chaque canal dans chaque centre (UBO,UNIGE,CHP). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre

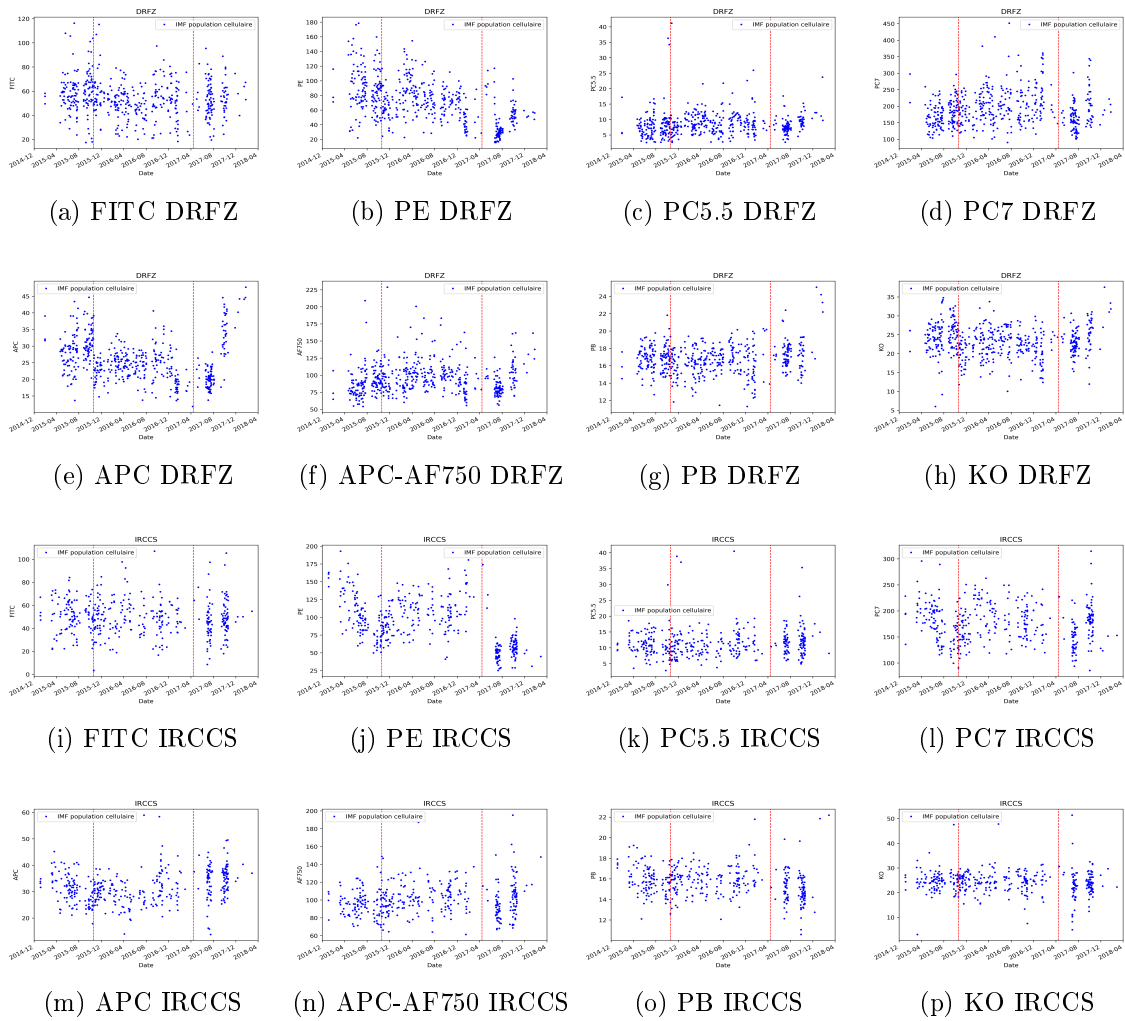


FIGURE A4: Évolution de l'IMF de chaque canal dans chaque centre (DRFZ,IRCCS). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre

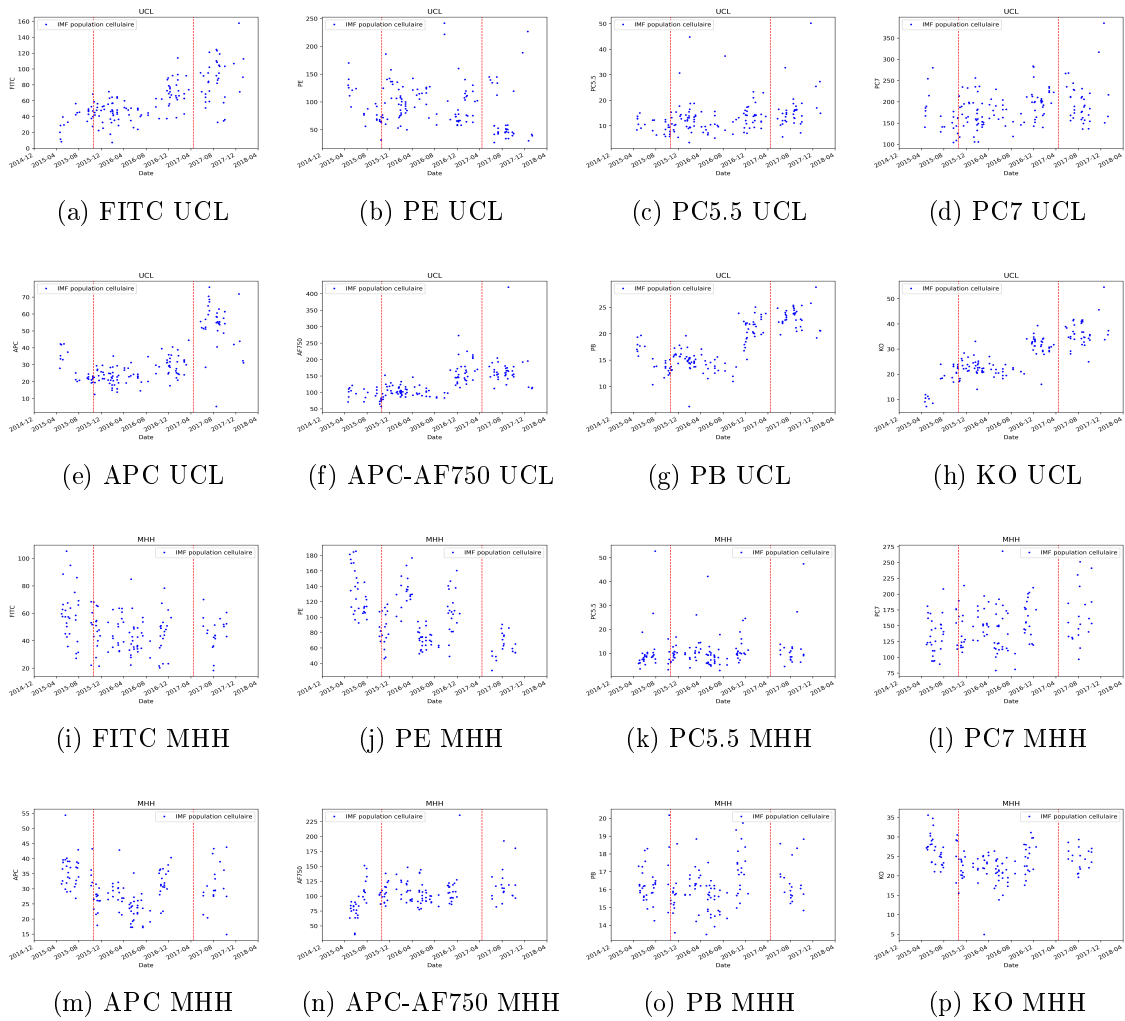


FIGURE A5: Évolution de l'IMF de chaque canal dans chaque centre (UCL,MHH). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre

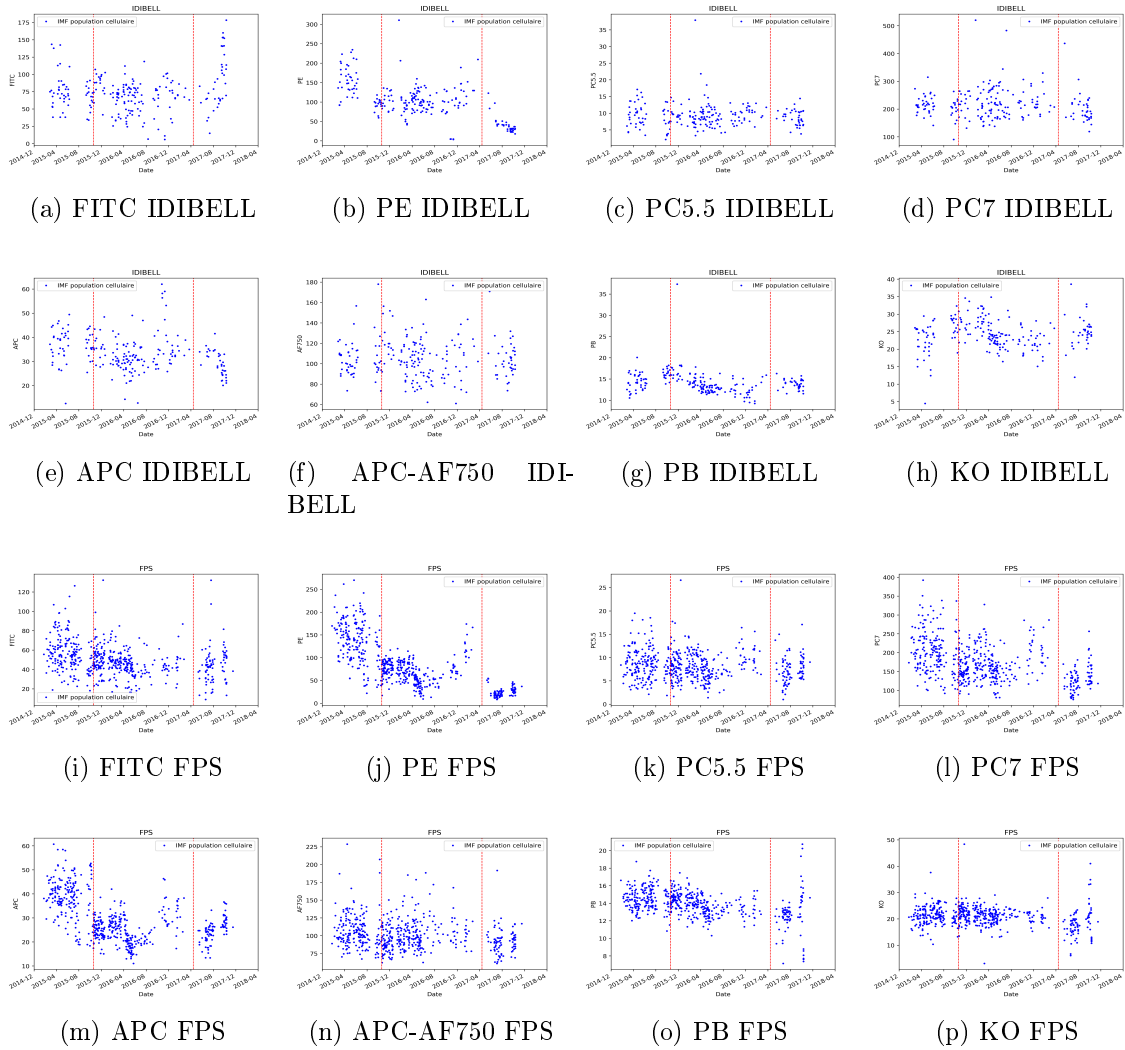


FIGURE A6: Évolution de l'IMF de chaque canal dans chaque centre (IDIBELL, FPS). Les lignes rouges correspondent aux deux changements de lot lors des calibrations inter centre

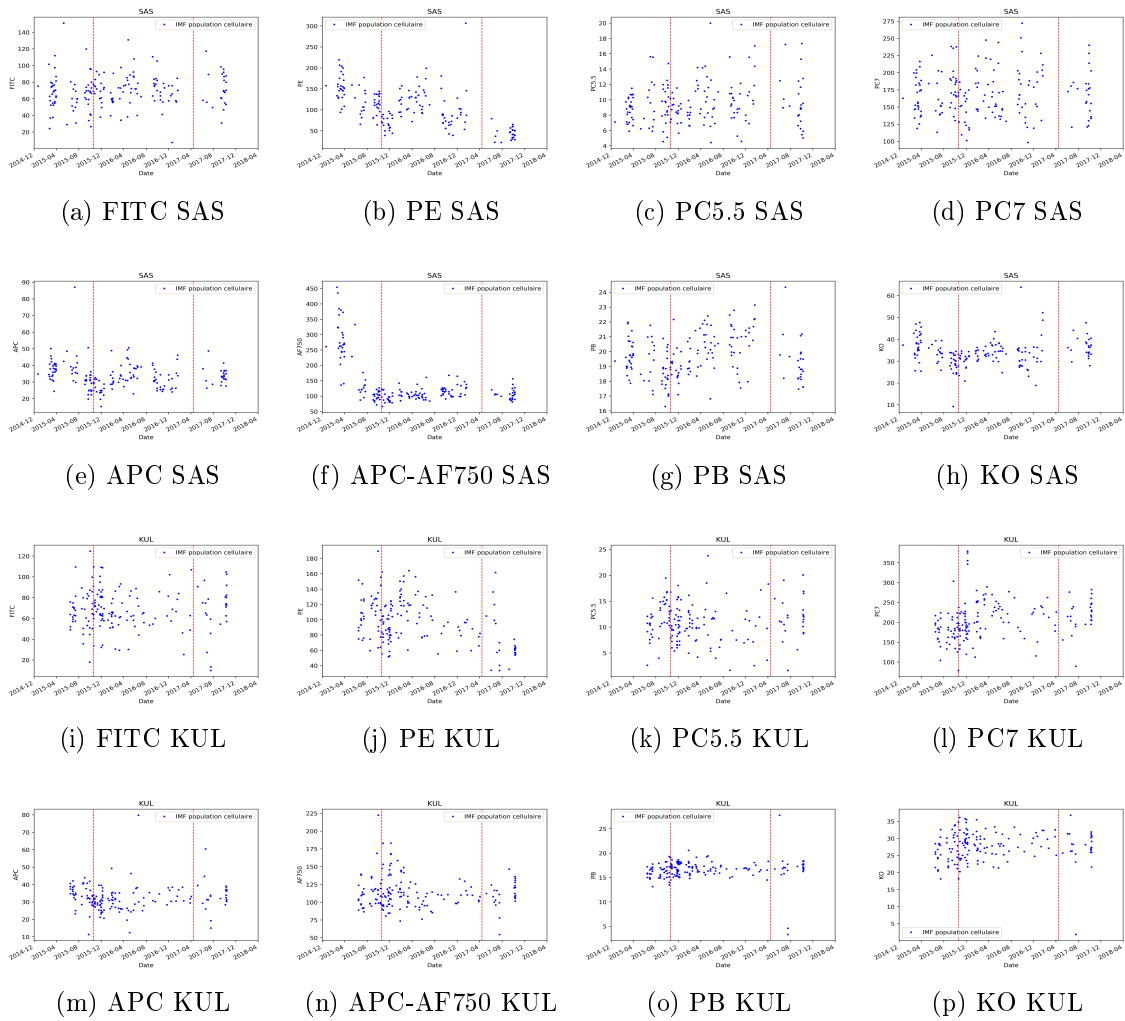


FIGURE A7: Évolution de l'IMF de chaque canal dans chaque centre (SAS,KUL). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre

Évolution de l'IMF du panel 1 au cours de PRECISESADS après correction :

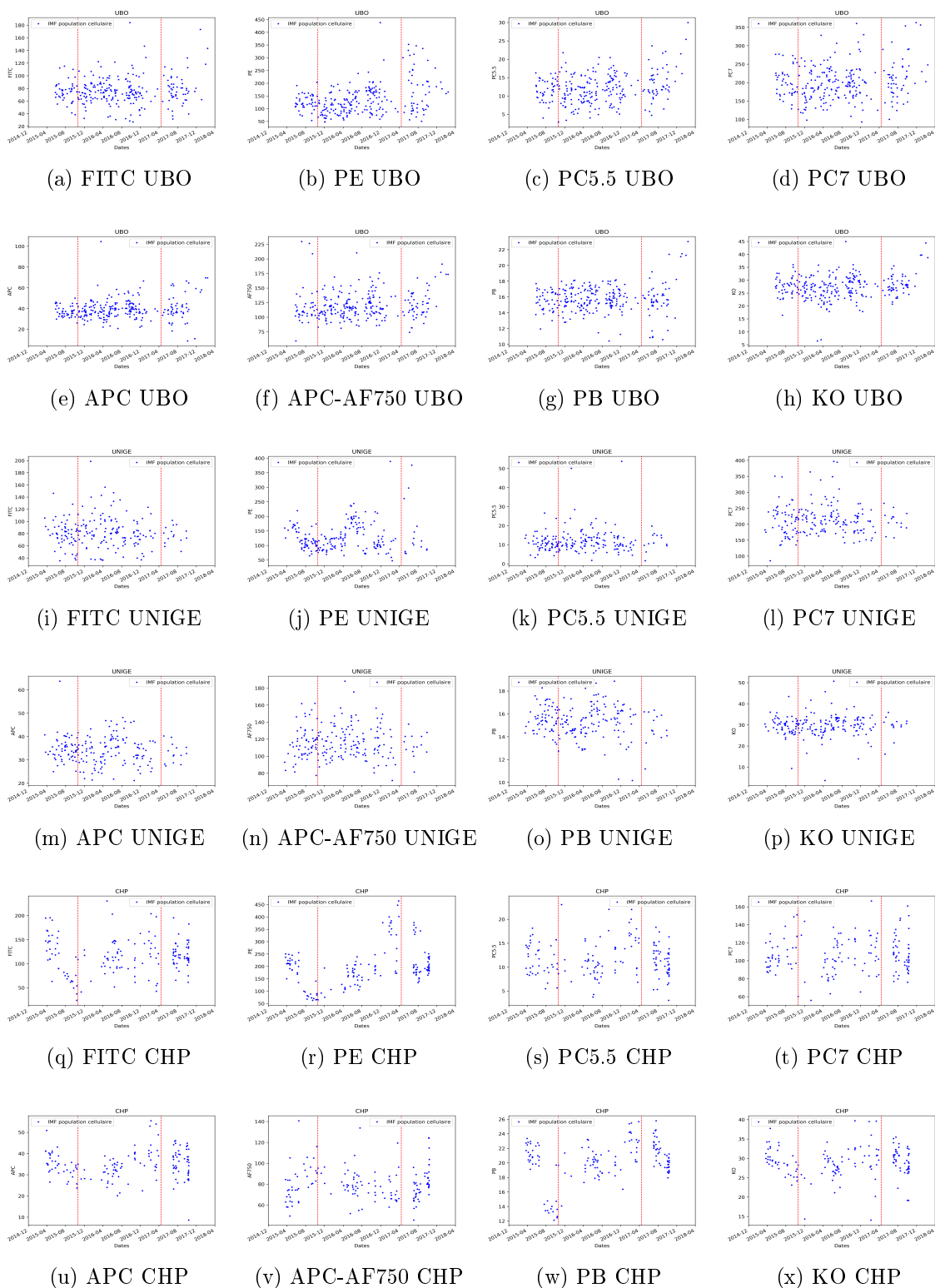


FIGURE A8: Évolution de l'IMF de chaque canal dans chaque centre (UBO,UNIGE,CHP). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre. L'application de la correction de l'effet lot a été appliqué.

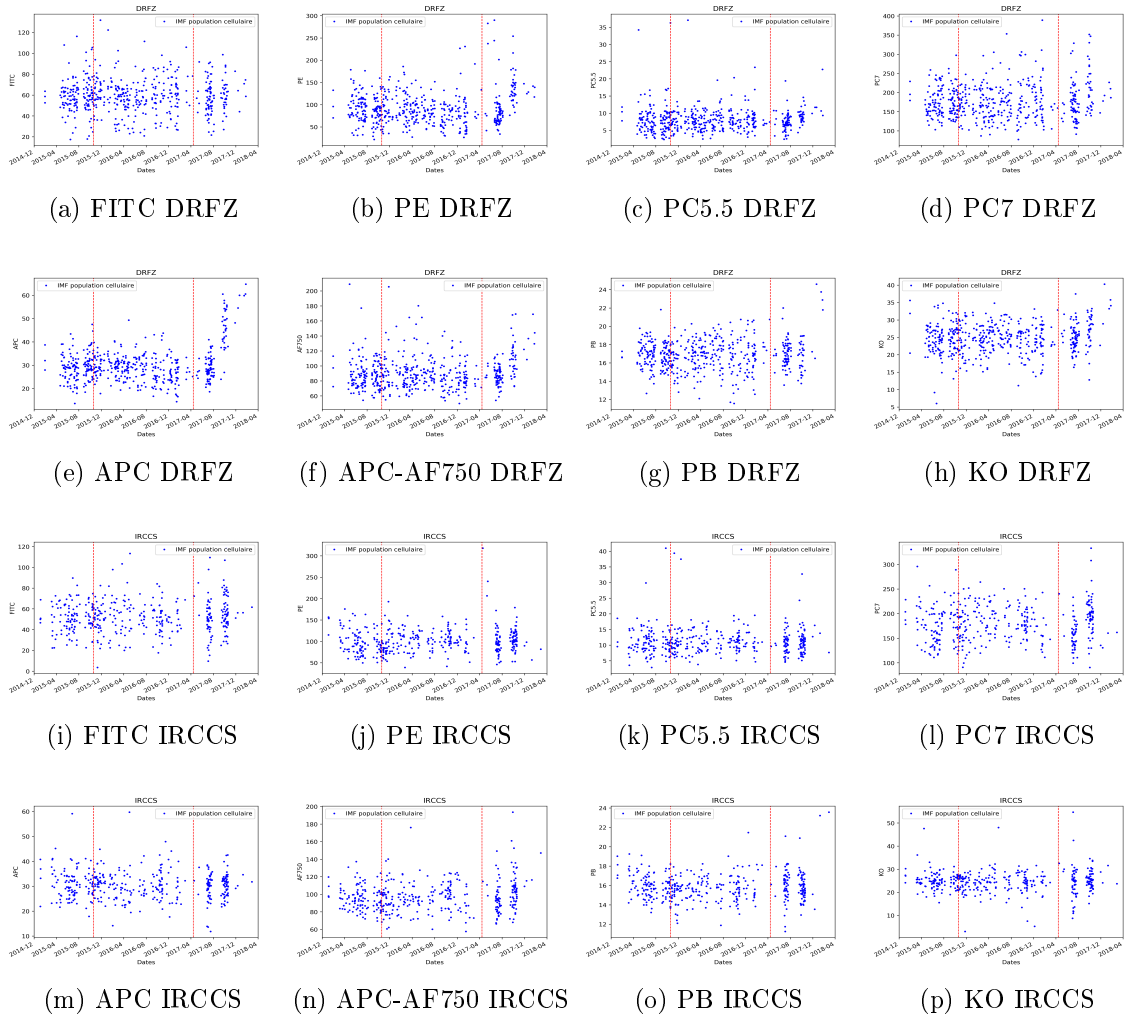


FIGURE A9: Évolution de l'IMF de chaque canal dans chaque centre (DRFZ,IRCCS). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre. L'application de la correction de l'effet lot a été appliqué.

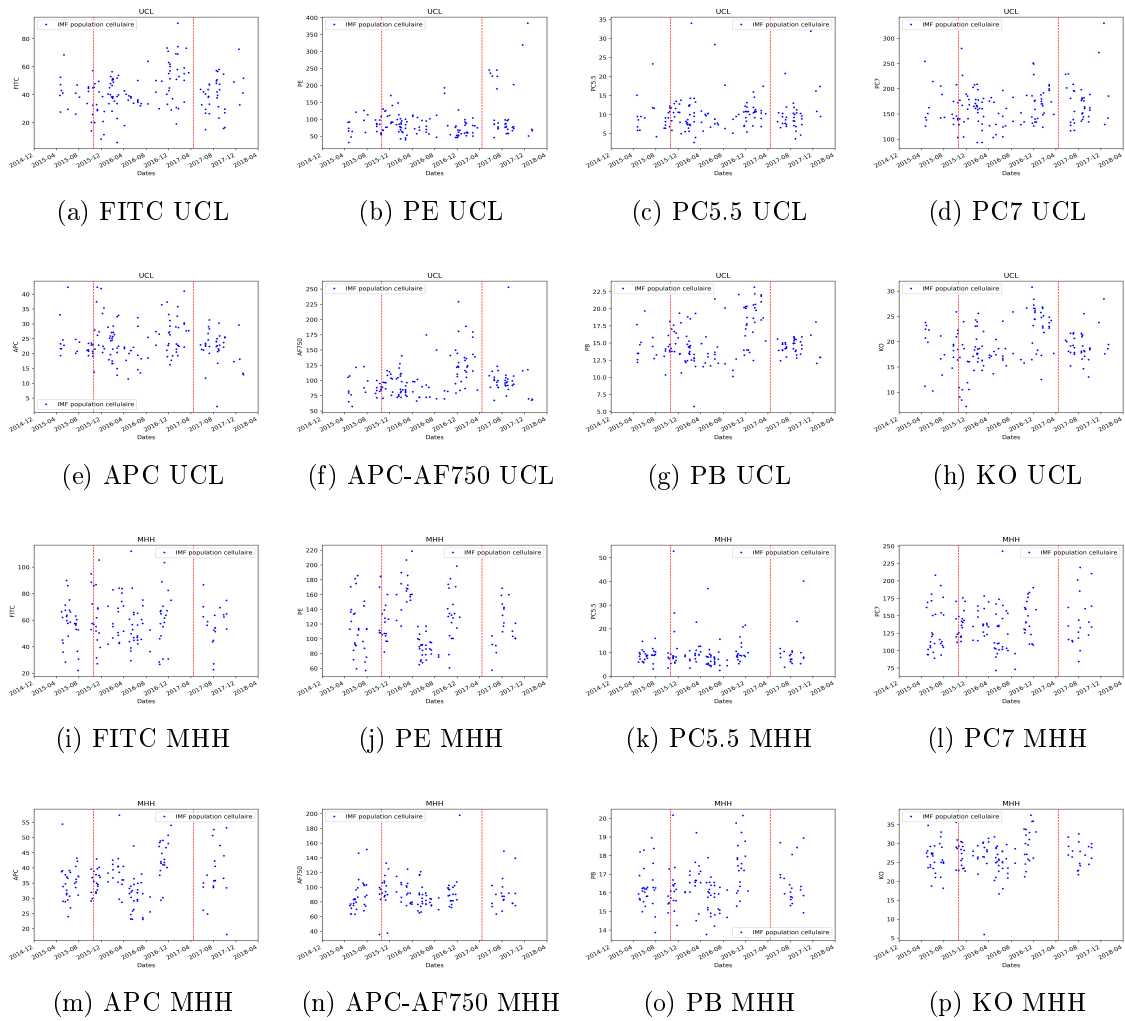


FIGURE A10: Évolution de l'IMF de chaque canal dans chaque centre (UCL,MHH). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre. L'application de la correction de l'effet lot a été appliqué.

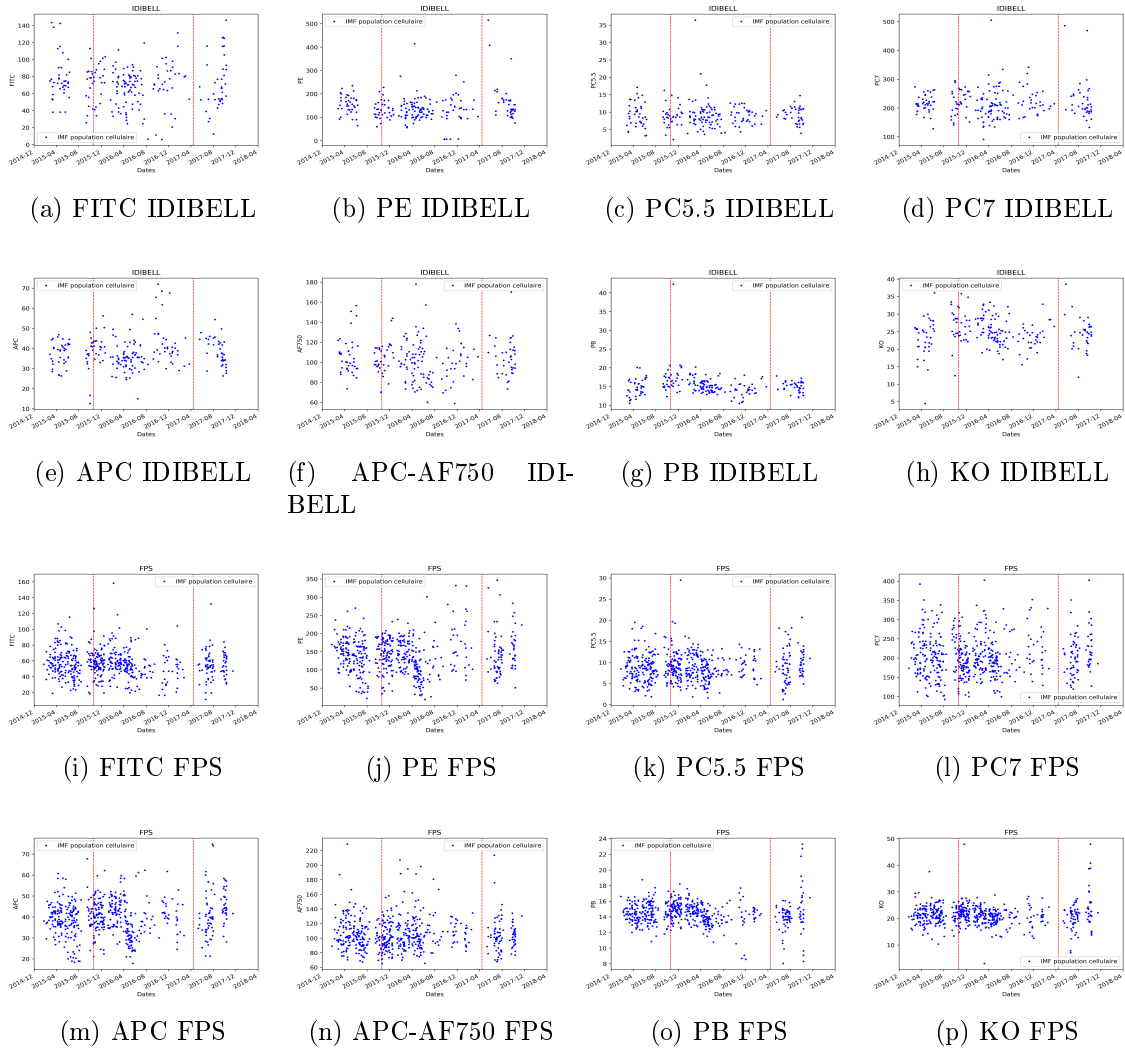


FIGURE A11: Évolution de l'IMF de chaque canal dans chaque centre (IDIBELL, FPS). Les lignes rouges correspondent aux deux changements de lot lors des calibrations inter centre. L'application de la correction de l'effet lot a été appliquée.

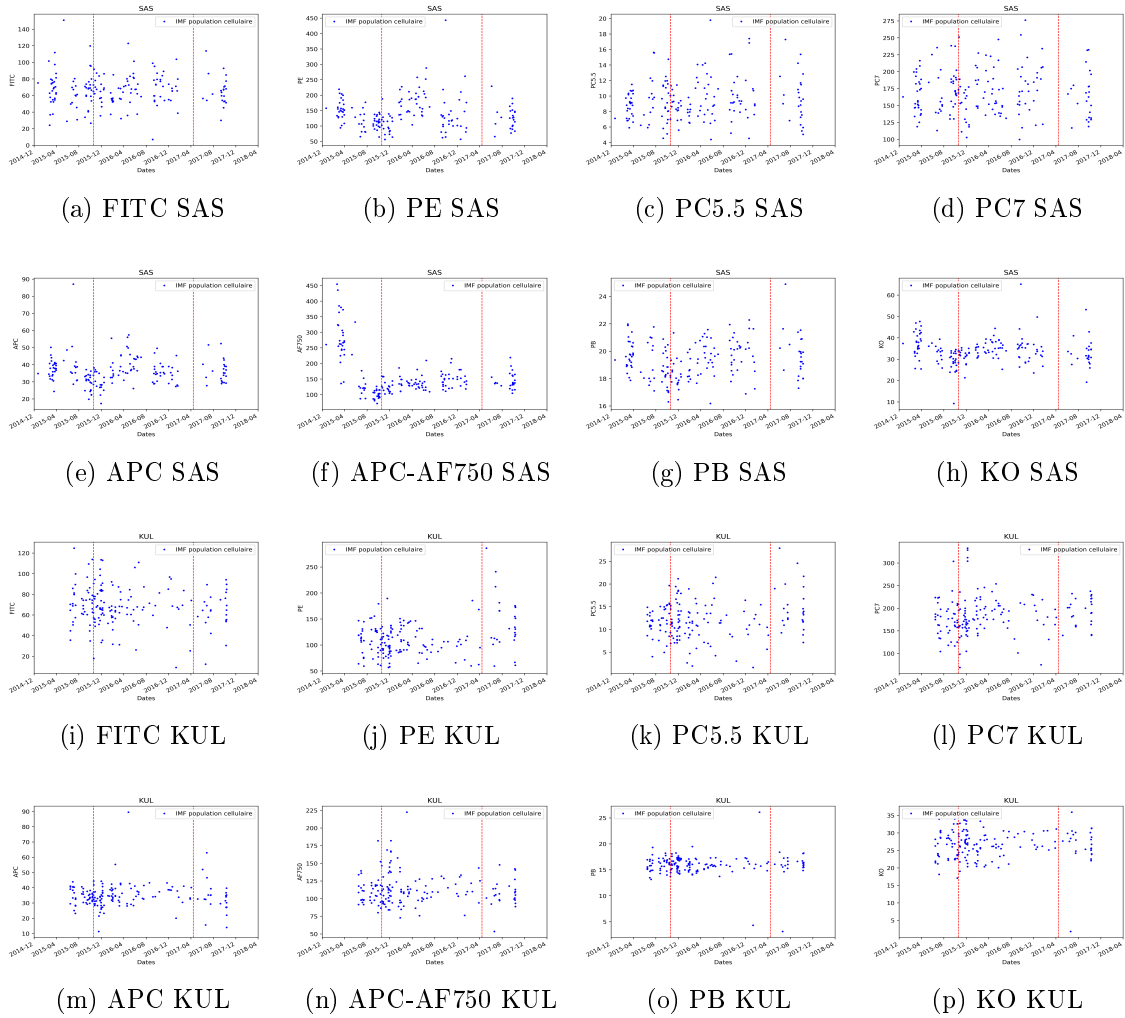


FIGURE A12: Évolution de l'IMF de chaque canal dans chaque centre (SAS,KUL). Les lignes rouges correspondent aux deux changement de lot lors des calibrations inter centre. L'application de la correction de l'effet lot a été appliqué.

A.3 Article en soumission

A standardization procedure for flow cytometry data harmonization in
prospective multicenter studies

Lucas Le Lann¹, PRECISESADS Flow Cytometry Study Group and Clinical Consortium, Pierre-
Emmanuel Jouve², Marta Alarcón-Riquelme³, Christophe Jamin^{1,4}, Jacques-Olivier Pers¹

¹Univ Brest, INSERM, UMR1227, Lymphocytes B et Autoimmunité, Brest, France,

²Altrabio SAS, Lyon, France

³GENYO, Centre for Genomics and Oncological Research Pfizer, University of
Granada, Andalusian Regional Government, PTS Granada, Granada, Spain

⁴Laboratoire d'Immunologie et Immunothérapie, CHU de Brest, Brest, France.

Corresponding Author: Christophe Jamin, UMR1227, CHRU Morvan, BP 824, F29609, Brest, France;
Tel : +332 98 22 33 84 ; Fax : +332 98 22 38 47 ; e-mail : christophe.jamin@univ-brest.fr

Grant support:

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under the grant agreement number 115565, resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007-2013) and EFPIA companies' in kind contribution. LLL was supported by the Agence Nationale de la Recherche under the "Investissement d'Avenir" program with the reference ANR-11-LABX-0016-001 and the Région Bretagne.

One of the challenges not yet fully resolved for clinical cytometry (1) is to compare, in prospective multicenter immunomonitoring trials, the frequencies and absolute values of various leukocyte populations, especially if the data are generated from different instruments and from different companies. The problem is even more difficult when it comes to comparing the expression of the markers (2) on freshly collected sample to avoid alteration of expression following freezing processes (3). If the fluorescence variations between different antibody lots remain a major problem, the divergent results can be seen as a result of inappropriate gating strategies between the centers (4) or because of imperfect fluorescence compensation (5, 6).

Recently, several studies have sought to move this issue forward without being able to meet all the criteria. Some have chosen to use equipment mainly from the same manufacturer (7, 8), others have preferred frozen samples (3, 9) or have resigned themselves to not doing prospective studies (10), while others have carried out inter-instrument comparability studies but on few samples (11).

As part of the IMI PRECISESADS study project, a multicenter analysis, using eleven different instruments (Navios, Gallios from Beckman Coulter, Canto II, Fortessa, Verse, Aria from BD Biosciences), prospectively on blood freshly collected from 3,301 individuals over a period of 5 years, was required. The objective was to compare the distribution (frequencies and absolute values) of leukocyte populations and the mean fluorescence intensities (MFIs) of the markers of the studied populations in order to establish a new classification of autoimmune diseases in relation to all the "omic" collection of data. We have preliminarily developed a standard operating procedure for the standardization of all instruments based on VersaComp Capture beads (Beckman Coulter) for inter-center harmonization and 8 peak beads (Beckman Coulter) for the intra-center daily QC, with the aim of obtaining inter-instrument coefficients of variation (CVs) of less than 5%. The relevance of this approach was demonstrated by an analysis of the same sample of control blood which highlighted a similar sensitivity of all devices in terms of frequencies of populations studied but also in terms of MFI markers studied (12).

The effective harmonization of the instruments through this procedure (Figure 1, step 1) allows the realization of large-scale multicentric phenotypic analyzes and assumes the stability of the 11 instruments throughout the duration of the study. Immunomonitoring was performed using 2 DuraClone antibody panels (Beckman Coulter) dedicated to the analysis of various populations of leukocytes and mononuclear cells in the peripheral blood of individuals (Supplementary Table 1).

Initially, a first R script was developed to normalize the results in intra-center based on the targets of the initial harmonization in order to correct variations potentially observed in the daily QC over the period of 5 years (Supplementary document 1). In order to validate this script, we voluntarily modified the PMTs of a Navios to provoke a measurement variation of the MFIs of the order of 10 to 15% on the 8 peak beads (Supplementary Figure 1a and Supplementary Table 2). As a result, these changes were made in the MFIs of the membrane markers studied during the acquisition of a blood test sample with DuraClone Panel 1 (Supplementary Figure 1b and Supplementary Table 3). The application of the R script on the LMD files of the 8 peak beads makes its MFI correspond to those obtained without modification of the PMTs, with coefficients of variation lower than 2.5% (Supplementary Figure 1c and Supplementary Table 2). Executed on the LMD files of the blood test sample stained with DuraClone Panel 1, the script also allowed MFIs to be found for membrane markers corresponding to those without modification, with coefficients of variation of less than 5% (Supplementary Figure 1d and Supplementary Table 3). Thus, this script makes it possible to standardize the results generated by the machine (Figure 1, step 2) and ensures the stability of its data throughout the duration of the project. This standardization script also makes it possible to ensure the comparability of results between the centers, as demonstrated by the analysis of the immunomonitoring of blood control stained with the DuraClone panel 1, carried out after the 5 years. The results show frequency variation coefficients measured between centers ranging from 2.3% for neutrophils to 17.7% for monocytes, and MFI variation coefficients ranging from 10.9% for the CD3 molecule to 30.9% for the CD15 molecule

(Supplementary Figure 2), confirming the stability of all instrument mirroring over the entire study period (12).

The inclusion of the study's 3,301 individuals was done along the way during the 5 years of the PRECISESADS project between January 2015 and December 2018. All FCS and LMD files from the acquisition of samples labeled with DuraClone Panel 1 and DuraClone Panel 2 were standardized with the R script. Then, the standardized file compensation matrices were verified and adjusted when necessary. This step was completed by the same single operator using Kaluza® software (Beckman Coulter) to minimize file preparation variations (Figure 1, step 3). Being extremely time-consuming, this step was followed by a second step dedicated to the analysis of the gating strategies in order to extract information from frequency and absolute number of the cell populations and from the MFI of the studied markers. To optimize this procedure, we opted for an automated analysis based on e-learning (13; 14). This analysis requires the development of an automaton through the immunomonitoring panel that learns the positioning of the gates as desired by the operator. The automaton operates in two steps. A first step that is specific to each of the instruments due to differences in event positioning in the Forward Scatter and Side Scatter (FS / SS) measures specific to each machine (Supplementary Figure 3a and 3b). The second step, commonly, is the extraction of information from gating strategies (Supplementary Figure 3c and 3d). In order to validate the e-learning of the panel 1 automaton and panel 2 automaton, we compared their results with a traditional "manual" analysis performed with the Kaluza® software, on 300 patients distributed throughout the 11 centers. The comparison of the results shows a very good correlation of the data of frequencies, absolute values and the MFIs (Supplementary Figures 4a-c), making it possible to validate the efficiency of the automata's learning (Figure 1, step 4). The flow cytometry data (frequencies, absolute values, and MFIs) of the 3,301 individuals included throughout the 5 years of the study are being extracted automatically by e-learning. A principal component analysis (PCA) of these data shows homogeneity of results between centers on frequencies (Figure 2a) but disparities between centers on MFIs (Figure 2b). What was not observed in analyses of a small sample (12 and

Supplementary Figure 2) has been unveiled with the data from the whole cohort. We, therefore, sought which parameter(s) could be responsible for these "center effects" despite the initial harmonization of the 11 cytometers and the intra-center normalization of each instrument to circumvent these dispersions.

First, the monoclonal antibodies chosen for immunomonitoring are used in DuraClone (Beckman Coulter) format. This dried antibody formulation provides stability over time, ease of storage, and avoids pipetting errors during sample preparation (15). Three different lots were used during the 5 years of the study. The MFI analysis of each panel 1 (Supplementary Figure 5a-c) and panel 2 (Supplementary Figure 5d and 5e) markers shows variations in batch changes. Therefore, we elaborated a new script under Python to correct the data between batches for each of the 11 instruments (Supplementary Document 2). Based on the alignment of the median values of the different batches with those of the first batch used for each center (Figure 1, step 5), the variations of the MFIs over time for each of the markers studied disappeared after application of the script (Supplementary Figure 5a-e). However, despite this correction, the PCA evaluation of the distribution of MFIs between instruments showed that a "center effect" persisted (Figure 2c). Intrinsic variations in the optical bench of each instrument remained and lead to measurement disparities in MFIs that are dependent on the target of the antibodies used and the associated fluorochromes (8). Assuming that the inclusions of individuals are similar in all centers, the medians of the MFIs on all measures must be equivalent regardless of the center. We therefore developed an additional Python script (Supplementary Document 3) to correct the medians of the MFIs of each marker between all the instruments (Figure 1, step 6), using the same Navios cytometer as a reference for all other machines throughout the project (12). After applying this script, the median values of each marker became insignificantly different regardless of the instrument with which the measurements were made (Supplementary Figure 6a and 6b), while maintaining the individual variations specific to each sample. A new PCA showed homogeneity of MFIs for all the markers studied in panel 1 as well as in panel 2 (Figure 2d).

Finally, the implementation of the workflow described in Figure 1 made it possible to compare all the data of frequencies, absolute values and MFIs resulting from the data generated by the immunomonitoring of 3,301 individuals included over a period of 5 years by 11 different flow cytometers. Biostatistical analyzes including these flow cytometry data with the set of "omic" data collected as part of the PRECISESADS project are now possible. More broadly, with all the validity required to follow the elaborated standardization procedures, this new workflow opens the possibility of carrying out any large-scale, prospective multicentric flow cytometry analysis whatever the duration, number and type of instruments required for the realization of such projects.

REFERENCES

1. Pedreira CE, Costa ES, Lecrevisse Q, van Dongen JJ, Orfao A; EuroFlow Consortium. Overview of clinical flow cytometry data analysis: recent advances and future challenges. *Trends Biotechnol.* 2013 Jul;31(7):415-25.
2. Böttcher S, van der Velden VHJ, Villamor N, Ritgen M, Flores-Montero J, Murua Escobar H, Kalina T, Brüggemann M, Grigore G, Martin-Ayuso M, Lecrevisse Q, Pedreira CE, van Dongen JJM, Orfao A. Lot-to-lot stability of antibody reagents for flow cytometry. 2017 Mar 30. pii: S0022-1759(17)30075-3. doi: 10.1016/j.jim.2017.03.018.
3. Ivison S, Malek M, Garcia RV, Broady R, Halpin A, Richaud M, Brant RF, Wang SI, Goupil M, Guan Q, Ashton P, Warren J, Rajab A, Urschel S, Kumar D, Streitz M, Sawitzki B, Schlickeiser S, Bijl JJ, Wall DA, Delisle JS, West LJ, Brinkman RR, Levings MK. A standardized immune phenotyping and automated data analysis platform for multicenter biomarker studies. *JCI Insight.* 2018 Dec 6;3(23).
4. Finak G, Langweiler M, Jaimes M, Malek M, Taghiyar J, Korin Y, Raddassi K, Devine L, Obermoser G, Pekalski ML, Pontikos N, Diaz A, Heck S, Villanova F, Terrazzini N, Kern F, Qian Y, Stanton R, Wang K, Brandes A, Ramey J, Aghaeepour N, Mosmann T, Scheuermann RH, Reed E, Palucka K, Pascual V, Blomberg BB, Nestle F, Nussenblatt RB, Brinkman RR, Gottardo R, Maecker H, McCoy JP. Standardizing Flow Cytometry Immunophenotyping Analysis from the Human ImmunoPhenotyping Consortium. *Sci Rep.* 2016 Feb 10;6:20686.

5. Kalina T, Flores-Montero J, Lecrevisse Q, Pedreira CE, van der Velden VH, Novakova M, Mejstrikova E, Hrusak O, Böttcher S, Karsch D, Sędek L, Trinquand A, Boeckx N, Caetano J, Asnafi V, Lucio P, Lima M, Helena Santos A, Bonaccorso P, van der Sluijs-Gelling AJ, Langerak AW, Martin-Ayuso M, Szczepański T, van Dongen JJ, Orfao A. Quality assessment program for EuroFlow protocols: summary results of four-year (2010-2013) quality assurance rounds. *Cytometry A*. 2015 Feb;87(2):145-56.

6. Kalina T, Brdickova N, Glier H, Fernandez P, Bitter M, Flores-Montero J, van Dongen JJM, Orfao A. Frequent issues and lessons learned from EuroFlow QA. *J Immunol Methods*. 2018 Sep 17. pii: S0022-1759(17)30140-0. doi:10.1016/j.jim.2018.09.008.

7. Streitz M, Miloud T, Kapinsky M, Reed MR, Magari R, Geissler EK, Hutchinson JA, Vogt K, Schlickeiser S, Kverneland AH, Meisel C, Volk HD, Sawitski B. Standardization of whole blood immune phenotype monitoring for clinical trials: panels and methods from the ONE study. *Transplant Res*. 2013 Oct 25; 2(1):17.

8 Kalina T, Flores-Montero J, van der Velden VH, Martin-Ayuso M, Böttcher S, Ritgen M, Almeida J, Lhermitte L, Asnafi V, Mendonça A, de Tute R, Cullen M, Sedek L, Vidriales MB, Pérez JJ, te Marvelde JG, Mejstrikova E, Hrusak O, Szczepański T, van Dongen JJ, Orfao A; EuroFlow Consortium. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia*. 2012 Sep;26(9):1986-2010.

9. Veluchamy JP, Delso-Vallejo M, Kok N, Bohme F, Seggewiss-Bernhardt R, van der Vliet HJ, de Gruijl TD, Huppert V, Spanholtz J. Standardized and flexible eight colour flow cytometry panels harmonized between different laboratories to study human NK cell phenotype and function. *Sci Rep.* 2017 Mar 10;7:43873.
10. Solly F, Rigollet L, Baseggio L, Guy J, Borgeot J, Guérin E, Debliquis A, Drenou B, Campos L, Lacombe F, Béné MC. Comparable flow cytometry data can be obtained with two types of instruments, Canto II, and Navios. A GEIL study. *Cytometry A.* 2013 Dec;83(12):1066-72.
11. Glier H, Heijnen I, Hauwel M, Dirks J, Quarroz S, Lehmann T, Rovo A, Arn K, Matthes T, Hogan C, Keller P, Dudkiewicz E, Stüssi G, Fernandez P; Swiss Cytometry Society. Standardization of 8-color flow cytometry across different flow cytometer instruments: A feasibility study in clinical laboratories in Switzerland. *J Immunol Methods.* 2017 Jul 29. pii: S0022-1759(17)30205-3. doi: 10.1016/j.jim.2017.07.013.
12. Jamin C, Le Lann L, Alvarez-Errico D, Barbarroja N, Cantaert T, Ducreux J, Dufour AM, Gerl V, Kniesch K, Neves E, Trombetta E, Alarcón-Riquelme M, Marañón C, Pers JO. Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. *Autoimmun Rev.* 2016 Nov;15(11):1038-1045.

13. Brinza L, Djebali S, Tomkowiak M, Mafille J, Loiseau C, Jouve PE, de Bernard S, Buffat L, Lina B, Ottmann M, Rosa-Calatrava M, Schicklin S, Bonnefoy N, Lauvau G, Grau M, Wencker M, Arpin C, Walzer T, Leverrier Y, Marvel J. Immune signatures of protective spleen memory CD8 T cells. *Sci Rep.* 2016 Nov 24;6:37651.

14. Grau M, Valsesia S, Mafille J, Djebali S, Tomkowiak M, Mathieu AL, Laubreton D, de Bernard S, Jouve PE, Ventre E, Buffat L, Walzer T, Leverrier Y, Marvel J. Antigen-Induced but Not Innate Memory CD8 T Cells Express NKG2D and Are Recruited to the Lung Parenchyma upon Viral Infection. *J Immunol.* 2018 May 15;200(10):3635-3646.

15. Hedley BD, Keeney M, Popma J, Chin-Yee I. Novel lymphocyte screening tube using dried monoclonal antibody reagents. *Cytometry B Clin Cytom.* 2015 Nov-Dec;88(6):361-70.

CONSORTIUM

The members of the PRECISESADS Flow Cytometry Study Group are Christophe Jamin¹, Concepción Marañón², Lucas Le Lann¹, Quentin Simon¹, Bénédicte Rouvière¹, Nieves Varela², Brian Muchmore², Aleksandra Dufour³, Montserrat Alvarez³, Carlo Chizzolini³, Jonathan Cremer⁴, Ellen De Langhe⁵, Nuria Barbarroja⁶, Chary Lopez-Pedraza⁶, Velia Gerl⁷, Laleh Khodadadi⁷, Qingyu Cheng⁷, Anne Buttgereit⁸, Zuzanna Makowska⁸, Aurélie De Groof⁹, Julie Ducreux⁹, Elena Trombetta¹⁰, Tianlu Li¹¹, Damiana Alvarez-Errico¹¹, Torsten Witte¹², Katja Kniesch¹², Nancy Azevedo¹³, Esmeralda Neves¹³, Maria Hernandez-Fuentes¹⁴, Pierre-Emmanuel Jouve¹⁵ and Jacques-Olivier Pers¹

¹U1227, Université de Brest, Inserm, Labex IGO, CHU de Brest, Brest, France.

²GENYO, Centre for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, PTS GRANADA, Granada, Spain.

³Immunology & Allergy, University Hospital and School of Medicine, Geneva, Switzerland

⁴Laboratory of Clinical Immunology, Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium.

⁵Division of Rheumatology, University Hospitals Leuven and Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium.

⁶IMIBIC, Reina Sofía Hospital, University of Cordoba, Cordoba, Spain.

⁷Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany.

⁸Bayer AG, Berlin, Germany.

⁹Pôle de Pathologies Rhumatismales Inflammatoires et Systémiques, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium.

¹⁰Laboratorio di Analisi Chimico Cliniche e Microbiologia - Servizio di Citofluorimetria, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milano, Italy.

¹¹Chromatin and Disease Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain

¹²Klinik für Immunologie und Rheumatologie, Medical University Hannover, Hannover, Germany.

¹³Serviço de Imunologia EX-CICAP, Centro Hospitalar e Universitário do Porto, Porto, Portugal

¹⁴UCB, Slough, UK

¹⁵AltraBio SAS, Lyon, France

The members of Precisesads Clinical Consortium are Beretta L, AguilarQuesada R¹, Aguirre-Zamorano MA², Callejas Rubio JL³, Castro-Villegas MC², Cervera R⁴, Chizzolini C⁵, Collantes E², Cornec D⁶, De Langhe E⁷, Devauchelle-Pensec V⁶, Escudero-Contreras A², Espinosa G⁴, Fernández Roldán MC⁸, Gomes Anjos T⁹, Hiepe F¹⁰, Jiménez Moleón I³, Jousse-Joulin S, Lauwerys B¹¹, López-Berrio A¹, Lories R⁷, Marovac J¹², Meroni PL¹³, Miranda B¹, Navarro-Linares H¹, Ortega-Castro R², Ortego N², Pers JO¹⁴, Ramón Garrido E¹⁵, Raya E³, Ríos Fernández R³, Rodríguez-Pintó I⁴, and Saraux A⁶.

¹Andalusian Public Health System Biobank, PTS GRANADA, Granada, Spain.

²Reina Sofia Hospital, University of Cordoba, Cordoba, Spain.

³Hospital Universitario San Cecilio, Servicio Andaluz de Salud, Granada, Granada, Spain.

⁴Hospital Clinic I Provincia, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain.

⁵University Hospital and School of Medicine, Geneva, Switzerland.

⁶Centre Hospitalier Universitaire de Brest, Hopital de la Cavale Blanche, Brest, France.

⁷Katholieke Universiteit Leuven, Belgium.

⁸Hospital Universitario San Cecilio, Servicio Andaluz de Salud, Granada, Spain.

⁹GENYO, Centre for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, PTS GRANADA, Granada, Spain.

¹⁰Deutsches Rheuma-Forschungszentrum Berlin, Germany.

¹¹Université catholique de Louvain, Brussels, Belgium.

¹²UCB, Slough, UK.

¹³Università degli studi di Milano, Italy.

¹⁴U1227, Université de Brest, Inserm, Labex IGO, CHU de Brest, Brest, France.

¹⁵Hospital Regional de Málaga, Servicio Andaluz de Salud, Málaga, Spain.

FIGURE LEGENDS

Figure 1. Workflow of the standardization procedure for the harmonization of flow cytometry data in multicentric prospective studies.

Flow cytometers are firstly harmonized using VersaComp capture beads to achieve the same reference for all instruments (step1). For the acquisition of blood samples, 8 peak beads are used as daily quality control. A R script allows the normalization of the data for each instrument to the reference during the period of inclusions (step 2). The compensation of all flow cytometry files are verified and adjusted to minimize disparities in the data file preparation (step 3). Frequencies of the populations of interest and the mean fluorescence intensities of the cell surface markers are automatically collected from all flow cytometry files by automaton having learned the gating strategy through e-learning (step 4). The mean fluorescence intensities of the cell surface markers are corrected by a Python script to adjust the median values and eliminate antibody batches variations in each instrument (step 5). The mean fluorescence intensities are finally corrected by an additional Python script to correct the variations of the median values between the instruments (step 6).

Figure 2. Evolution of Principal Component Analysis during the workflow of the standardization procedure of flow cytometry data in multicentric prospective studies.

Peripheral blood of 3,301 individuals was labeled with the dry panel 1 and panel 2 antibody formulations and then analyzed by flow cytometry using 11 different, previously harmonized instruments. The data of each instrument was then standardized by the R script. The frequencies of the leukocyte populations (panel 1) and mononuclear cells (panel 2) **(a)** and the mean fluorescence intensities of the cell surface markers **(b)** were collected from all flow cytometry files by automaton having learned the gating strategy through e-learning and were analyzed by principal component analysis (PCA). The mean fluorescence intensities of the cell surface markers were corrected by a Python script to adjust the median values and eliminate antibody batches variations in each instrument, before being analyzed by PCA **(c)**. The mean fluorescence intensities were finally

corrected by an additional Python script to eliminate the variations of the median values between the instruments, before being analyzed by PCA (**d**).

Supplementary Table 1. Composition of the two panels of the dry formulation of antibodies

(DuraClone). Fluorochromes indicated in the line title are classical fluorochromes from Beckman Coulter. The DuraClone panel compositions (Beckman Coulter) indicate the target molecule and common aliases, with the antibody clones mentioned in the brackets. FITC: fluorescein isothiocyanate; PE: phycoerythrin; PC5.5: PE-cyanin 5.5; PC7: PE-cyanin 7; APC: allophycocyanin; APC-AF750: APC-alexa fluor 750; PB: pacific blue; KRO: krome orange.

Supplementary Table 2. Evaluation of the normalization using a script R with the 8 peak beads

fluorescence. The mean fluorescence intensity of the most intense peak from the 8 peak beads is measured before (Reference) and after (Shift) the modification of the PMT values of a Navios flow cytometer. The variation with the reference after the shift is indicated. The application of the R script on the LMD files after the shift of the PMTs restores the initial values of the mean fluorescence intensity with abolition of the variations. PMT: photomultiplying tube; FITC: fluorescein isothiocyanate; PE: phycoerythrin; PC5.5: PE-cyanin 5.5; PC7: PE-cyanin 7; APC: allophycocyanin; APC-AF750: APC-alexa fluor 750; PB: pacific blue; KRO: krome orange. Mean \pm SD of 3 experiments.

Supplementary Table 3. Evaluation of the normalization of the cell surface fluorescence using a

R script. The peripheral blood cells from a control individual were labeled with the dry panel 1 DuraClone formulation. The mean fluorescence intensity of the positive cells for each marker was measured before (Reference) and after (Shift) the modification of the PMT values of a Navios flow cytometer. The variation with the reference after the shift is indicated. The application of the R script on the LMD files after the shift of the PMTs restores the initial values of the mean fluorescence intensity with abolition of the variations. PMT: photomultiplying tube; FITC: fluorescein

isothiocyanate; PE: phycoerythrin; PC5.5: PE-cyanin 5.5; PC7: PE-cyanin 7; APC: allophycocyanin; APC-AF750: APC-alexa fluor 750; PB: pacific blue; KRO: krome orange. Mean±SD of 3 experiments.

Supplementary Document 1. R script for the data normalization of a single flow cytometer.

R script runs for the intra-center normalization of the daily QC 8 peak beads of a flow cytometer on its 8 peak beads reference obtained after the harmonization of all the instruments. The coefficients of normalization are then carried out to the data files from peripheral blood stained with panel 1 and panel 2.

Supplementary Document 2. Python script for the data correction of a single flow cytometer.

Python script runs for the intra-center correction of the mean fluorescence intensities of the markers analyzed from panel 1 and panel 2 stainings generated by a single flow cytometer to adjust the variations induced by the different batches of fluorochrome-associated antibodies during the duration of the study to the first batch used.

Supplementary Document 3. Python script for the data correction between all flow cytometers.

Python script runs for the inter-center correction of the mean fluorescence intensities of the markers analyzed from panel 1 and panel 2 stainings generated by all flow cytometers to adjust the variations of the median values of all individuals obtained in one center to the median values of a reference center.

Supplementary Figure 1. Validation of the R script for the data normalization of a single flow

cytometer. Fluorescence profile of the 8 peaks beads before (Fluorescence Reference) and after (Fluorescence with 15% deviation) modification of the PMTs of each of the 8 channels of a Navios flow cytometer **(a)**. The peripheral blood of a control was labeled with the dry panel 1 antibody formulation and analyzed with the Navios flow cytometer before and after modification of the PMTs

of each of the channels **(b)**. The LMD files of the 8 peaks beads **(c)** and the blood sample **(d)** were standardized by the R script (15% deviated fluorescence normalization) and the fluorescence profiles compared to their respective references. FITC: fluorescein isothiocyanate; PE: phycoerythrin; PC5.5: PE-cyanin 5.5; PC7: PE-cyanin 7; APC: allophycocyanin; APC-AF750: APC-alexa fluor 750; PB: pacific blue; KRO: krome orange.

Supplementary Figure 2. Checking the stability of the harmonization of the eleven flow

cytometers. In all centers, the same batch of control blood was labeled with the dry panel 1 antibody formulation and analyzed by each of the eleven flow cytometers. The frequencies of the population targeted by the markers and the mean fluorescence intensities of the surface markers were obtained by an analysis with the Kaluza® software. The dotted lines indicate the threshold of 5% of variation compared to the Navios-1 reference flow cytometer. FITC: fluorescein isothiocyanate; PE: phycoerythrin; PC5.5: PE-cyanin 5.5; PC7: PE-cyanin 7; APC: allophycocyanin; APC-AF750: APC-alexa fluor 750; PB: pacific blue; KRO: krome orange.

Supplementary Figure 3. Gating strategies of the automatons. For all instruments, the data from the flow cytometry files are analyzed with a similar strategy by one automaton for the panel 1 and another automaton for the panel 2 **(a)**, and then specifically for each instrument from the gate [S4] to account for the variability of FSC and SSC signals **(b)**. The desired cell populations are identified by gating strategies identical for all instruments for panel 1 **(c)** and panel 2 **(d)** stainings.

Supplementary Figure 4. Comparisons of results obtained by traditional analysis and

automated analysis. The peripheral blood of 300 individuals was labeled with the dry panel 1 and panel 2 antibody formulations and then analyzed by flow cytometry using 9 different, previously harmonized instruments. The data of each instrument was then standardized by an R script. The frequencies of the cell populations **(a)**, the absolute values **(b)** and the mean fluorescence intensities

of the cell surface markers (c) were obtained by a simple manual analysis using Kaluza® software, and by automaton having learned the gating strategy through e-learning and compared.

Supplementary Figure 5. Intra-center correction of the fluorescence intensities of all the markers according to the batches formulation used. The peripheral blood of 3,301 individuals was labeled with the dry panel 1 (a-c) and panel 2 (d, e) antibody formulations and analyzed by flow cytometry using 11 different, previously harmonized instruments. The data of each instrument was then standardized by an R script. The mean fluorescence intensities of the cell surface markers were collected from all flow cytometry files by automaton having learned the gating strategy through e-learning (Before) and then treated with a Python script to adjust the median values (After) for the correction of the variations due to the batches of antibodies in each instrument.

Supplementary Figure 6. Comparisons between instruments of the mean fluorescence intensities after inter-center correction of the results. The peripheral blood of 3,301 individuals was labeled with the dry panel 1 (a) and panel 2 (b) antibody formulations, and analyzed by flow cytometry using 11 different, previously harmonized instruments. The data of each instrument was then standardized by an R script. The mean fluorescence intensities of the cell surface markers from leukocyte populations (a) and from mononuclear cells (b) were collected by an automaton having learned the gating strategy through e-learning, then corrected by a Python script to eliminate antibody batches variations in each instrument and finally corrected by an additional Python script to correct the fluorescence medians between instruments. The results collected by the 11 instruments were compared with each other before and after the additional Python script. *p<0.05; **p<0.01; ***p<0.001; ns: not significant.

A.4 Article Harmonisation

Autoimmunity Reviews 15 (2016) 1038–1045



Contents lists available at ScienceDirect

Autoimmunity Reviews

journal homepage: www.elsevier.com/locate/autrev



Review

Multi-center harmonization of flow cytometers in the context of the European “PRECISESADS” project



Christophe Jamin ^{a,*}, Lucas Le Lann ^a, Damiana Alvarez-Errico ^b, Nuria Barbarroja ^c, Tineke Cantaert ^d, Julie Ducreux ^e, Aleksandra Maria Dufour ^{f,g}, Velia Gerl ^h, Katja Kniesch ⁱ, Esmeralda Neves ^j, Elena Trombetta ^k, Marta Alarcón-Riquelme ^{l,m}, Concepción Marañón ^l, Jacques-Olivier Pers ^a

^a INSERM ERI29, EA2216, Université de Brest, Labex IGO, CHRU Morvan, Brest, France

^b Chromatin and Disease Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat Barcelona, Spain

^c IMIBIC/Reina Sofia Hospital/University of Cordoba, Córdoba, Spain

^d Laboratory of Tissue Homeostasis and Disease, Faculty of Medicine, KU Leuven, Leuven, Belgium

^e Pôle de Pathologies Rhumatismales, Institut de Recherche Expérimentale et Clinique, Université Catholique de Louvain, Brussels, Belgium

^f Department of Immunology and Allergy, University Hospital and School of Medicine, Geneva, Switzerland

^g Department of Pathology and Immunology, School of Medicine, Geneva, Switzerland

^h Department of Rheumatology and Clinical Immunology, Charité Hospital, Berlin, Germany

ⁱ Klinik für Immunologie und Rheumatologie, Medizinische Hochschule Hannover, Hannover, Germany

^j Serviço de Imunologia EX-CICAP, Centro Hospitalar do Porto, Porto, Portugal

^k Laboratorio di Analisi Chimico Cliniche e Microbiologia—Servizio di Citofluorimetria, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Italy

^l GENYO, Centre for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, PTS GRANADA, Granada, Spain

^m Institute for Environmental Medicine, Karolinska Institutet Stockholm, Sweden

ARTICLE INFO

Article history:

Received 10 July 2016

Accepted 12 July 2016

Available online 1 August 2016

Keywords:

Multi-parametric flow cytometry

Multi-center harmonization

Standard operating procedure

PRECISESADS

ABSTRACT

The innovative medicine initiative project called PRECISESADS will study 2,500 individuals affected by systemic autoimmune diseases (SADs) and controls. Among extensive OMICS approaches, multi-parameter flow cytometry analyses will be performed in eleven different centers. Therefore, the integration of all data in common bioinformatical and biostatistical investigations requires a fine mirroring of all instruments. We describe here the procedure elaborated to achieve this prerequisite. One flow cytometer chosen as reference instrument fixed the mean fluorescence intensities (MFIs) of 8 different fluorochrome-conjugated antibodies (Abs) using VersaComp Ab capture beads. The ten other centers adjusted their own PMT voltages to reach the same MFIs. Subsequently, all centers acquired Rainbow 8-peak beads data on a daily basis to follow the stability of their instrument overtime. One blood sample has been dispatched and concomitantly stained in all centers. Comparison of leukocytes frequencies and cell surface marker MFIs demonstrated the close sensitivity of all flow cytometers, allowing a multicenter analysis. The effective multi-center harmonization enables the constitution of a workable wide flow cytometry database for the identification of specific molecular signatures in individuals with SADs.

© 2016 Elsevier B.V. All rights reserved.

Contents

1. Introduction	1039
2. Multiparametric flow cytometry analyses	1039
3. First step: setting PMT voltage values	1040
4. Second step: monitoring instrument stabilization	1041
5. Verification of the harmonization	1041
6. Conclusion	1042
Take-home messages	1043

Abbreviations: APC, Allophycocyanin; APC-AF750, Allophycocyanin-Alexa Fluor 750; CV, Coefficient of variation; FITC, Fluorescein Isothiocyanate; KO, Krome orange; MFI, Mean fluorescence intensity; PB, Pacific blue; PC5.5, Phycoerythrin-cyanin 5.5; PC7, Phycoerythrin-cyanin 7; PE, Phycoerythrin; PMT, Photomultiplier tube; SAD, Systemic autoimmune disease.

* Corresponding author at: Laboratoire d'Immunothérapies et Pathologies Lymphocytaires B, Université de Brest, CHRU Morvan, BP824, F29609, Brest, France.

E-mail address: christophe.jamin@univ-brest.fr (C. Jamin).

<http://dx.doi.org/10.1016/j.autrev.2016.07.034>

1568-9972/© 2016 Elsevier B.V. All rights reserved.

Grant support	1043
Acknowledgments	1044
References	1044

1. Introduction

Inflammatory autoimmune diseases affect 1–3% of the population. There is numerous evidence suggesting that many of these conditions such as rheumatoid arthritis, systemic lupus erythematosus or Sjögren's syndrome might be incorrectly classified [1–4]. Furthermore, these diseases are hard to diagnose, and prevention of severe and fatal outcomes are difficult to reach. Although new immunotherapies are being developed with promising effects, it remains unclear which patients will benefit from which treatment [5–13]. Moreover, separate disease classification means that patients with clinically identified autoimmune diseases cannot benefit from treatment used for other diseases although common molecular basis can be found [14–19]. It becomes clear that a new classification of these diseases would be useful. Based on molecular signatures, the reclassification of patients will allow more personalized treatments to offer more effective and adapted medical care.

The PRECISESADS project aims to gather patients with systemic autoimmune diseases (SADs) into clusters taking in account OMICS information collected from peripheral blood cells, sera and urines of 2.500 people. In this context, all individuals will benefit from flow cytometry analyses that will be performed in eleven different European centers using different instruments from different manufacturers. Therefore, a fine inter-cytometer harmonization and intra-cytometer calibration before starting the inclusion of individuals is a prerequisite for the integration of the data into bioinformatical and biostatistical analyses of all OMICS information.

2. Multiparametric flow cytometry analyses

Multiparameter flow cytometry is becoming the method of choice to determine extensive phenotypic identity of peripheral blood leukocytes [20] but also of cells found in the bone marrow [21], in secondary lymphoid organs [22] or in ectopic lymphoid organs [23]. This procedure tends to be used also as a diagnostic tool [24] in broad

pathological situations such as leukemia [25], myelodysplastic syndrome [26], infections [27], or autoimmune diseases [28–30].

This complexified procedure needs growing expertise and the requirement of standardization approaches and harmonization methods appear henceforth all the more inevitable [31–34]. Between the different OMICS studies of the PRECISESADS project, 9 panels of 7 or 8 cell surface markers have been settled to perform a multiparametric flow cytometry analysis. This immunophenotypic characterization of the 2.500 individual samples will be carried out in eleven centers equipped with different instruments. To ensure comparison of the results, i.e., frequencies and absolute numbers of cell population and mean fluorescence intensities (MFIs) of cell surface markers avoiding center effect or disparities of the data due to mis-standardization, a specific harmonization procedure has been developed. To this end, the 9 panels of multicolor stainings have been designed based on the compatibility of all flow cytometers regarding their optical configurations. These multiparametric analyses will supply information related to the distribution of the leukocyte populations and the different lymphocyte subsets, either according to their maturation stage or according to their activation status (Table 1). Among the critical points reducing the reproducibility during flow cytometry studies are the instability of the fluorochromes conjugated to the antibodies (Abs) mainly of the tandem fluorochromes, and the pipetting errors of the reagents that may lead to changes in staining levels [20,33]. To bypass these problems, we used Duraclone tubes (Beckman Coulter) specifically designed and optimized for the PRECISESADS study. These tubes correspond to ready-to-use unitized, dry format Ab cocktails. They eliminate errors due to manual Ab preparations, they improve the stability compared to liquid reagents, avoiding tandem breakdown, and are room temperature stable, thus excluding the need to manage varying expiration date and revalidations among single color liquid Abs [35].

In the eleven sites responsible for the flow cytometry acquisition, a Navios flow cytometer (Beckman Coulter) was used in three centers, a Gallios (Beckman Coulter) in one center, a FACS Canto II (BD

Table 1
The PRECISESADS antibody panels, specificities and clones.

	FITC	PE	PC5.5	PC7	APC	APC-AF-750	PB	KO
Panel 1	CD16 (3G8)	CD15 (80H5)	CD56 (N901)	CD14 (RMO52)	CD19 (J4.119)	CD3 (UCHT1)	CD4 (13B8.2)	CD8 (B9.11)
Panel 2	CD1c (L161)	Lineage (various)	CD141 (M80)	CD11c (BU15)	CD123 (SSDCLY4107D2)	DRAQ7	HLA-DR (IMMU357)	
Panel 3	CD57 (NC1)	CD45RA (ALB11)	CD27 (1A4CD27)	CD62L (DREG56)	CD38 (LS198)	CD3 (UCHT1)	CD4 (13B8.2)	CD8 (B9.11)
Panel 4	TCR PAN γ/δ (IMMU510)	TCR PAN α/β (IP26A)	CD25 (B1.49.9)	CD127 (R34.34)	CD69 (TP1.55.3)	CD3 (UCHT1)	CD4 (13B8.2)	CD8 (B9.11)
Panel 5	IgD (IA6-2)	CD267 (1 A1)	CD5 (BL1a)	CD27 (1A4CD27)	CD19 (J4.119)	CD24 (ALB9)	CD38 (LS198)	
Panel 6	CD43 (DFT1)	CD69 (TP1.55.3)	CD5 (BL1a)	CD27 (1A4CD27)	CD11b (BEAR 1)	CD20 (B9E9)	CD3 (UCHT1)	
Panel 7	CD64 (22)	CD32 (2E1)	CD18 (7E4)	CD14 (RMO52)	CD19 (J4.119)	CD11b (BEAR 1)	CD16 (3G8)	
Panel 8	CD35 (J3D3)	CD59 (P282E)		CD14 (RMO52)	CD19 (J4.119)	CD3 (UCHT1)	CD21 (BL13)	CD16 (3G8)
Panel 9	CD46 (J4.48)	CD55 (JS11KSC2.3)		CD14 (RMO52)	CD19 (J4.119)	CD3 (UCHT1)		CD16 (3G8)

FITC, fluorescein isothiocyanate; PE, phycoerythrin; PC5.5, Phycoerythrin-Cyanin 5.5; PC7, Phycoerythrin-Cyanin 7; APC, Allophycocyanin; APC-AF750, Allophycocyanin-Alexa Fluor 750; PB, Pacific Blue; KO, Krome Orange.

Biosciences) in four centers and a FACS Aria III, a FACS Verse and a LSR Fortessa (BD Biosciences) in one center each. All instruments are equipped with three lasers emitting at 405/407, 488, and 633/635 nm and with optical filter configuration permitting the detection of FITC, PE, PC5.5, PC7, APC, APC-AF750, PB, and KO fluorochromes. VersaComp Ab Capture Bead kit (Beckman Coulter) is used for the photomultiplier tube (PMT) adjustments and the determination of the target MFI values applied for the multicenter harmonization procedure. Eight-peak Rainbow bead calibration particles (Spherotech) are utilized over the 5-year duration of the study for the daily checks as monocenter verification of the instrument stability. The same lots of VersaComp capture beads (#4,131,003 K) and of 8-peak Rainbow beads (#AF01) were ordered in all centers. If a new lot of beads needs to be ordered, the procedure described below will be repeated and the targets adjusted according to the PMT voltages fixed with the primary lot.

3. First step: setting PMT voltage values

To warrant reproducible measurements over the duration of the project in one instrument at the same center and to assure comparable measurements between all instruments at distinct sites throughout the project, specific operating procedures have been established. One flow cytometer (a Navios) has been designated as a reference instrument with which the 9 combinations of multicolor immunophenotypic panels have been validated. The PMT voltages

Table 2

Target mean fluorescence intensity (MFI) values for each fluorescence channel obtained with the reference instrument that must be reached in the other flow cytometers to fix their PMT voltages.

Fluorochrome channel	Single-color Duraclone	MFI values	
		Beckman Coulter instruments	BD Biosciences instruments
FITC	CD16	16.20	4147.20
PE	CD15	4.00	1024.00
PC5.5	CD56	31.05	7948.80
PC7	CD14	57.16	14,632.96
APC	CD19	42.47	10,872.32
APC-AF750	CD3	88.88	27,753.28
PB	CD4	21.95	5619.20
KO	CD8	12.23	3130.88

Abbreviations: FITC, fluorescein isothiocyanate; PE, phycoerythrin; PC5.5, Phycoerythrin-Cyanin 5.5; PC7, Phycoerythrin-Cyanin 7; APC, Allophycocyanin; APC-AF750, Allophycocyanin-Alexa Fluor 750; PB, Pacific Blue; KO, Krome Orange; MFI, mean fluorescence intensity; PMT, photomultiplier tube.

on this reference flow cytometer have been placed so that the 9 panels used the same PMT settings. These optimal voltages are used to determine the target fluorescence intensities of the VersaComp Ab capture beads stained with single-color Duraclone tubes (Fig. 1A and Table 2). Subsequently, using the same lot of

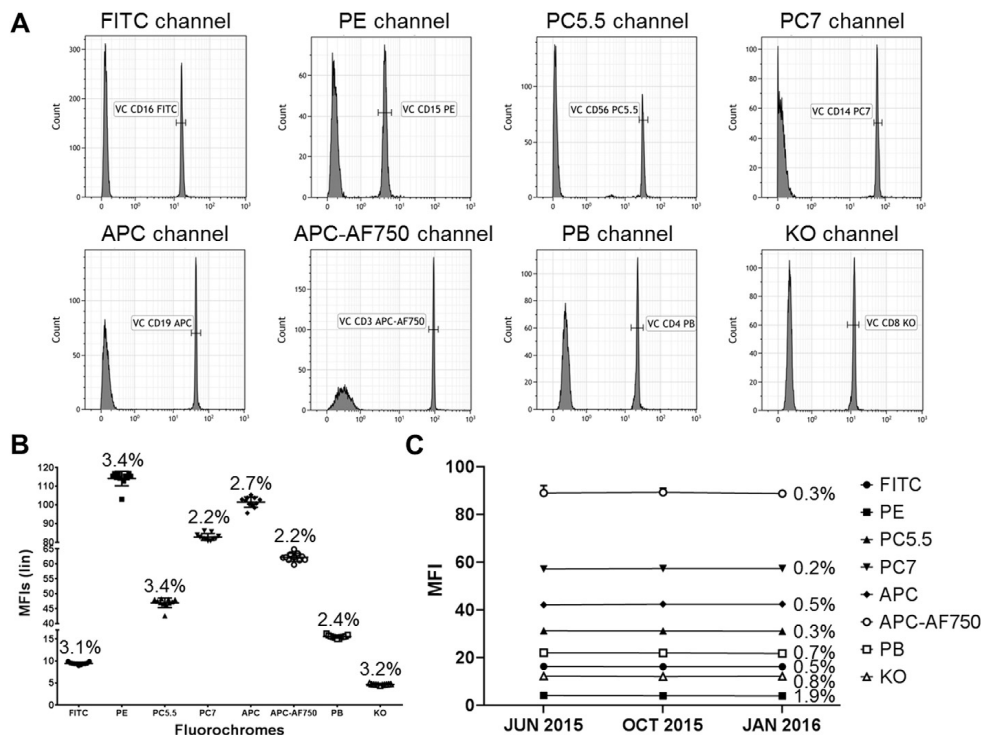


Fig. 1. Target fluorescence intensities of the VersaComp capture beads. (A) The beads are stained with single-color Duraclone tubes and analyzed on the Navios reference instrument (Center 1). (B) All centers have fixed the PMT values of their individual flow cytometer to get the same targeted fluorescence intensities. Coefficients of variation (CV) are indicated. (C) Long-term evaluation of the harmonization stability between all flow cytometers. CVs are indicated.

Table 3
PMT values of each instrument to obtain the same target mean fluorescence intensities for each fluorochrome.

		FITC	PE	PC5.5	PC7	APC	APC-AF-750	PB	KO
Center 1	Navios	421	467	484	562	524	631	414	382
Center 2	Navios	396	400	499	596	507	588	400	387
Center 3	Navios	425	435	460	505	530	550	400	415
Center 4	Gallios	480	440	479	562	576	561	380	408
Center 5	FACS Canto II	362	379	483	498	576	591	374	443
Center 6	FACS Canto II	375	457	459	541	536	647	364	392
Center 7	FACS Canto II	523	464	564	631	582	714	444	507
Center 8	FACS Canto II	394	435	550	576	645	640	416	375
Center 9	FACS Aria III	451	397	479	547	533	687	443	474
Center 10	FACS Verse	366	346	388	470	454	511	422	339
Center 11	LSR Fortessa	433	433	457	570	642	648	405	467

FITC, fluorescein isothiocyanate; PE, Phycoerythrin; PC5.5, Phycoerythrin–Cyanin 5.5; PC7, Phycoerythrin–Cyanin 7; APC, Allophycocyanin; APC-AF750, Allophycocyanin–Alexa Fluor 750; PB, Pacific Blue; KO, Krome Orange.

VersaComp beads stained with the same lot of single-color Duraclone tubes corresponding to panel 1 Abs, PMT voltages of all instruments have been set to obtain the same target fluorescence intensities (Table 3). Previous international standardization adopted acceptance criteria for deviations of up to 15% from the target MFI values. These stringent criteria allow the determination of intra- and inter-sample differences even for small MFI changes in individual markers [20]. In the PRECISEADS study, for the target peaks to be accepted, it has been decided that PMT voltages for each fluorochrome should be adjusted so that less than 5% CV of MFI was obtained with the reference instrument. Thus, PMT voltage settings lead to reach nearly identical MFI values for each single-color Duraclone tube in all instruments with CV lower than 3.4% (Fig. 1B).

To evaluate the fluctuation of the instruments overtime and to avoid deviation, the standard operating procedure using the VersaComp Ab capture beads with the same single-color Duraclone tubes is repeated every 3 months. If required, PMT voltages are adjusted to maintain identical MFI values of the target fluorescence intensities. In that case, the PMT values are set in the 9 multiparameter

panels serving as new intra-instrument reference assessments for the inclusion of future individual samples. The little variations of each fluorochrome-associated PMT detector in all instruments are shown Fig. 2. This procedure maintains identical MFI values demonstrating the stability of the harmonization with stringent CVs lower than 2.0% over long-term evaluation (Fig. 1C).

4. Second step: monitoring instrument stabilization

In addition to the inter-instrument harmonization, a supplementary standard operating procedure has been adopted prior to the inclusion of any patient in each center. This daily quality control is performed with the 8-peak Rainbow beads, using the data obtained after the PMT voltage setting as internal reference. Thereby, in all instruments, specific target MFI on low and high fluorescence peaks for each fluorochrome channel are recorded and fixed as “low” and “high” gates (Fig. 3). Again, for each instrument to pass the check, the deviation of the MFI values within the two gates of every fluorochrome must be <5% compared to the internal reference. In cases where instrument performance failed, while cleaning, de-gassing flow cell and laser delay are verified, PMT values are modified to adjust the position of the “low” and “high” fluorescence peaks in their gates. Revised PMTs are then reported in the 9 panels of the immunophenotyping. Compensation matrices will be impacted but adjusted later during the analysis procedure. On a daily basis, the MFI values of the “low” and “high” peaks of each fluorochrome channel for each individual instrument are stated and scrutinized with the Kaluza software (Beckman Coulter) to verify that the stringent deviation below 5% is maintained throughout the study (Fig. 4).

5. Verification of the harmonization

To demonstrate the efficient harmonization of the eleven flow cytometers, one blood sample has been dispatched in the eleven centers and stained with the Duraclone tubes corresponding to panel 1 and panel 3. Panel 1 is dedicated to the analysis of the different leukocyte populations (Table 4) following a lyse/no-washed procedure for the determination of the absolute cell number using Flow-Count fluorospheres (Beckman Coulter). Panel 3 has been designed for the frequency evaluation of the different T-cell subsets (Table 4) following a lyse/washed procedure (Table 5). Percentages

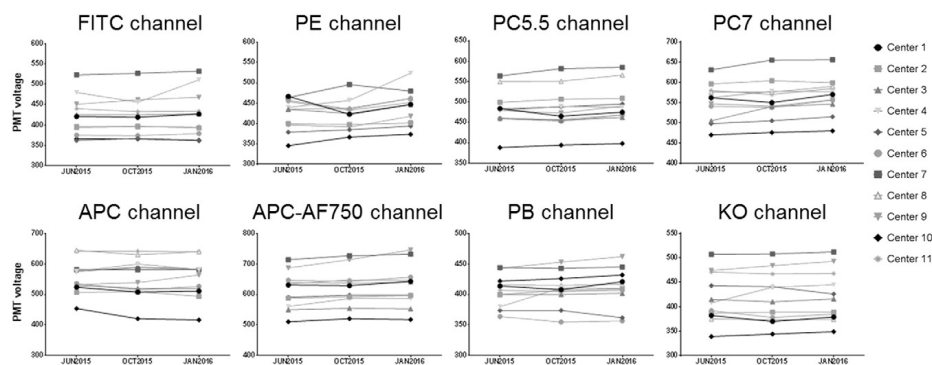


Fig. 2. Overtime variation of PMT voltages for each fluorescence channel in all individual flow cytometers. PMT, photomultiplier tube.

of leukocytes determined in the eleven centers are reproducible with small CVs varying from 1.8% for the neutrophil population to 22.8% for the monocyte population in panel 1 (Fig. 5A) and from 2.7% for the naive CD4⁺ T cell subset to 23.9% for the effector memory CD8⁺ T cell subset in panel 3 (Fig. 5C). Importantly, MFIs of cell surface markers assessed in the eleven centers are also comparable with CVs ranging from 12.4% for CD4 (PB fluorochrome) to 22.3% for CD15 (PE fluorochrome) but 49.4% for CD16 (FITC fluorochrome) in panel 1 (Fig. 5B). The higher CD16 variation is likely due to delayed blood transport (over 48 h) that might damage neutrophils and affect their CD16 expression [36]. Thus, no FITC disparity is found in panel 3 where the CV for CD57 (FITC fluorochrome) is 11.2% (Fig. 5D). Weak deviations are also observed for the other T cell markers. CVs varied from 7.9% for CD27 (PC5.5 fluorochrome) to 16.9% for CD38 (APC fluorochrome). Generally speaking, the lower the frequency or the MFI, the greater the variation, but it still remained in a range of values allowing the comparison of the data obtained with all individual instruments.

6. Conclusion

Differences between flow cytometers such as laser power output, sharpness of the optical filter's edges or hardware-associated variables may account for deviations [37,38]. It is therefore of the utmost importance for a multicenter study such as the PRECISESADS project to elaborate a harmonization procedure aimed at minimizing these differences. Our approach using Duraclone tubes provide stable reagents and our two-step stringent calibration of the eleven instruments generate reliable and comparable results. Thus, the PRECISESADS method for the inter-calibration of all flow cytometers yielded high-reproducible data with nearly identical frequencies of leukocyte populations and cell subsets, as well as very close MFIs of cell surface markers when a blood sample is analyzed in the different centers. Flow cytometry analyses of the 2.500 people who will participate in the project can now be safely gathered together to constitute a unique common flow cytometry database. Whichever the instrument used for data

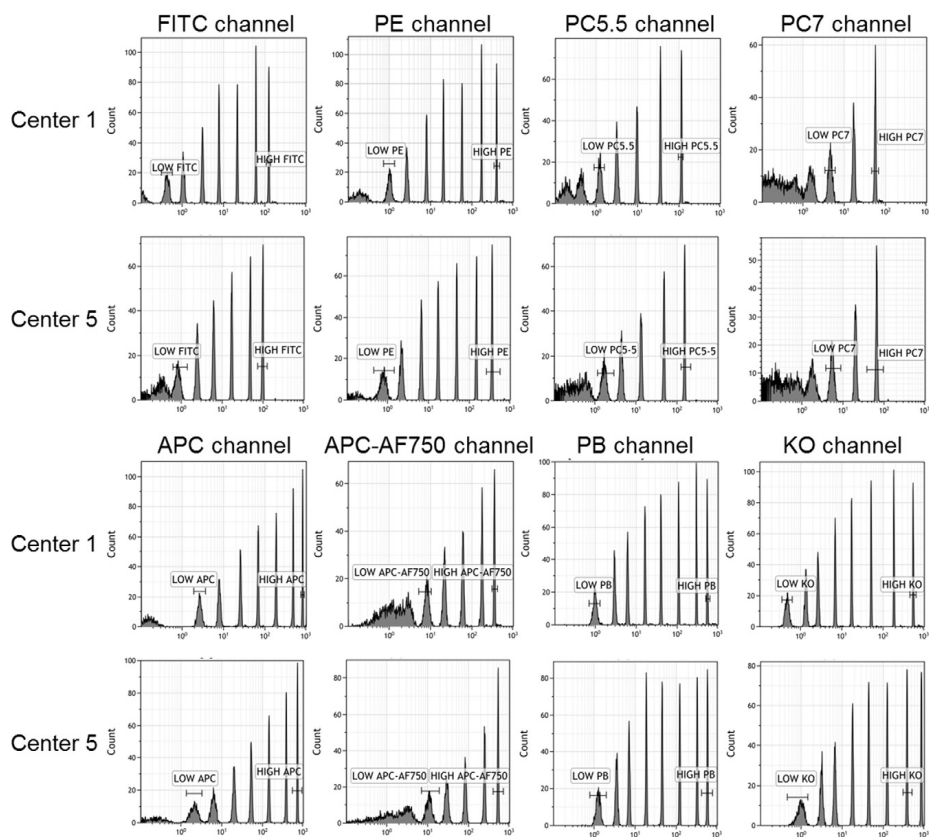


Fig. 3. Profile of 8-peak Rainbow beads. Low and high target fluorescence intensities in each channel are shown. Representative example of a Navios instrument (Center 1) and a FACScanto II (Center 5) instrument.

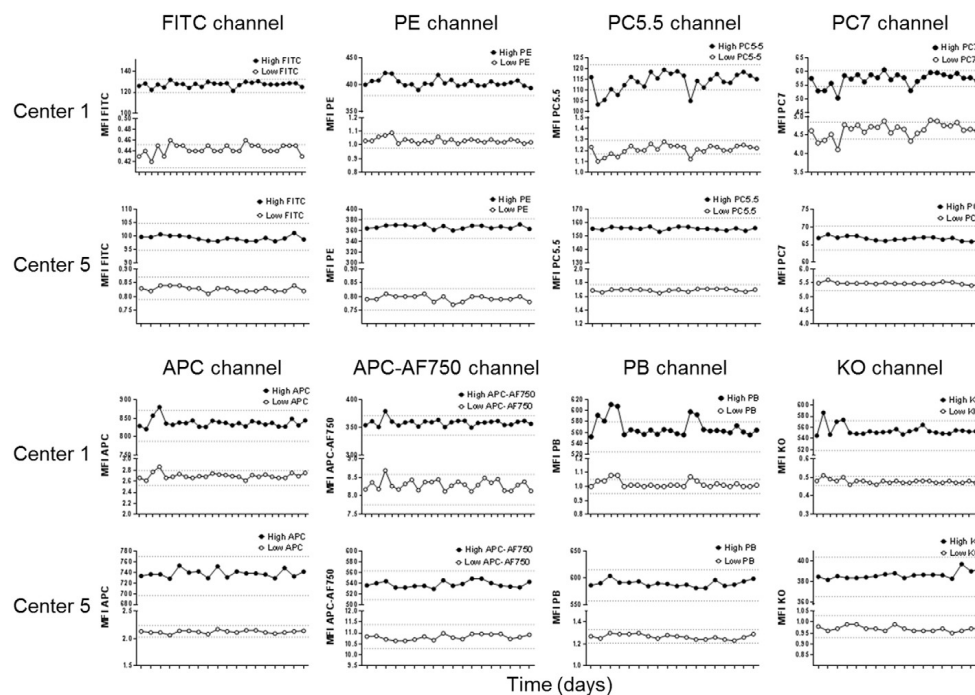


Fig. 4. Overtime low and high target fluorescence intensities of the 8-peak Rainbow beads analyzed after fixation of PMT voltages. Representative example of a Navios instrument (Center 1) and a FACS Canto II instrument (Center 5). Acceptable $\pm 5\%$ range for each channel is illustrated by dotted lines. PMT, photomultiplier tube.

acquisition, this pool of immunophenotypic information is a prerequisite to perform deconvolution studies involving also the different OMICS data. Thereafter, cluster identification of individuals sharing similar molecular signatures for their diseases can be safely investigated and the reclassification of SADs confidently considered.

Take-home messages

- A multi-parametric flow cytometry study performed in different centers requires a fine harmonization of all instruments.
- An effective multi-center calibration of flow cytometers can be achieved following a specific and rigorous procedure.
- This preliminary critical mirroring was indispensable to the inclusion of the flow cytometry data into large bioinformatical analyses of OMICS information for the PRECISESADS project.

Grant support

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115565, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. LLL is supported by Agence Nationale de la Recherche under the "Investissement d'Avenir" program with the reference ANR-11-LABX-0016-001 and the Région Bretagne. CM was funded by the

Instituto de Salud Carlos III (PI10/0552.PI13/0522) partly supported by European FEDER funds. The authors declare that no conflict of interest exists.

Table 4

Identification of the cell population and gating strategy for the analyses with panel 1 and panel 3 Duraclone tubes.

Populations identified in Panel 1	Gating strategy
CD3 + T cells	CD3 + CD19 –
CD4 + T cells	CD3 + CD4 + CD8 –
CD8 + T cells	CD3 + CD4 – CD8 +
B cells	CD3 – CD19 +
NK cells	CD3 – CD56 +
Monocytes	CD4low CD14 +
Neutrophils	CD15 – CD16 +

Populations identified in Panel 3	Gating strategy
Naïve CD4 + T cells	CD3 + CD4 + CD8 – CD45RA + CD62L + CD27 +
Effector CD4 + T cells	CD3 + CD4 + CD8 – CD45RA + CD62L – CD27 –
Central memory CD4 + T cells	CD3 + CD4 + CD8 – CD45RA – CD62L + CD27 +
Effector memory CD4 + T cells	CD3 + CD4 + CD8 – CD45RA – CD62L – CD27 –
Naïve CD8 + T cells	CD3 + CD4 – CD8 + CD45RA + CD62L + CD27 +
Effector CD8 + T cells	CD3 + CD4 – CD8 + CD45RA + CD62L – CD27 –
Central memory CD8 + T cells	CD3 + CD4 – CD8 + CD45RA – CD62L + CD27 +
Effector memory CD8 + T cells	CD3 + CD4 – CD8 + CD45RA – CD62L – CD27 –
Activated T cells	CD3 + CD4 – CD8 + CD38 +
Cytotoxic T cells	CD3 + CD4 – CD8 + CD57 +

Table 5

Detailed PRECISESADS standard operating procedures for sample staining with the Duraclone tubes.

Staining with panel 1 and panel 2 using a no-washed procedure	Staining with panel 3 to panel 9 using a washed procedure.
<ol style="list-style-type: none"> Prepare the IOTest 3 lysing 1×/fixative solution (Beckman Coulter): <ol style="list-style-type: none"> Dilute the IOTest 3 Lysing Solution 10× in distilled water. Add the IOTest 3 Fixative Solution according to the manufacturer's instructions. Mix well. Drop by reverse pipetting the required volume of blood samples (either 50 µL or 100 µL) as indicated in the Standard Operating Procedure into the corresponding Duraclone tubes: <ol style="list-style-type: none"> Mix for 10 s to rehydrate the antibodies. Incubate the tubes for 20 min at room temperature in the dark. Add 2 mL of the IOTest 3 lysing 1×/fixative solution at room temperature. <ol style="list-style-type: none"> Mix for 2 s. Incubate the tubes for 20 min at room temperature in the dark. Add the FlowCount fluorospheres (Beckman Coulter). <ol style="list-style-type: none"> Mix the vial for 5 s and shake it 5 times by inversion. Add by reverse pipetting 50 µL or 100 µL of fluorospheres (same volume as blood with the same pipet). Mix for 2 s. Tubes are ready for acquisition. <ol style="list-style-type: none"> Keep at 4 °C in the dark until acquisition. Acquire the cells within 2 h after preparation. 	<ol style="list-style-type: none"> Prepare the IOTest 3 lysing 1×/fixative solution (Beckman Coulter): <ol style="list-style-type: none"> Dilute the IOTest 3 Lysing Solution 10× in distilled water. Add the IOTest 3 Fixative Solution according to the manufacturer's instructions. Mix well. Drop by reverse pipetting the required volume of blood samples (either 50 µL or 100 µL) as indicated in the Standard Operating Procedure into the corresponding Duraclone tubes. <ol style="list-style-type: none"> Mix for 10 s to rehydrate the antibodies. Incubate the tubes for 20 min at room temperature in the dark. Add 2 mL of the IOTest 3 lysing 1×/fixative solution at room temperature. <ol style="list-style-type: none"> Mix for 2 s. Incubate the tubes for 20 min at room temperature in the dark. Centrifuge at 400 g for 5 min at room temperature. Discard the supernatant by returning the tube, and keep the tube upside down. Carefully absorb the liquid on paper towels to eliminate all the droplets. Add 4 mL of PBS, put the cap, and mix gently by reversing the tubes 4 times. Centrifuge at 400 g for 5 min at room temperature. Discard the supernatant by returning the tube, and keep the tube upside down. Carefully absorb the liquid on paper towels to eliminate all the droplets. Dispense 500 µL of PBS. Mix for 2 s. Tubes are ready for acquisition. <ol style="list-style-type: none"> Keep at 4 °C in the dark until acquisition. Acquire the cells within 2 h after preparation.

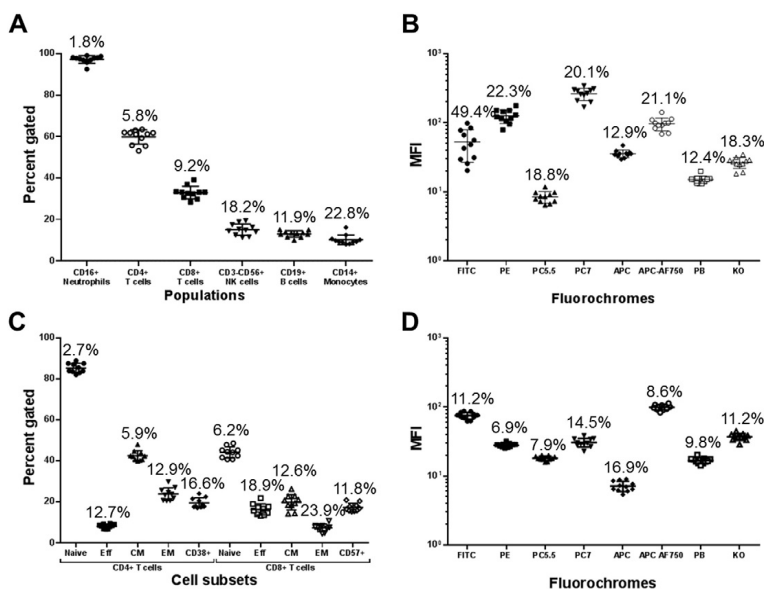


Fig. 5. Comparison of flow cytometry data acquired in the eleven centers. Percentages (A, C) and mean fluorescence intensities (MFI) of cell surface markers (B, D) identified in panel 1 (A, B) and panel 3 (C, D) analyses after staining the same blood sample. Coefficient of variations are indicated. Eff, effector; CM, central memory; EM, effector memory.

Acknowledgments

Thanks are due to Geneviève Michel and Simone Forest for typing the manuscript.

References

- Ramos-Casals M, Brito-Zeron P, Seror R, et al. Characterization of systemic disease in primary Sjogren's syndrome: EULAR-SS Task Force recommendations for articular, cutaneous, pulmonary and renal involvements. *Rheumatology (Oxford)* 2015;54:2230–8.

- [2] Abreu MM, Danowski A, Wahl DC, et al. The relevance of “non-criteria” clinical manifestations of antiphospholipid syndrome: 14th International Congress on Antiphospholipid Antibodies Technical Task Force Report on Antiphospholipid Syndrome Clinical Features. *Autoimmun Rev* 2015;14:401–14.
- [3] Maslinska M, Przygodzka M, Kwiatkowska B, Sikorska-Siudek K. Sjogren's syndrome: still not fully understood disease. *Rheumatol Int* 2015;35:233–41.
- [4] Calixto OJ, Franco JS, Anaya JM. Lupus mimickers. *Autoimmun Rev* 2014;13:865–72.
- [5] Relle M, Weimann-Menke J, Scorletti E, Cavagna L, Schwarting A. Genetics and novel aspects of therapies in systemic lupus erythematosus. *Autoimmun Rev* 2015;14:1005–18.
- [6] Liao J, Chang C, Wu H, Lu Q. Cell-based therapies for systemic lupus erythematosus. *Autoimmun Rev* 2015;14:43–8.
- [7] Kamal A, Khamashta M. The efficacy of novel B cell biologics as the future of SLE treatment: a review. *Autoimmun Rev* 2014;13:1094–101.
- [8] Doria A, Gatto M, Zen M, Iaccarino L, Punzi L. Optimizing outcome in SLE: treating-to-target and definition of treatment goals. *Autoimmun Rev* 2014;13:770–7.
- [9] Dimitroulas T, Nikas SN, Trontzas P, Kitis GD. Biologic therapies and systemic bone loss in rheumatoid arthritis. *Autoimmun Rev* 2013;12:958–66.
- [10] Thomas R. Dendritic cells and the promise of antigen-specific therapy in rheumatoid arthritis. *Arthritis Res Ther* 2013;15:204.
- [11] Zhang H, Chambers W, Sciascia S, Cuadrado MJ. Emerging therapies in systemic lupus erythematosus: from clinical trial to the real life. *Expert Rev Clin Pharmacol* 2016;1–14.
- [12] Her M, Kavanagh A. Alterations in immune function with biologic therapies for autoimmune disease. *J Allergy Clin Immunol* 2016;137:19–27.
- [13] Barnado A, Crofford LJ, Oates JC. At the bedside: neutrophil extracellular traps (NETs) as targets for biomarkers and therapies in autoimmune diseases. *J Leukoc Biol* 2016;99:265–78.
- [14] Cornec D, Jamin C, Pers JO. Sjogren's syndrome: where do we stand, and where shall we go? *J Autoimmun* 2014;45:109–14.
- [15] Kourilovitch M, Galarza-Maldonado C, Ortiz-Prado E. Diagnosis and classification of rheumatoid arthritis. *J Autoimmun* 2014;48:48–49:26–30.
- [16] Mosca M, Tani C, Vagnani S, Carli L, Bombardieri S. The diagnosis and classification of undifferentiated connective tissue diseases. *J Autoimmun* 2014;48–49:50–2.
- [17] Chung CP, Rohan P, Krishnaswami S, McPheeters ML. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. *Vaccine* 2013;31(Suppl. 10):K41–61.
- [18] Bossini-Castillo L, Lopez-Isac E, Martin J. Immunogenetics of systemic sclerosis: defining heritability, functional variants and shared-autoimmunity pathways. *J Autoimmun* 2015;64:53–65.
- [19] Distler O, Cozzio A. Systemic sclerosis and localized scleroderma-current concepts and novel targets for therapy. *Semin Immunopathol* 2016;38:87–95.
- [20] Kalina T, Flores-Montero J, van der Velden VH, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia* 2012;26:1986–2010.
- [21] Rajab A, Porwit A. Screening bone marrow samples for abnormal lymphoid populations and myelodysplasia-related features with one 10-color 14-antibody screening tube. *Cytometry B Clin Cytom* 2015;88:253–60.
- [22] Simon Q, Pers JO, Cornec D, et al. In-depth characterization of CD24CD38 transitional human B cells reveals different regulatory profiles. *J Allergy Clin Immunol* 2015;137:1577–84.
- [23] Pers JO, Devauchelle V, Daridon C, et al. BAFF-modulated repopulation of B lymphocytes in the blood and salivary glands of rituximab-treated patients with Sjogren's syndrome. *Arthritis Rheum* 2007;56:1464–77.
- [24] Wood B. 9-color and 10-color flow cytometry in the clinical laboratory. *Arch Pathol Lab Med* 2006;130:680–90.
- [25] van Dongen JJ, Orfao A, EuroFlow C. EuroFlow: resetting leukemia and lymphoma immunophenotyping. Basis for companion diagnostics and personalized medicine. *Leukemia* 2012;26:1899–907.
- [26] Bellos F, Kern W. Flow cytometry in the diagnosis of myelodysplastic syndromes and the value of myeloid nuclear differentiation antigen. *Cytometry Part B* 2014;00B:000.
- [27] Bene MC, Le Bris Y, Robillard N, et al. Flow cytometry in hematological nonmalignant disorders. *Int J Lab Hematol* 2016;38:5–16.
- [28] Binard A, Le Pottier L, Devauchelle-Pensec V, et al. Is the blood B-cell subset profile diagnostic for Sjogren syndrome? *Ann Rheum Dis* 2009;68:1447–52.
- [29] Laakso SM, Laurinli TT, Rossi LH, et al. Regulatory T cell defect in APECED patients is associated with loss of naive FOXP3(+) precursors and impaired activated population. *J Autoimmun* 2010;33:351–7.
- [30] Kaminski DA, Wei C, Rosenberg AF, Lee FE, Sanz I. Multiparameter flow cytometry and bioanalytics for B cell profiling in systemic lupus erythematosus. *Methods Mol Biol* 2012;900:109–34.
- [31] Maecker HT, McCoy JP, Consortium FHL, et al. A model for harmonizing flow cytometry in clinical trials. *Nat Immunol* 2010;11:975–8.
- [32] Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol* 2012;12:191–200.
- [33] Hasan M, Beitz B, Rouilly V, et al. Semi-automated and standardized cytometric procedures for multi-panel and multi-parametric whole blood immunophenotyping. *Clin Immunol* 2015;157:261–76.
- [34] Finak G, Langweiler M, Jaimes M, et al. Standardizing flow cytometry immunophenotyping analysis from the human immunophenotyping consortium. *Sci Rep* 2016;6:20686.
- [35] Hedley BD, Keeney M, Popma J, Chin-Yee I. Novel lymphocyte screening tube using dried monoclonal antibody reagents. *Cytometry B Clin Cytom* 2015;88:361–70.
- [36] Dransfield I, Buckle AM, Savill JS, et al. Neutrophil apoptosis is associated with a reduction in CD16 (Fc gamma RIII) expression. *J Immunol* 1994;153:1254–63.
- [37] Maecker HT, Trotter J. Flow cytometry controls, instrument setup, and the determination of positivity. *Cytometry A* 2006;69:1037–42.
- [38] Perfetto SP, Ambrozak D, Nguyen R, Chattopadhyay P, Roederer M. Quality assurance for polychromatic flow cytometry. *Nat Protoc* 2006;1:1522–30.

A.5 Article PRECISEADS

Integrative Analysis Reveals A Molecular Stratification of Systemic Autoimmune Diseases

Nature

Guillermo Barturen¹, Sepideh Babaei², Francesc Catalá-Moll³, Daniel Toro-Domínguez¹, Zuzanna Makowska², Pedro Carmona-Sanz⁴, Manuel Martínez-Bueno¹, Elena Carnero-Montoro¹, Martin Kerick⁵, Lucas Le Lann⁶, Christophe Jamin⁶, Javier Rodríguez-Ubrevá³, Antonio García-Gómez³, Jorge Kageyama², Anne Buttgereit², Sikander Hayat², The PRECISEADS Clinical Consortium**, The PRECISEADS Flow Cytometry Consortium**, Robert J. Benschop⁷, Chris Chamberlain^{8***}, Ernst R. Dow⁷, John Ioannou⁸, Laurence Laigle⁹, Jacqueline Marovac⁸, Fiona McDonald², Jerome Wojcik¹⁰, Esteban Ballestar³, Lorenzo Beretta¹¹, Maria Orietta Borghi¹², Johan Frostegård¹³, Javier Martín⁵, Joerg Mueller², Ralf Lesche², Yves Renaudineau⁶, Jacques-Olivier Pers⁶, and Marta E. Alarcón-Riquelme^{1,14,*}

(1) Department of Medical Genomics, Center for Genomics and Oncological Research (GENYO), Granada, Spain. (2) Pharmaceuticals Division, Bayer Pharma Aktiengesellschaft, Berlin, Germany. (3) Cancer Epigenetics and Biology Program, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. (4) Department of Bioinformatics, Center for Genomics and Oncological Research (GENYO), Granada, Spain. (5) Institute of Parasitology and Biomedicine López Neyra, Spanish National Research Council, Granada, Spain. (6) Université de Brest, INSERM, Labex IGO, CHU de Brest, Brest, France. (7) Eli Lilly and Company, Indianapolis, IN, United States of America. (8) UCB Pharma, Slough, United Kingdom. (9) Institut de Recherches Internationales Servier, Suresnes, France. (10) QuartzBIO, SA, Geneva, Switzerland. (11) Scleroderma Unit, Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca'Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy. (12) University of Milan, IRCCS Istituto Auxologico Italiano, Milan, Italy. (13) Unit for Chronic Inflammatory Diseases, Institute for Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

Key words: autoimmunity, systemic lupus erythematosus, rheumatoid arthritis, systemic sclerosis, primary Sjogren's syndrome, mixed connective tissue disease, primary antiphospholipid syndrome, undifferentiated connective tissue disease, SNF, stratification, classification criteria.

**the members of the Clinical and Flow Cytometry Consortia are listed in the supplementary materials.

***New affiliation: Head of Experimental Medicine, Merck Healthcare KGaA, Germany.

SUMMARY

Clinical heterogeneity, a hallmark of systemic autoimmune diseases (SADs) impedes early diagnosis and effective treatment options, issues that may be addressed if patients could be grouped into a molecular defined stratification. With the aim of reclassifying SADs indistinctively of the clinical diagnoses, unsupervised clustering of integrated whole blood transcriptome and methylome cross-sectional data of 955 patients with 7 SADs and 267 healthy controls was undertaken. Four clusters were identified, with one showing no specific molecular pattern, where healthy controls were assigned. The remaining three were 'acute phase inflammatory', 'T cell immunity', and 'interferon', each including all diagnoses, defined by genetic, clinical, serological and cellular features, each prone to reversal by different drugs. An inception cohort prospectively followed for 6 and 14 months, showed that most patients are either assigned always to the same cluster or move

from the healthy-like cluster to a single aberrant cluster resembling the relapsing-remitting dynamic of these diseases, suggesting that single aberrant molecular signatures may underpin each individual's patient relapses. These results have important implications on mechanisms and analysis of disease trajectories, therapy design across diseases, and mark a paradigm shift in our view of SADs and their taxonomy.

INTRODUCTION

The systemic autoimmune diseases (SADs) are entities diagnosed based on different clinical and laboratory criteria. The diseases are highly heterogeneous with varied progression of disease severity. The search for specific and sensitive biomarkers that may help predict the development of tissue damage have been an important focus of clinical research with little success. In general, the time from disease onset to diagnosis can be of many years, leading to damage accrual and worst prognosis. Some individuals, never fulfill the clinical criteria and remain undiagnosed for years or a lifetime. These are known as undifferentiated connective tissue disease (UCTD).

Previous genetic studies have shown that many of the SADs share susceptibility genes¹ and patients may also share, to a larger or lesser degree a gene expression signature known as the interferon signature, that is, the increased expression of interferon inducible genes^{2,3}. Clinical features can also be shared for instance a number of patients with systemic lupus erythematosus (SLE) may develop erosions and joint deformities in hands and feet, similar to those found in rheumatoid arthritis (RA)⁴. Further, the entity known as mixed connective tissue disease (MCTD) is controversial because patients may have clinical manifestations usually observed in SLE, RA or in systemic sclerosis (SSc)^{5,6}. While patients with SLE and RA may present with secondary Sjögren's syndrome (SjS), there are those having what is known as the primary entity (pSjS), without evidence of RA or SLE⁷. Similarly, SLE patients may have secondary antiphospholipid syndrome, but there are also patients with primary antiphospholipid syndrome (PAPS) who tend not to develop SLE (hence, primary) even after many years follow-up⁸.

The heterogeneous nature of these diseases is also observed at cellular and molecular levels. Not all patients with SLE have an interferon signature while the signature is more prevalent in childhood-onset disease. Some patients with SSc have a disease limited to the skin⁹, and not all patients who fulfill the diagnostic criteria for RA have anti-citrullinated peptide antibodies (~70%)¹⁰. Patients with SLE and pSjS share the presence of anti-SSA and anti-SSB antibodies, and in this regard, also share alleles of the HLA class II gene DRB1*0301¹¹.

The heterogeneity of the diseases also impedes the identification of new therapies, and worst, it has consequences in our selection of response endpoints and the overall results of clinical trials. One very clear example is the recent reported failure of Anifrolumab (anti-type I IFNR) in a phase III trial, while in a phase II it was successful when patients had been divided into IFN hi and IFN low^{12,13}. It is clear that as we advance, those patients having an IFN signature may benefit most from this new biological, but so far we know nothing of those patients who do not have an IFN signature and remain drug orphan.

In an unprecedented study in autoimmunity, high dimensional molecular data from whole blood shows here how the seven conditions (SLE, RA, SSc, SjS, MCTD, PAPS, and UCTD) distribute among clusters of specific molecular patterns, and how for each individual the patterns are stable over time.

RESULTS

Integrative molecular analysis redefines the distribution of systemic autoimmune diseases into functional clusters that are independent of the diagnoses

A total of 1521 patients had information from all three OMIC types. After individual platform quality control, 86% of the samples remained. Out of these, 6% showed discrepancies in terms of genetic information between platforms or gender compared with the clinical information, and were discarded (Supplementary Table 11). Following complete quality control, 955 patients and 267 healthy controls were available for analysis (diagnosis distribution is shown in Supplementary Table 12). For the inception cohort, of 173 patients with verified European ancestry, 148 (86%) had data from at least one of the three time points with information for all OMIC types. Three patients were lost after evaluation of discrepancies (detailed sample evaluation can be found in Supplementary Table 13). The numbers of patients with complete information for the three time points and the distribution of their diagnoses can be consulted in Supplementary Table 14.

With the aim of reclassifying the SADs indistinctively of the clinical diagnoses, genome wide transcriptome and methylome information was used in an unsupervised protocol design to perform an integrative molecular analysis (see supplementary materials). Cases from the cross-sectional cohort were divided into a discovery set (80%) and an independent (3 separate centers) validation set (20% of the samples). The discovery set cases were used in the next steps. The feature selection procedure resulted in 1821 genes and 4144 CpGs (Supplementary Figures 2A and 2B). Using similarity Network Fusion (SNF) algorithm¹⁴, two optimal configurations of 2 and 4 clusters with similar stability values were obtained. Next, using weighted correlation network analysis, WGCNA¹⁵ genes were grouped into 5 significant modules, while CpGs formed 3 modules, considering that module 0 is defined as a non-significant noisy module by WGCNA developers (Supplementary Figure 3). Interestingly, 32% of the CpGs were included in the CpG module 0, while only 10% of genes were grouped in gene module 0. The difference in the number of noisy features identified by WGCNA between both datasets is possibly due to the dynamic nature of DNA methylation, which might result in variation over time that could disrupt the stringent correlation needed for inclusion into the significant WGCNA modules.

The 2 and 4 cluster configurations totally agreed with patient assignments in clusters 1 and 3, but not for clusters 2 and 4, where patients presented different assignments in the 2 cluster configuration. Additionally, an independent signature defined cluster 4, and cluster 2 did not present a pattern as strong as the ones defining clusters 1 and 3. Thus, the patterns arising from the feature module analysis better fit with a 4 cluster configuration (see 2 cluster and 4 cluster annotations in Figure 1A). Otherwise, clusters 2 and 4 might be subdivided by the 2 cluster configuration, but such further subdivision had low stability (see Supplementary Figures 2D and 2E). These molecular patterns were clearly replicated in the validation set, showing that the molecular patterns, which defined the clusters, were stable in two independent datasets coming from different recruitment centers (Figure 1A).

The clusters of patients were defined by specific molecular signatures (Figure 1A). Clusters 1 and 3 had opposite patterns of gene modules 1 and 3 and CpG modules 1 and 2. Cluster 1 presented over-expression of gene module 3 and hypo-methylation of CpG module 1, while cluster 3 showed over-expression of gene module 1 and hypo-methylation of CpG module 2. Additionally, cluster 3 also showed slight over-expression of gene module 5. Cluster 4 had over-expression of gene module 2 and hypo-methylation of CpG module 3. Cluster 2 however, had no discernible pattern, showing a heterogeneous signature across the modules defined by WGCNA. Cluster 2 grouped around 40% of the patients.

Patients with all clinical diagnoses could be found in all 4 clusters (Figure 1B). Clusters 2 and 4 presented a somewhat biased distribution of some diagnoses (Figure 1C). Cluster 2 had an increased number of RA, SSc and less expected, PAPS. Cluster 4 was enriched in SLE, pSjS, and also MCTD (Figure 1B). Interestingly, UCTD patients were the only ones not significantly enriched

or depleted in any cluster (Figure 1C), even though the number of UCTD patients was higher than the number of MCTD and PAPS patients (Supplementary Table 12). For MCTD the results are quite remarkable. MCTD has strong clinical similarities with RA, SSc and SLE, but not pSjS, and its existence as a disease entity has been controversial. In fact, for some, MCTD patients are in a prior stage of disease and move toward any of the three disease entities¹⁶, for others, MCTD is a separate disease^{6,17}. That the great majority of patients fell in the interferon cluster is unexpected.

Except for systemic antibiotics (SABIO), most of the covariates associated with the transcriptome and the methylome principal components (Supplementary Figure 1), were unevenly distributed across the clinical diagnoses (Figure 1C). Importantly, most of these covariates did not show a dependency with the molecular clusters (Figure 1C), and the covariates that remained significantly associated to the clusters (mainly 3 treatments: antimalarials, AM; biologicals, BIO and steroids, STED) were related to the enrichment of the clinical diagnoses in those clusters (Figure 1C). For example, antimalarial (hydroxychloroquine) treated patients were enriched in cluster 4, and this association was driven by the enrichment of SLE, pSjS and MCTD (Supplementary Table 3). In short, treatments did not define the clusters.

Gene and CpG modules share and complement specific molecular functions that stratify the systemic autoimmune diseases

Gene and CpG modules that defined the clusters were highly concordant in their functionality. These showed a relationship, where overexpressed gene modules and hypomethylated CpG modules in the same clusters were enriched with the same functionalities as defined by Chaussabel et al,¹⁸ and Li et al,¹⁹ module enrichments (Figure 2A). Cluster 1 was defined by overexpression and hypomethylation of genes and CpGs included in inflammatory modules driven by monocytes and neutrophils (gene module 3 and CpG module 1). Cluster 3 was represented by T lymphocyte and natural killer (NK) cell functions (gene module 1 and CpG module 2), while cluster 4 was defined by interferon signature modules and dendritic cell functions (gene module 2 and CpG module 3). As previously shown, cluster 2, had no clear functional module signature (Figure 2A). Gene modules 4 and 5 further complemented the molecular information. Cell cycle and transcription upregulation (gene module 4) was associated with the interferon cluster (cluster 4) and B lymphocyte functions (gene module 5) were particularly observed in clusters 3 and 4 (Figure 2A).

Direct significant associations between CpGs and genes revealed different regulatory relationships between functionally-related modules. *Cis* significant associations linked CpG modules with their counterpart gene modules, while *trans* significant associations were evenly distributed without showing major relationships between homologous functional modules (Figure 2B). Nearly all *cis* associated CpGs (99%) in the interferon module (CpG module 3) were associated with the interferon gene module (gene module 2), 89% of the *cis* associations from the inflammatory gene module (gene module 3) were associated with CpGs in the inflammatory module (CpG module 1), and 75% of the *cis* associated CpGs from the T lymphocyte module (CpG module 2) were associated with its functionally-related gene module (gene module 1). However, the *cis* associations revealed a more complex relationship between DNA methylation and gene expression, where all CpG modules presented significant *cis* associations with more than one gene module. In particular, the inflammatory CpG module presented *cis* associations with every gene module (Figure 2B), which is not surprising as the CpGs englobed in the inflammatory module were enriched in the immune generic activation module defined by Li et al., enrichment shared with three of the gene modules (Figure 1A). However, and a major difference between clusters is that more than 80% of the CpGs included in the interferon module were associated in *cis* with genes included in the interferon transcriptome module, whereas most of the features in the rest of the modules had few (11 to 17%) significant *cis* associations (Figure 2C). Thus, the complex relationships shown between layers of information revealed that features selected from both

datasets shared common molecular functions but were not necessarily directly associated, providing a deeper view of the molecular state of the patients than that given by a single layer.

In order to understand how DNA methylation regulates the molecular patterns that clustered different clinical diagnoses, cell-type specific histone marks and transcription factor binding sites (TFBSs) regulated by the CpG modules were analyzed. In general, the histone mark cell-type enrichments were clearly related with the functional annotations of the modules. CpG module 1 (inflammation) was enriched in active histone marks (H3K4me3 and H3K27ac) in promoters and enhancers in monocytes, neutrophils, and dendritic cells. Surprisingly, H3K4me1, a histone mark for distal regulatory elements and enhancers, was significantly enriched in CpG module 1 for almost all cell-types included in the analysis. This difference between distal and proximal active histone marks in the CpG enrichments, suggests different regulation of the inflammatory response in different cell-types mediated by DNA methylation. On the other hand, the CpG module 2 (T cell immunity) showed a pattern opposite to that of CpG module 1, with all active marks (including H3K4me1) significantly depleted for the cell-types encompassing this module and with enrichments for lymphocytes and NK cells. Both CpG module 3 (interferon) and 0 showed uniform patterns across all cell-types, but with a different consequence. The interferon module was enriched in all active histone marks of all cell-types, and depleted for the H3K27me3 gene expression repression mark. This result shows that the activation of the interferon signature is a general response for all immune cell-types. On the other hand, the CpG module 0 was depleted in active and enriched in inactive histone marks.

TFBSs significantly enriched in each CpG module were also related with the functional annotation of the modules. The TFBS for NFIL3, ATF4, CHOP, PU.1, CEBP, and SPIB were specifically enriched for CpGs in the inflammatory module. Among these, ATF4 and CHOP have been described as regulators of the innate inflammatory response mediated via endoplasmic reticulum stress and the promotion of apoptotic processes²⁰, while all of them have been shown to be regulators of inflammatory processes in different autoimmune conditions²¹⁻²³. The TFBSs for EGRF2 and TCF3 were enriched in CpG module 2, which may play a role during the regulation of adaptive immunity²⁴⁻²⁶, while TFBS for IRF2, IRF3, and IRF8 (enriched in CpG module 3) are known regulators of interferon inducible genes and under the control of the type I IFN receptor²⁷.

Low dimensional layers of information show specific patterns in the molecular stratification

Having validated the clusters and established their functionality, their characterization from the clinical and serological point of view was next. For this, data from pre-selected autoantibodies, cytokines, small lipid moiety (natural) autoantibodies that were measured, and cell surface antigen markers detected in real-time with flow-cytometry collected from the recruited individuals at the time of sampling were used²⁸. The serology characterization showed four patterns of associations with clusters: anti-citrullinated peptide, anti-centromere B and IgM anti-phosphatidylcholine natural autoantibodies were slightly increased in clusters 2 and 3, with a strong depletion in the interferon cluster 4; most of the autoantibodies (anti-dsDNA, anti-Sm, anti-SSA, anti-SSB, anti-U1RNP), immunoglobulin markers (protein free light chains), and cytokine levels (IP-10, BAFF, MCP-2 and TNF- α) were enriched in the interferon cluster (cluster 4); the inflammatory cluster 1 was enriched in MMP-8 and C-reactive protein; while the enrichment of IL-1RA and BLC was shared by the inflammatory and interferon clusters (Figure 3A). In general, the enrichment of serological markers followed the molecular functions defined for the clusters. For example, cluster 4 was enriched in cytokines regulated by type I IFNs, such as IP-10 and BAFF, but also in TNF α , which, under some situations, was shown to induce interferon type I production mediating the innate immune response²⁹. On the other hand, cluster 1 patients were producing elevated levels of C-reactive protein and MMP-8, both markers of acute inflammatory processes^{30,31}.

Determining the cell population composition of the clusters revealed the enrichment in neutrophils in cluster 1, a certain enrichment of NK cells in cluster 2, T cells, B cells, NK cells and

NKT cells in cluster 3, and with the exception of some B cell enrichment, cluster 4 did not show any particular cell composition being instead expressed in all cell types as shown for the functionality of the features (above) (Figure 3B), all these following the previous molecular characterization of the clusters.

We then investigated if the clusters showed any differences in the genetic contribution of risk alleles for autoimmune diseases. For this, risk alleles discovered in different GWAS performed on all the four main SADs were included in the analysis (Supplementary Figure 5D and 5E). This analysis showed that the interferon cluster was particularly enriched in HLA (Human Leukocyte Antigen) class II risk alleles, but no other cluster showed significant enrichment. In order to validate this result, even if the number of individuals is far from the GWAS standards, four GWAS were performed comparing each cluster and healthy controls. Again, the only signal below genome-wide significance level (5×10^{-8}) came from alleles located in HLA class II genes in the interferon cluster analysis (Supplementary Figure 6). The weak HLA associations found in the other clusters were primarily HLA class I (HLA-B), and some suggestive signals were observed. Also suggestive was the association with IRF5 in cluster 4. Moreover, when GWAS were performed using the clinical diagnoses as covariates in the model, the associated signal remained (data not shown), verifying that the significant signal found was independent of the disease composition of the cluster. This result implies that the usual genetic associations observed for each of the systemic autoimmune diseases actually reflect the molecular and cellular mechanisms occurring in individuals whose molecular disease pathway is the type I interferon pathway. Some evidence exists to this, as recently it was observed that HLA class II was strongly associated in patients with SLE positive for the interferon signature³², and additionally in a plethora of specific autoantibodies as it was shown in cluster 4³³.

Most important is the clinical relevance of the clusters. The clinical information of the patients was summarized into principal components and differential associations were found for each molecular cluster (Figure 3C and 3D). The interferon cluster 4 was associated with some of the most extreme phenotypes in SADs such as kidney function abnormalities (including nephritis), thrombosis, nervous system involvement, and leukopenia, in addition to other minor comorbidities. Fibrosis complications in both skin and muscle skeletal organs were enriched in the inflammatory cluster 1 and in cluster 2, while these clusters were different in terms of kidney clinical features associations (enriched in the inflammatory cluster 1). Some comorbidities related with abnormal lipid metabolism and dyslipidemias were enriched in cluster 2. The T cell immune cluster 3, in general, presented less aggressive phenotypes than the other clusters, with also an enrichment of dyslipidemias, presence of abdominal pain, diarrhea, and constipation. Association with sicca syndrome was also found in this cluster.

Cluster 2 shows a healthy-like molecular pattern

Cluster 1, 3 and 4 were found to be defined and characterized by specific molecular patterns, however, cluster 2 presented an undefined pattern from the molecular point of view. To gain insight on the type of patients that were being grouped in this cluster, different analyses were performed. Healthy individuals were assigned to the molecular clusters by means of the trained molecular model, and 74% of them were grouped in cluster 2, compared to a 12%, 11% and 3% of assignments to 1, 3 and 4, respectively (Figure 4A). Additionally, differential expression analysis was performed between each cluster and healthy controls. The highest number of differentially expressed genes (DEGs) was found for the inflammatory cluster 1 (2898 DEGs) with a preponderance of down-regulated genes (2477 DEGs), followed by the interferon cluster 4 (820 DEGs) and the T cell immune cluster 3 (294 DEGs), while only 9 genes were found to be differentially expressed in cluster 2. The low amount of DEGs in cluster 2 compared with healthy controls suggested a close related molecular pattern, without differentially regulated pathways between both groups of individuals. We reasoned that this could be due to two non-mutually exclusive reasons. On the one hand, some of the most enriched diseases in this cluster, RA and SSc

could be undergoing processes in the target tissues (synovia and skin, respectively), and that blood was unable (with the exception of NK cells), to identify a defined molecular pattern. Potentially, cluster 2 might also be grouping patients undergoing remission or having low disease activity. To test this, disease activity was compared between clusters. Disease activity indexes were considered for each clinical diagnosis through different scores measuring clinical manifestations, traditionally associated with relapses, but are not performed across all diseases. For this analysis 138 SLEDAI-scored (out of 222 SLE patients), 79 ESSDAI-scored (out of 190 pSjS patients) and 17 DAS28-scored (out of 206 RA patients) were available. No activity score was available for SSc. In both, SLEDAI and ESSDAI higher and similar disease activity scores were shown in clusters 1, 3, and 4 as compared to cluster 2 (Figure 4C and 4D). Significant differences were found for cluster 1 and 4 when compared with cluster 2 in the SLEDAI analysis ($p=0.04$ and $p=0.02$, respectively, Wilcoxon rank sum test), while the other showed suggestive p -values. The results suggest that in some instances, low disease activity could lie behind the undifferentiated patterns of cluster 2, but possibly also the lack of a sufficiently large cell population from which the data would be extracted when using whole blood. Nevertheless, RA and SSc are known to have prolonged periods of low activity or remission, which could also explain the results.

Specific drugs are found that might revert pathogenic molecular patterns

In order to investigate if the clustering found could have therapeutic implications, drug reposition analysis was conducted for each of the aberrant clusters (1, 3, and 4). None of the drugs found with the lowest connectivity scores that theoretically might revert the abnormal gene expression patterns were of common use in any of the SADs (Figure 4E). Nevertheless, some of the drugs identified with the best scores were described to potentially be effective to treat autoimmune disorders. For instance, topical use of isoliquiritigenin in psoriasis³⁴, arvanil in multiple sclerosis³⁵, or a PKC β -inhibitor which was associated by GWAS analysis with inflammatory bowel disease and RA³⁶. Protein kinase C β has been also associated with SLE development in mouse models, thus an inhibitor might prevent developing the disease³⁷. While a few drugs did not differentiate clusters, most drugs did have cluster specificity (Figure 4E). One of the identified drugs was shared across all three clusters, indirubin, which is a traditional Chinese medicine that was clinically used in the treatment of autoimmune disorders. Recently, some studies have been published showing its potential beneficial effects in psoriasis³⁸ and immune thrombocytopenia³⁹, and some clinical trials were performed in psoriasis patients (clinicaltrials.gov). The inflammatory cluster 1 presented the highest number of significantly associated drugs, and it is also the cluster with the largest diversity of MoAs (Mechanisms of Action), without any common MoA for all the drugs (Supplementary figure 7B), probably explained by the general nature of the inflammatory processes. The T cell cluster 1 drugs presented a couple of common MoAs, including NFK β and EGFR pathway inhibitors among others (in agreement with the EGRF2 TFBSs enrichment shown above, Figure 2E), while the interferon cluster drugs showed common mechanisms related to the inhibition of apoptosis pathways.

Pathological molecular patterns are stable over time

It is possible that the clusters represent the disease state of the individual patients and that these could move, with time, to different clusters as disease progresses. Furthermore, although we have clearly adjusted for confounders, the long time of disease combined with years of treatment (12 years mean since diagnosis for the CS cohort) could lie behind the configuration of clusters that we observe. In order to define, and as a second validation, if the clusters could be observed by means of the same trained model in patients with newly diagnosed disease, the inception cohort was analyzed. The same clusters were observed, and each of the three sampling points was independently assigned to one molecular cluster.

In order to analyze whether the molecular assignments were stable over time, the assignment for each time point were summarized by patient. The Jaccard index for patients with complete molecular information at recruitment and 6 ($n=103$) or 14 months ($n=78$) samplings showed similar results (Figure 5A and 5B). In both comparisons, most of the patients remained in

the same cluster after recruitment (62-63%) and 34-33% of them moved from a pathological cluster to the healthy-like cluster or vice versa, while only 4% of patients were assigned to a different pathological cluster. The analysis of all three time points together (n=68) showed that only 4 patients (6%) jumped between different pathological clusters, while 33 patients (48%) remained in the same cluster over all three time points (Figure 5C). The remaining patients (46%) showed a behavior that resembled a relapse-remission dynamic of SADs from a molecular point of view: their pathological cluster was stable (never assigned to a different pathological cluster) but could be assigned in a given time point to the healthy-like cluster (Figure 5C). In brief, when changing cluster, patients would only move to and from cluster 2 to a single pathological cluster. This is similar to what we observed in our recent work on longitudinal stratification of SLE⁴⁰. Our results suggest that patients' cluster assignment is stable in time and may signify that each patient has a single mechanism of disease pathology out of 3 possible, when detectable in an integrated fashion through transcriptome and methylome in blood.

First steps towards systemic autoimmune diseases precision medicine

In summary, we observe primarily 3 pathological clusters that we may describe as an acute phase inflammatory cluster 1 with a pattern coming from neutrophils and monocytes, a T cell immunity cluster 3 enriched in lymphocytes, particularly T cells, NK cells, NKT cells but also B cells, and an interferon cluster 4, spread across all cell types but with a slight enrichment of NK cells. Cluster 2 had no particular molecular pattern where healthy controls were assigned when using the same trained model and indeed with very few differentially expressed genes as compared to controls.

This study shows for the very first time and in an unprecedented number of individuals with several layers of high and low molecular data that systemic autoimmune disease patients from 7 different diseases share molecular clusters defined by specific molecular patterns that are stable in time. The clusters have, each, specific clinical and serological characteristics, but also quite different regulatory profiles. The results obtained in this study are a first step towards laying the foundations of the personalized medicine in systemic autoimmune diseases.

METHODS

Samples, Data types and OMIC Platforms

Two cohorts of individuals with 7 different systemic autoimmune diseases (SADs) were recruited: a cross sectional (CS) cohort composed of 2003 patients and 617 healthy controls (HC) and an inception cohort of 215 patients followed and sampled at 0, 6 and 14 months. Inclusion and exclusion criteria are detailed in Supplementary Table 2. Demographic information and treatment can be consulted in Supplementary Tables 3-6. A detailed description of both cohorts can be found in the Supplementary Materials.

Blood and serum samples were obtained from all patients in order to gather several types of molecular information. High-dimensional genome wide genotype, transcriptome, DNA methylome (OMIC datasets) and proportions of relevant cell types (using flow cytometry custom marker panels) were analyzed from whole blood samples. Low dimensional information was obtained from serum samples, including selected serology information such as SADs related autoantibodies, cytokines, chemokines and inflammatory mediators. A detailed description of all the protocols and methods can be found in the supplementary materials.

Each OMIC dataset was quality controlled and controlled the integrity of the patient origin. Briefly, genetic information was obtained from RNA sequencing and from the single nucleotide polymorphisms (SNPs) included in the DNA methylation arrays. The concordance between all OMIC datasets was assessed, and discordant samples were discarded from further analysis. Detailed information on the sample integrity analysis can be consulted in supplementary materials.

Cluster Discovery and Validation Analysis

Design

The quality-controlled dataset of patients with SADs was split into a discovery set (~80% cases and healthy controls) and a validation set (~20% cases). The validation set was composed of all patients from three independent recruitment centers (France, UBO; Belgium, UCL and Germany, UKK). The discovery set was corrected for **batch and potential confounder effects** (gender, age, treatment effects and center of recruitment biases) using linear regression models but preserving the differences between clinical conditions using the *removeBatchEffect* from the *limma R package*⁴¹. The potential confounders were assessed using principal component association with the sample annotation. Annotations associated (linear regression p-value ≤ 0.05) with any of the transcriptome and/or methylome first ten principal components were considered confounders and included in the models (Supplementary Figure 1). Identification analysis of the confounders was performed using the *swamp R package*⁴². After correction, samples that were identified as **outliers** either for transcriptome or methylome were removed from the analysis. Outliers were assessed by multidimensional scaling and a fixed threshold based on mahalanobis distance cut-off of 8 for both transcriptome and methylome (*MASS R package*⁴³). In total, from the discovery cohort, 44 samples were discarded (28 due to transcriptome, and 16 samples for methylome). Outliers were not associated with any clinical feature or meta-information (data not shown).

Feature Selection

The **selection of features** on high-throughput datasets is aimed at selecting a subset of informative features to be used in the clustering analysis. Feature selection prevents uninformative features (that usually are the majority in high-throughput datasets) that could mask the signal of interest, and additionally the reduction of features makes the use of machine learning approaches in these large datasets feasible. In general, feature selection techniques are based on comparisons to obtain the features that maximize the differences between groups (supervised feature selection techniques).

When we began, preliminary analyses of the results to find the best way to perform feature selection and integrative clustering, we observed that the feature selection performed in all cancer studies published to date would not allow us to find any clusters or groups. In cancer reclassification studies, features with the highest variability across cancer types are chosen for reclassification purposes in an unsupervised way⁴⁴, but the natural variability of the immune-system makes this approach inappropriate for systemic autoimmune diseases because of the dynamics of the immune system and its dependence on natural variability⁴⁵. We took a completely new approach where only features with statistically increased variability in cases (being cases those individuals diagnosed with any kind of SAD included in the analysis) compared to healthy controls were chosen, allowing us to identify the clusters. Expression of genes and DNA methylation sites variability contributes to disease susceptibility^{46,47}, which implies that our feature selection might be able to select features which discriminate between differential molecular patterns, and at the same time remove high variable features not related with the pathologies included in the study. Additionally, we chose unsupervised feature selection techniques because supervised techniques can give features that discriminate between diagnoses, but important features that could subdivide those diagnoses into new groups would be less prone to be selected. Thus, a FDR Levene-test below 0.1 and a case-versus-control variability fold-change higher than 0.5 were used to select the informative features from transcriptome and methylome data from the cases. Importantly, confounding correction, outlier detection, and feature selection were only performed on the discovery cohort.

Clustering

An integrated clustering approach was chosen for molecular clustering analysis of the data. With this approach merging of all OMIC data together is the first aim followed by a single clustering. High-dimensional OMIC datasets were used for clustering analysis, while low-

dimensional molecular data was used for characterization of the molecular clusters obtained. Similarity Network Fusion (SNF)¹⁴ was selected to obtain the clusters of SAD patients. SNF is based on network theory allowing to find shared and complementary patterns across layers of information (different OMICs), as it deals well with the differential dimensionality of the layers⁴⁸. Moreover, the SNF algorithm can be used in a discovery / validation approach, as it allows the training of a mathematical model that can be validated with an independent dataset.

Transcriptome and methylome selected features were integrated using SNF and the clustering model was trained on the discovery set by means of a 10 times nested cross-validation approach (Supplementary Figure 2C). After background structure subtraction, two optimal configurations arised for 2 and 4 clusters of patients with similar stability values in the inner cross-validation (Supplementary Figure 2E) and validated in the outer cross-validation (Supplementary Figure 2D).

Training and testing of the clustering model were done as follows. Healthy controls were not used in the clustering model. Ten times nested cross-validation was run on the discovery set (without controls) in order to set up the optimal hyperparameters for the SNF algorithm. Inside each nested cross-validation, an inner 5-fold cross-validation was performed for hyperparameter model optimization with each training subset of the outer 10-fold cross-validation (Supplementary Figure 2C). Multiple combinations of different values of the recommended main SNF hyperparameters were then tested: 10 to 30 neighbors in the K-nearest neighbors (K) with a step of 5, 0.3 to 0.8 hyperparameter used in constructing similarity network with a step of 0.1 and 2 to 20 numbers of clusters. The hyperparameters were optimized through maximizing the proportion of cluster assignments that can be recovered from the test datasets using the trained SNF model, and a model trained using each dataset independently (considered as the ground truth). The average proportion of labels recovered for each combination of hyperparameters were compared with the results of the randomized matrixes (random background structure), the results can be seen in Supplementary Figure 2D. The clustering results obtained from the discovery set were validated using the validation set (Supplementary Figure 2C).

Functional Characterization of the Clusters and their Features

In order to assess the molecular signatures that drive the clusters and their relationships across layers of information, selected genes and CpGs were independently grouped into functional modules (groups of correlated features) using weighted correlation network analysis, WGCNA¹⁵. For genes and CpGs selected and included in the clusters, meta-features were identified using default parameters and “signed hybrid” network type¹⁵. Two immunological databases were used for meta-feature molecular characterization: Chaussabel et al.¹⁸ and Li et al.¹⁹ by means of the *tmod R package*⁴⁹. The enrichment of genes in those modules was tested and functional terms below a hyper-geometric enrichment FDR < 0.01 were selected. The genes related with CpGs were obtained from the Illumina probe annotation.

CpG - Gene Associations

The association between CpG DNA methylation and gene expression was assessed using the *Matrix eQTL R package* implementation⁵⁰. Associations were only tested on selected features. Briefly, methylation data was categorized in five bins (0.2 range each) and considered as *loci* in the analysis, while gene log10(TPM+1) were used as the quantitative feature (emQTL, expression-methylation Quantitative Trait Loci). Batch effects and other potential confounders were included in the linear regression model (see above). Significant associations considered were those with FDR < 0.01 and they were classified into *cis*-regulatory associations (distance between CpG and associated gene < 10kb) or *trans*-regulatory associations (≥10kb).

Epigenetic Characterization of CpG Modules

CpGs selected were characterized in terms of transcription factor binding site motifs and histone modification co-localization. **Transcription factor** enrichment analysis was performed using HOMER software v4.10⁵¹. In order to determine the relative enrichment of known motifs in

the methylation modules, known motif enrichment was tested in the DNA sequence contained in 500 bp windows around CpGs for each module. **Histone mark modification** enrichment analysis was performed using a subset of blood cells and histone modifications from the BLUEPRINT Consortium⁵². BLUEPRINT enriched regions for H3K4me3, H3K27ac, H3K4me1, H3K36me3, H3K27me3 and H3K9me3 histone modifications in monocytes (CD14+ CD16-), neutrophils (CD45+ CD66b+ CD16+), naïve CD4 T cells (CD3+ CD4+ CD45RA+), naïve CD8 T cells (CD3+ CD8+ CD62L+ CD45RA+), naïve B cells (CD19+ CD27- IgD+), dendritic cells, NK cells (CD3- CD56dim CD16+) and eosinophils (CD45+ CD66b+ CD16-) were downloaded from BLUEPRINT data analysis portal⁵³. Histone modification co-localization enrichment was tested for the CpGs in each module compared to all CpGs selected, the enrichment significance was calculated by means of Fisher's exact test and FDR was calculated for multiple test correction.

Healthy pattern characterization of SADs clusters

Healthy individuals were mapped to the clusters in order to define where in the model do the healthy molecular conditions fit. In this sense, the selected gene and CpG features were extracted from healthy individuals' OMIC datasets and tested with the pre-defined model. Additionally, differential expression analysis was performed between the clusters and the healthy individuals. Transcriptome expression dataset samples and genes were filtered and cleaned as described before. Read count expression values were normalized using voom algorithm from the *limma R package*⁴¹. Known batch and potential confounders previously defined were included in the linear model. *limma package* was used for differential expression analysis. An $FDR < 1e-5$ and $\log_2FC > |0.5|$ was the cut-off to consider transcripts as differentially expressed between healthy controls and SADs patients by cluster.

Cluster Characterization for Clinical and Low-dimensional Data

The distribution across clusters of clinical annotations (diagnosis and clinical symptoms) and low-dimensional data (auto-antibodies, serum cytokine protein levels, cell subsets, and genetic risk markers) were analyzed. Numerical and categorical low-dimensional dataset associations with clusters were assessed using linear and logistic regression models, respectively. Features with Analysis of Variance (ANOVA) p-values below 0.01 were considered unevenly distributed across clusters, and z-scored beta coefficients were used to assess enrichments and depletions. *R packages car*⁵⁴ and *multcomp*⁵⁵ were used in the analysis. Genetic risk loci and clinical features were reduced to Principal Components (PCs) prior to regression analysis, only PCs with variance explained higher than 5 times the standard deviation plus the average of 1000 randomizations of the original matrix were tested (First 64 and 11 PCs were selected for risk associated SNPs and clinical features respectively, see Supplementary Figure 4A and Supplementary Figure 4B). Clinical features with less than 5% informative value (non-negative and non-missing values) were filtered before the analysis (Supplementary Figure 4C) and the remaining missing values were imputed using *missMDA R package*⁵⁶. Briefly, significant PCs were assessed by means of *estim_ncpPCA* function, and they were used for the imputation of the missing values using *imputePCA* function and regularized method.

Genome-wide association analyses

Genome-wide association analyses were performed for each cluster. Genotype data filtering and association analysis were conducted using *PLINK v1.9*⁵⁷ applying the following criteria: minimum total call rate per sample of 90%, minimum call rate per marker of 98%, minor allele frequency threshold of 1%, Hardy-Weinberg Equilibrium p-value at a minimum of $1e-5$. Analysis of association for each cluster versus HC were performed by logistic regression under additive allelic model. Genetic stratification was not detected by genomic inflation factor (lambdaGC): lambdaGC cluster 1 versus HC = 1.025, lambdaGC cluster2 versus HC = 1.049, lambdaGC cluster3 vs HC = 1.038 and lambdaGC cluster4 versus HC = 1.037).

Drug repurposing analysis

Gene expression signatures from the most significant differentially expressed genes (top and bottom 150 DEGs), from the differential expression analysis against HC, for clusters 1, 3 and 4 were queried independently into the *CLUE database*⁵⁸. Thus, drugs that might potentially revert the aberrant molecular patterns of each cluster were assessed (drug repurposing analysis). The CLUE repurposing tool retrieves a similarity score (ranged from 100 to -100), where positive values represent related molecular patterns and negative values represent opposite molecular patterns between drug signatures and gene expression signatures. Drugs with a connectivity score below -90 were considered significant (Figure 4A and Supplementary Figure 7A).

In order to acquire insight into the relationship between the molecular effects of the significant drugs, and direct links to possibly shared mechanisms of action (MoA) recorded in the CLUE database, the CLUE touchstone tool was used to obtain ranked lists of the mechanisms of action (MoAs) related with the significant drugs. Touchstone tool retrieves a ranked list of related drugs, and the MoA for each related drug was obtained parsing drug-MoA links from the CLUE database. Significant drug-MoAs ranked lists were imputed into *fgsea R package* in order to obtain a score of association between each significant drug and all the MoAs related. Only MoAs with p -value ≤ 0.01 and gsea (gene set enrichment analysis) similarity score ≥ 0.99 with at least one other drug were represented in Supplementary Figure 7B.

Inception cohort analysis

Inception cohort time points were assigned individually to clusters by means of the SNF model trained with the cross-sectional cohort. The results were analyzed by pair of samplings and samples with complete information: patients with information for M000 and M006 samplings, $n=103$; M000 and M014 samplings, $n=78$ and complete sampling points, $n=68$. The classification of patients for complete samples was as follows: *stable patients*, patients with the same cluster assignment for all the time points; *relapse – remission patients*, patients with assignments only to an aberrant cluster (clusters 1, 3 or 4) and assignments to healthy-like individual cluster (cluster 2) at any time point; *unstable patients*, patients with assignments to different aberrant clusters across time.

Data availability

Data will be transferred to ELIXIR⁵⁹ and made available upon request and in agreement with the privacy rules of the European Union. Additionally, gene expression and DNA methylation datasets can be explored at <http://bioinfo.genyo.es/precisesadsdata/>.

REFERENCES

- 1 Li, Y. R. *et al.* Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nature medicine* **21**, 1018-1027, doi:10.1038/nm.3933 (2015).
- 2 Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2610-2615, doi:10.1073/pnas.0337679100 (2003).
- 3 Bennett, L. *et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *The Journal of experimental medicine* **197**, 711-723, doi:10.1084/jem.20021553 (2003).
- 4 Tani, C. *et al.* Rhupus syndrome: assessment of its prevalence and its clinical and instrumental characteristics in a prospective cohort of 103 SLE patients. *Autoimmunity reviews* **12**, 537-541, doi:10.1016/j.autrev.2012.09.004 (2013).
- 5 Alarcon-Segovia, D. & Cardiel, M. H. Comparison between 3 diagnostic criteria for mixed connective tissue disease. Study of 593 patients. *The Journal of rheumatology* **16**, 328-334 (1989).

- 6 Sharp, G. C., Irvin, W. S., Tan, E. M., Gould, R. G. & Holman, H. R. Mixed connective tissue disease--an apparently distinct rheumatic disease syndrome associated with a specific antibody to an extractable nuclear antigen (ENA). *The American journal of medicine* **52**, 148-159 (1972).
- 7 Skopouli, F. N., Drosos, A. A., Papaioannou, T. & Moutsopoulos, H. M. Preliminary diagnostic criteria for Sjogren's syndrome. *Scandinavian journal of rheumatology. Supplement* **61**, 22-25 (1986).
- 8 Wilson, W. A. *et al.* International consensus statement on preliminary classification criteria for definite antiphospholipid syndrome: report of an international workshop. *Arthritis and rheumatism* **42**, 1309-1311, doi:10.1002/1529-0131(199907)42:7<1309::AID-ANR1>3.0.CO;2-F (1999).
- 9 Peterson, L. S., Nelson, A. M. & Su, W. P. Classification of morphea (localized scleroderma). *Mayo Clinic proceedings* **70**, 1068-1076, doi:10.4065/70.11.1068 (1995).
- 10 Murphy, D., Matthey, D. & Hutchinson, D. Anti-citrullinated protein antibody positive rheumatoid arthritis is primarily determined by rheumatoid factor titre and the shared epitope rather than smoking per se. *PloS one* **12**, e0180655, doi:10.1371/journal.pone.0180655 (2017).
- 11 Teruel, M. & Alarcon-Riquelme, M. E. Genetics of systemic lupus erythematosus and Sjogren's syndrome: an update. *Current opinion in rheumatology* **28**, 506-514, doi:10.1097/BOR.0000000000000310 (2016).
- 12 Vukelic, M., Li, Y. & Kyttaris, V. C. Novel Treatments in Lupus. *Frontiers in immunology* **9**, 2658, doi:10.3389/fimmu.2018.02658 (2018).
- 13 Casey, K. A. *et al.* Type I interferon receptor blockade with anifrolumab corrects innate and adaptive immune perturbations of SLE. *Lupus science & medicine* **5**, e000286, doi:10.1136/lupus-2018-000286 (2018).
- 14 Wang, B. *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nature methods* **11**, 333-337, doi:10.1038/nmeth.2810 (2014).
- 15 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
- 16 Ciang, N. C., Pereira, N. & Isenberg, D. A. Mixed connective tissue disease-enigma variations? *Rheumatology* **56**, 326-333, doi:10.1093/rheumatology/kew265 (2017).
- 17 Alarcon-Segovia, D. Mixed connective tissue disease: some statements. *Clinical rheumatology* **1**, 81-83 (1982).
- 18 Chaussabel, D. *et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* **29**, 150-164, doi:10.1016/j.immuni.2008.05.012 (2008).
- 19 Li, S. *et al.* Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nature immunology* **15**, 195-204, doi:10.1038/ni.2789 (2014).
- 20 Zhang, K. & Kaufman, R. J. From endoplasmic-reticulum stress to the inflammatory response. *Nature* **454**, 455-462, doi:10.1038/nature07203 (2008).
- 21 Schlenner, S. *et al.* NFIL3 mutations alter immune homeostasis and sensitise for arthritis pathology. *Annals of the rheumatic diseases* **78**, 342-349, doi:10.1136/annrheumdis-2018-213764 (2019).
- 22 Sokhi, U. K. *et al.* Dissection and function of autoimmunity-associated TNFAIP3 (A20) gene enhancers in humanized mouse models. *Nature communications* **9**, 658, doi:10.1038/s41467-018-03081-7 (2018).
- 23 Smith, J. A. Regulation of Cytokine Production by the Unfolded Protein Response; Implications for Infection and Autoimmunity. *Frontiers in immunology* **9**, 422, doi:10.3389/fimmu.2018.00422 (2018).

- 24 Pollack, B. P. EGFR inhibitors, MHC expression and immune responses : Can EGFR inhibitors be used as immune response modifiers? *Oncoimmunology* **1**, 71-74, doi:10.4161/onci.1.1.18073 (2012).
- 25 Wohner, M. *et al.* Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. *The Journal of experimental medicine* **213**, 1201-1221, doi:10.1084/jem.20152002 (2016).
- 26 Best, J. A. *et al.* Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. *Nature immunology* **14**, 404-412, doi:10.1038/ni.2536 (2013).
- 27 Jefferies, C. A. Regulating IRFs in IFN Driven Disease. *Frontiers in immunology* **10**, 325, doi:10.3389/fimmu.2019.00325 (2019).
- 28 Jamin, C. *et al.* Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. *Autoimmunity reviews* **15**, 1038-1045, doi:10.1016/j.autrev.2016.07.034 (2016).
- 29 Yarilina, A. & Ivashkiv, L. B. Type I interferon: a new player in TNF signaling. *Current directions in autoimmunity* **11**, 94-104, doi:10.1159/000289199 (2010).
- 30 Sproston, N. R. & Ashworth, J. J. Role of C-Reactive Protein at Sites of Inflammation and Infection. *Frontiers in immunology* **9**, 754, doi:10.3389/fimmu.2018.00754 (2018).
- 31 Manicone, A. M. & McGuire, J. K. Matrix metalloproteinases as modulators of inflammation. *Seminars in cell & developmental biology* **19**, 34-41, doi:10.1016/j.semcdb.2007.07.003 (2008).
- 32 Miller, S. *et al.* Hypomethylation of STAT1 and HLA-DRB1 is associated with type-I interferon-dependent HLA-DRB1 expression in lupus CD8+ T cells. *Annals of the rheumatic diseases* **78**, 519-528, doi:10.1136/annrheumdis-2018-214323 (2019).
- 33 Fernando, M. M. *et al.* Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS genetics* **4**, e1000024, doi:10.1371/journal.pgen.1000024 (2008).
- 34 Wu, Y. *et al.* Isoliquiritigenin prevents the progression of psoriasis-like symptoms by inhibiting NF-kappaB and proinflammatory cytokines. *Journal of molecular medicine* **94**, 195-206, doi:10.1007/s00109-015-1338-3 (2016).
- 35 Malfitano, A. M. *et al.* Arvanil inhibits T lymphocyte activation and ameliorates autoimmune encephalomyelitis. *Journal of neuroimmunology* **171**, 110-119, doi:10.1016/j.jneuroim.2005.09.005 (2006).
- 36 Altman, A. & Kong, K. F. Protein kinase C inhibitors for immune disorders. *Drug discovery today* **19**, 1217-1221, doi:10.1016/j.drudis.2014.05.008 (2014).
- 37 Oleksyn, D. *et al.* Protein kinase Cbeta is required for lupus development in Sle mice. *Arthritis and rheumatism* **65**, 1022-1031, doi:10.1002/art.37825 (2013).
- 38 Sekhon, S. & Koo, J. Indirubin: a novel topical agent in the treatment of psoriasis. *The British journal of dermatology* **178**, 21, doi:10.1111/bjd.16074 (2018).
- 39 Zhao, Y. *et al.* Indirubin modulates CD4(+) T-cell homeostasis via PD1/PTEN/AKT signalling pathway in immune thrombocytopenia. *Journal of cellular and molecular medicine* **23**, 1885-1898, doi:10.1111/jcmm.14089 (2019).
- 40 Toro-Dominguez, D. *et al.* Stratification of Systemic Lupus Erythematosus Patients Into Three Groups of Disease Activity Progression According to Longitudinal Gene Expression. *Arthritis & rheumatology* **70**, 2025-2035, doi:10.1002/art.40653 (2018).
- 41 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 42 swamp: Visualization, Analysis and Adjustment of High-Dimensional Data in Respect to Sample Annotations. (2018).
- 43 Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S.* (2002).

- 44 Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929-944, doi:10.1016/j.cell.2014.06.049 (2014).
- 45 Brodin, P. *et al.* Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37-47, doi:10.1016/j.cell.2014.12.020 (2015).
- 46 Li, J., Liu, Y., Kim, T., Min, R. & Zhang, Z. Gene expression variability within and between human populations and implications toward disease susceptibility. *PLoS computational biology* **6**, doi:10.1371/journal.pcbi.1000910 (2010).
- 47 Heyn, H. *et al.* DNA methylation contributes to natural human variation. *Genome research* **23**, 1363-1372, doi:10.1101/gr.154187.112 (2013).
- 48 Barturen, G., Beretta, L., Cervera, R., Van Vollenhoven, R. & Alarcon-Riquelme, M. E. Moving towards a molecular taxonomy of autoimmune rheumatic diseases. *Nature reviews. Rheumatology* **14**, 75-93, doi:10.1038/nrrheum.2017.220 (2018).
- 49 Weiner 3rd, J. & Domaszewska, T. tmod: an R package for general and multivariate enrichment analysis. *PeerJ Preprints* **4**, e2420v2421, doi:10.7287/peerj.preprints.2420v1 (2016).
- 50 Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358, doi:10.1093/bioinformatics/bts163 (2012).
- 51 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 52 Stunnenberg, H. G., International Human Epigenome, C. & Hirst, M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1897, doi:10.1016/j.cell.2016.12.002 (2016).
- 53 Fernandez, J. M. *et al.* The BLUEPRINT Data Analysis Portal. *Cell systems* **3**, 491-495 e495, doi:10.1016/j.cels.2016.10.021 (2016).
- 54 Fox, J. & Weisberg, S. *An {R} Companion to Applied Regression*. 2 edn, (Sage, 2011).
- 55 Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biometrical journal. Biometrische Zeitschrift* **50**, 346-363, doi:10.1002/bimj.200810425 (2008).
- 56 Josse, J. & Husson, F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software* **70**, 1-31, doi:doi:10.18637/jss.v070.i01 (2016).
- 57 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).
- 58 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417, doi:10.1016/j.cell.2017.10.049 (2017).
- 59 Bousfield, D. *et al.* Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources. *F1000Research* **5**, doi:10.12688/f1000research.7911.1 (2016).

ACKNOWLEDGMENTS

This work has been supported through a grant from the Innovative Medicines Initiative Joint Undertaking No. 115565 and in-kind and in-cash contributions from the EFPIA partners. The authors would like to particularly express their gratitude to the patients, nurses and many others who helped directly or indirectly in the consecution of this study.

AUTHOR CONTRIBUTIONS

MEAR, LL, FMcD, EB, LB, JM, J-OP designed the study; **LL and LB** organized the electronic Clinical Report Form and supervised the clinical information; **The PRECISEADS Clinical Consortium** agreed on the clinical data, recruited the patients, performed their clinical assessment, and obtained the samples; **JM and LL** monitored the patient recruitment, the informed consents and the ethical approvals; **LleL, CJ, J-OP and the Flow Cytometry Consortium** optimized and calibrated the flow cytometers across all recruitment centers, normalized, gated and procured the data; **GB, SB, ZM, ECM, MK, JRU, AGG, JK, AB, SH, JM, RL** performed RNASequencing, genotyping, methylation, data curation, and data processing; **MOB, JF, YR** performed the cytokine, small lipid moiety antibodies and coordinated the routine antibody laboratory tests; **JW** organized the data and prepared the data for analysis; **GB, RJB, CC, ERD, JI, LL, JM, FMcD, SR, MEAR, JF, JM, LB, J-OP** oversaw the progress of the study; **GB, FCM, RJB, JI, FMcD, MEAR, LB, some members of the PRECISEADS Clinical Consortium** discussed and interpreted the results; **GB, FCM, DTD, MMB** performed the analyses; **LL** compiled the clinical information and wrote the materials and methods section for the clinical data; and **GB and MEAR** wrote the manuscript. All authors approved of the content of the manuscript.

AUTHOR INFORMATION

Correspondence should be directed to: Marta E. Alarcón-Riquelme, Department of Medical Genomics, Center for Genomics and Oncological Research (GENYO) Pfizer-University of Granada-Andalusian Regional Government. Health Sciences Technology Park, 18016, Granada, Spain. E-mail: marta.alarcon@genyo.es

Disclosures:

Employees of BAYER Pharma: Sepideh Babaei, Zuzanna Makowska, Jorge Kageyama, Anne Buttgerreit, Sikander Hayat, Fiona McDonald, Joerg Mueller, Ralf Lesche.

Employee of Eli Lilly and Co: Robert J. Benschop, Ernst R. Dow

Employee of UCB: Chris Chamberlain, John Ioannou, María Juárez, Jacqueline Marovac, Ian White.

Employee of Sanofi: Emanuele De Rinaldis, Sambasiva Rao.

Employee of IRIS Servier: Laurence Laigle.

FIGURE LEGENDS

Figure 1. Molecular pattern distribution is limited to 4 clusters and does not follow the clinical diagnoses. (A) Heatmap showing the distribution of gene and CpG functional modules across the 4 autoimmune disease clusters. In columns patients are grouped by cluster assignment and in rows the functional modules of the features are shown with their scaled median values. Each subset of patients represented (discovery and validation sets) is presented separately. For the transcriptome, red represents over-expression and blue represents under-expression. For the methylome, purple represents hypo-methylation and orange represents hyper-methylation. At the top of the figure the annotation shows: two configurations of clusters for 4 (4 CLUSTERS) and 2 (2 CLUSTERS) groups, each of the treatment groups for each individual (SABIO, systemic antibiotics; STED, steroids; BIO, biologics; IMS, immunosuppressors and AM, antimalarials), recruitment centers distribution, disease activity as physician global assessment (PGA), duration of the disease since diagnosis, age at onset, gender, age and diagnosis. The abbreviations for the recruitment centers can be found in Supplementary Table 1. (B) Mosaic plot showing the distribution of diseases in each cluster (C1-C4, clusters 1 to 4) with the number of individuals inside each block. Each disease is represented by a color in columns. (C) Association of covariates with clinical diagnosis, molecular clusters, and the associations between them. The significance of the associations is represented as a scale ranging from blue (non-significant) to red (significant). The direction of the association is shown as the z-scored beta coefficients where orange is enrichment and purple is depletion.

Figure 2. OMIC layers of information functionally characterize each of the molecular clusters. (A) Annotation of the selected features according to the hypergeometric enrichment of their modular functional assignment. The module annotations were obtained using the blood immunological signature databases of Chaussabel et al. (Ch) and Li et al (Li). Only significant enrichment results are shown (q -value < 0.01). Significant modules are shown in columns and their annotation in rows. (B) Association between CpG and gene modules ($FDR < 0.01$ was considered significant). Left and right alluvial diagrams show cis and trans associations (cis associations were considered at distances < 10 Kb). (C) Proportion of significant and non-significant cis-associations across modules. Significant ones are annotated by module. (D) Enrichment of histone marks (rows) in CpG modules according to blood cell types (in columns). Significant enrichments (q -value < 0.01) are colored by their enrichment score (red means enriched and blue depleted). (E) Enrichment of transcription factor binding sites in the CpG modules. Only significant enrichments (q -value < 0.01) with absolute values of enrichment score > 1 are shown.

Figure 3. Each cluster is associated with specific low-dimensional layers of information. Z-scored beta coefficients of the significant ANOVA associations ($FDR < 0.01$) for (A) serology, (B) flow-cytometry, and (C) clinical data, are shown. The serology information included autoantibodies (blue), cytokines (green), and antibodies against small lipid moieties or natural autoantibodies (red). Abbreviations are as follows: CCP2, anti-citrullinated peptide antibodies; CENTB, anti-centromere B; DNA, anti-dsDNA; SM, anti-Sm; SSB, anti-SSB or anti-La; U1RNP, anti-U1RNP; PFLC, Protein free light chains; SSA, anti-SSA or anti-Ro; PC.IGM, IgM anti-phosphatidylcholine. Hierarchical clustering of the results expressed as Z-scored beta coefficients are represented in the circle plots (A, B, and C). The four autoimmune molecular clusters are shown concentrically from 1, the outermost, to 4, the innermost circle. Clinical information was summarized using principal components (PC) and the PCs most significantly associated with each cluster are shown. (D) Only those clinical items having a significant contribution to each significant PC (observed contribution higher than five times the expected) are depicted.

Figure 4. Three of the clusters present aberrant molecular patterns that may be reverted by specific drugs. (A) Distribution of healthy individual assignments to the SADs molecular classification. The pie diagram shows that nearly 74% of controls are similar to patients in cluster 2. (B) Differentially expressed genes (DEGs) between clusters and healthy controls. Top barplot (black bars) shows shared DEGs across clusters, right barplot (colored bars) represents the number of DEGs by condition. The dot connectivity plot represents the intersections across clusters. (C) SLEDAI boxplots by molecular cluster, $n=138$ SLE patients. (D) ESSDAI boxplots by molecular cluster, $n=79$ SjS patients. In both boxplots, the lower and upper hinges correspond to the first and third quartiles, while whiskers represent the $1.5 \times$ interquartile ranges. (E) Heatmap showing the drugs (columns) with the highest probability of reverting the aberrant molecular pattern of each pathological cluster (rows). Only drugs with a connectivity score below -90 for at least one of the clusters are shown (Red values, related molecular profiles between clusters and drugs; green values, opposite molecular profiles between clusters and drugs). Euclidean distance and complete linkage methods were used for drug clustering. The mechanisms of action for each drug can be consulted in Supplementary Figure 7.

Figure 5. Aberrant molecular patterns are stable in time and related to relapsing moments of disease. (A) The stability Jaccard index between molecular cluster assignments at baseline (rows) and at 6 months (columns). (B) Stability jaccard index between molecular cluster assignments at baseline (rows) and at 14 months (columns). (C) Patient classification considering the 3 time point cluster assignments. The definitions are as follows: stable patients, patients assigned to exactly the same cluster in the 3 time points; relapse-remission patients, patients assigned to only one pathological cluster (cluster 1, 3 or 4) but have been at any given time point in the healthy-like cluster 2; unstable patients, patients assigned to more than one aberrant cluster at any one time point.

A.6 Standard operating procedure

SOP calibration

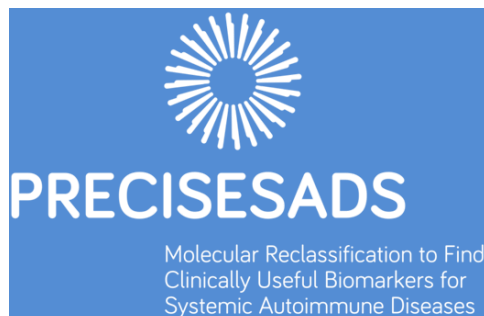


Work Package 5

Standard Operating Procedure #1

for the QC 8 Peaks Beads protocol

for Beckman Coulter users





Evaluation Test Procedure for Multi-Color Duraclone Precisesads project

PURPOSE

The purpose of this document is to provide instructions and procedures required for the generation of the proper settings with 8 Peaks beads on a Navios and Gallios flow cytometers.

PREREQUISITE

It is absolutely required to create a **specific user** in your flow cytometer that will be dedicated to the PRECISESADS project.

Thus, new names should be assigned to the parameters:

	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	FITC-A
P8	FL2 INT Log (PE)	PE-A
P9	FL4 INT Log (PC5.5)	PC5-5-A
P10	FL5 INT Log (PC7)	PC7-A
P11	FL6 INT Log (APC)	APC-A
P12	FL8 INT Log (APC-Alexa750)	APC-AF750-A
P13	FL9 INT Log (PB)	PB-A
P14	FL10 INT Log (Krome Orange)	KO-A
P15	TIME	TIME

These names will be the same in all centers and all cytometers (BD and BC flow cytometers) to enable the comparisons of the data during the whole project.



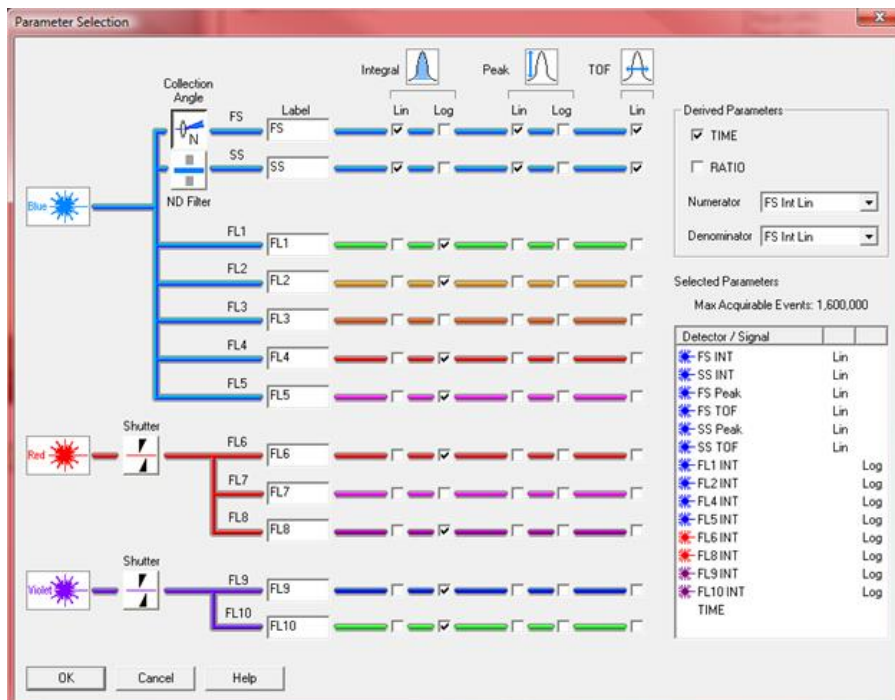
SUMMARY

Part 1: Creation of the QC 8 Peaks protocol	page 4
Part 2: Adjustment of PMTs with Capture Beads	page 9
Part 3: Acquisition of the QC 8 Peaks beads	page 14
Annex 1: Examples of beads acquisition	page 17
Annex 2: Full addresses of the participants	page 20

TO START: OPEN YOUR SPECIFIC PRECISESADS CONFIGURATION

Part 1: Creation of the QC 8 Peaks protocol

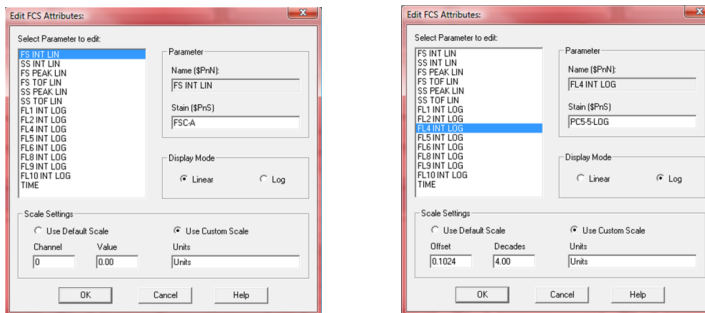
1. To create the new protocol, go to "Cytometer":
 - a. select "Cytometer Controls"
 - b. click on "Parameters"
 - c. select the "N" Collection Angle
 - d. select all the parameters, as described below, in the following order:



This ordered selection of the parameters will facilitate the comparison of the data with the BD flow cytometers.

2. Change the description of the different parameters:
 - a. Open "File"
 - b. Select "Edit FCS Header Attributes"
 - c. Select the Parameter to Edit
 - d. In the "Stain(\$PnS)" box, change the name as required (see the table in f).

Examples



- e. The table mentions all the parameters that must be selected in the indicated order, with their specific name for the PRECISESADS project:

	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	FITC-A
P8	FL2 INT Log (PE)	PE-A
P9	FL4 INT Log (PC5.5)	PC5-5-A
P10	FL5 INT Log (PC7)	PC7-A
P11	FL6 INT Log (APC)	APC-A
P12	FL8 INT Log (APC-Alexa 750)	APC-AF750-A
P13	FL9 INT Log (PB)	PB-A
P14	FL10 INT Log (Krome Orange)	KO-A
P15	TIME	TIME



3. The 8 Peaks Beads (reference A79016, lot AF01):

- a. The beads are excited by Blue, Red and Violet laser with a fluorescence emission on all channels. The corresponding dyes are indicated in the following table

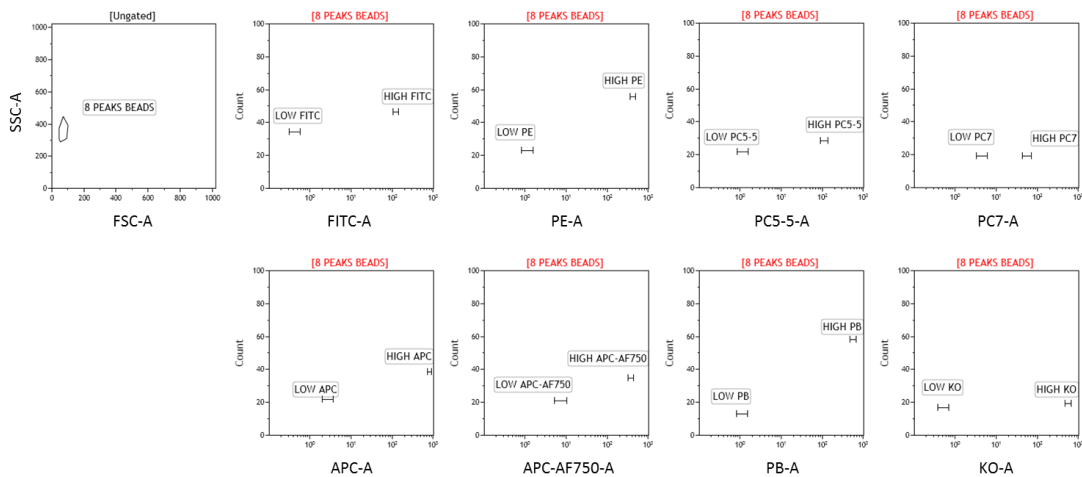
Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
FITC	PE	PC5.5 PerCP Cy5,5	PC7	APC	APC AF750 AF750 APC H7	PB V450	KO V500

- b. The target values for the 8 Peaks beads have been designed at Brest on a reference Navios flow cytometer. These target values are optimized for the Beckman Coulter Duraclone QC.

They should allow obtaining the same settings in every center (caution, not necessarily with the same PMT voltages!).

4. Create the plots of the **QC 8 Peaks protocol**:

- a. 1 ungated FSC-A/SSC-A bi-parametric plot, containing a region called **"8 PEAKS BEADS"**
- b. 8 mono-parametric Log fluorescence plots (one for each fluorescence)
 - i. All are gated on the **"8 PEAKS BEADS"** region
 - ii. All contain 2 linear region called **"HIGH fluorescence"** and **"LOW fluorescence"**, i.e.
 - FITC-A plot contains 2 regions called **"LOW FITC"** and **"HIGH FITC"**.
 - PE-A plot contains 2 regions called **"LOW PE"** and **"HIGH PE"**.
 - PC5-5-A plot contains 2 regions called **"LOW PC5-5"** and **"HIGH PC5-5"**.
 - PC7-A plot contains 2 regions called **"LOW PC7"** and **"HIGH PC7"**.
 - APC-A plot contains 2 regions called **"LOW APC"** and **"HIGH APC"**.
 - APC-AF750-A plot contains 2 regions called **"LOW APC-AF750"** and **"HIGH APC-AF750"**.
 - PB-A plot contains 2 regions called **"LOW PB"** and **"HIGH PB"**.
 - KO-A plot contains 2 regions called **"LOW KO"** and **"HIGH KO"**.





- c. The HIGH and LOW regions enable to determine the targets of the 8 peaks beads. They must be positioned at the following median values:

Target values of the 8 Peaks beads (ref# A79016, lot AF01) for the 8 fluorescence parameters

	SPECIFIC NAMES	REGION	MEDIAN values
P7	FITC-A	LOW FITC	0.45
		HIGH FITC	131.17
P8	PE-A	LOW PE	1.18
		HIGH PE	453.94
P9	PC5-5-A	LOW PC5-5	1.23
		HIGH PC5-5	115.7
P10	PC7-A	LOW PC7	4.64
		HIGH PC7	57.6
P11	APC-A	LOW APC	2.63
		HIGH APC	832.07
P12	APC-AF750-A	LOW APC-AF750	8.29
		HICH APC-AF750	358.89
P13	PB-A	LOW PB	1.13
		HIGH PB	632.04
P14	KO-A	LOW KO	0.51
		HIGH KO	590.62

5. Set the PMT voltages to detect the 8 peaks beads.

	PARAMETERS	SPECIFIC NAMES	PMT voltages	Gain
P1	FSC INT Lin	FSC-A	1000	2.0
P2	SSC INT Lin	SSC-A	473	5.0
P7	FL1 INT Log (FITC)	FITC-A	421	1.0
P8	FL2 INT Log (PE)	PE-A	467	1.0
P9	FL4 INT Log (PC5.5)	PC5-5-A	484	1.0
P10	FL5 INT Log (PC7)	PC7-A	562	1.0
P11	FL6 INT Log (APC)	APC-A	524	1.0
P12	FL8 INT Log (APC-Alexa 750)	APC-AF750-A	631	1.0
P13	FL9 INT Log (PB)	PB-A	414	1.0
P14	FL10 INT Log (Krome Orange)	KO-A	382	1.0

Note: These PMT voltages are only indications determined on Navios. Depending on your Beckman flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.



Part 2: Adjustment of PMTs with Capture Beads

Important Note: FLOW CYTOMETER SETTINGS

The PMT voltages for each channel set in the Part 1 have been initially determined to have equivalent fluorescence intensities between BD and BC flow cytometers.

These settings based on the 8 Peaks Beads fluorescences seem appropriate for fluorescence intensities.

However, due to intrinsic differences between flow cytometers (electronic, filters, mirrors...) intermediate fluorescence intensities may vary between BD and BC cytometers.

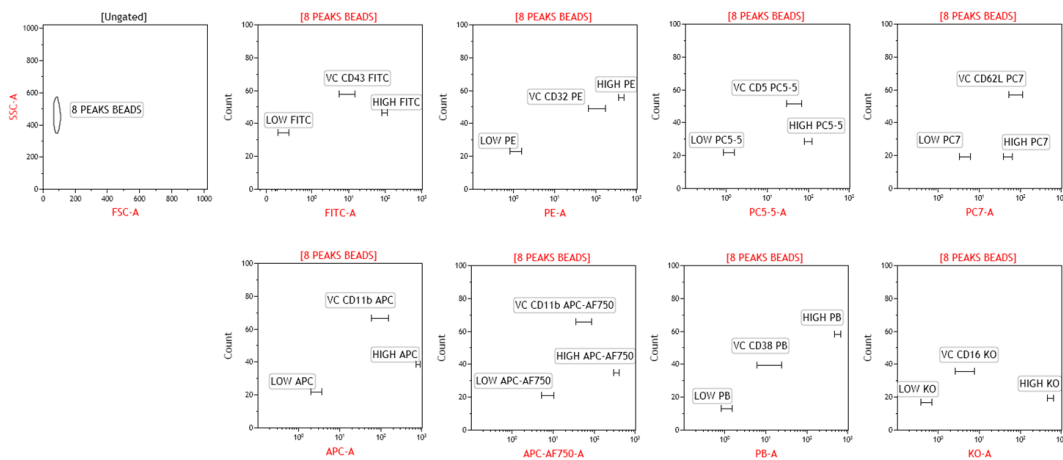
Adjustments are needed to achieve the fine comparisons required in the Precisesads project.

MATERIALS REQUIRED

- VersaComp Antibody Capture Bead Kit (#B22804, Beckman Coulter)
- Staining buffer: 1x PBS pH 7.4 with 0.2% BSA

1. Load the **QC 8 Peaks protocol**:

- a. Insert 1 linear region called "VC CDxx fluorescence" in each plot, i.e.
 - FITC-A plot contains 1 region called "**VC CD43 FITC**".
 - PE-A plot contains 1 region called "**VC CD32 PE**".
 - PC5-5-A plot contains 1 region called "**VC CD5 PC5-5**";
 - PC7-A plot contains 1 region called "**VC CD62L PC7**".
 - APC-A plot contains 1 region called "**VC CD11b APC**".
 - APC-AF750-A plot contains 1 region called "**VC CD11b APC-AF750**".
 - PB-A plot contains 1 region called "**VC CD38 PB**".
 - KO-A plot contains 1 regions called "**VC CD16 KO**".



- b. The VC CDxx regions enable to determine the targets of the capture beads. They must be positioned at the following median values:

Target values of the VersaComp Antibody Capture beads (ref# B22804) for the 8 fluorescence parameters

	SPECIFIC NAMES	REGION	MEDIAN values
P7	FITC-A	VC CD43 FITC	9.28
P8	PE-A	VC CD32 PE	115.12
P9	PC5-5-A	VC CD5 PC5-5	47.20
P10	PC7-A	VC CD62L PC7	82.05
P11	APC-A	VC CD11b APC	102.63
P12	APC-AF750-A	VC CD11b APC-AF750	61.05
P13	PB-A	VC CD38 PB	15.34
P14	KO-A	VC CD16 KO	4.57

Note: These values were obtained using monoclonal antibodies which enter into the composition of the Duraclone tubes.



2. Prepare 8 cytometer tubes for the **Capture beads**
 - a. In each cytometer tube, put 45 μ l of negative beads and 45 μ l of positive beads (ref# B22804)
 - b. Add 1 single antibody in each tube
 - i. Tube 1: 5 μ l of anti-CD16 (KO-conjugated)
 - ii. Tube 2: 5 μ l of anti-CD38 (PB-conjugated)
 - iii. Tube 3: 5 μ l of anti-CD11b (APC-AF750-conjugated)
 - iv. Tube 4: 5 μ l of anti-CD11b (APC-conjugated)
 - v. Tube 5: 5 μ l of anti-CD62L (PC7-conjugated)
 - vi. Tube 6: 5 μ l of anti-CD5 (PC5.5-conjugated)
 - vii. Tube 7: 10 μ l of anti-CD32 (PE-conjugated)
 - viii. Tube 8: 10 μ l of anti-CD43 (FITC-conjugated)
3. Vortex 2 sec.
4. Incubate 20 min at room temperature in the dark.
5. Add 1mL of Staining Buffer
 - a. Vortex 2 sec
 - b. Centrifuge at 300xg for 6 minutes
6. Discard the supernatant
7. Resuspend the beads in 600 μ l of Staining buffer.



8. Load the **QC 8 Peaks** protocol

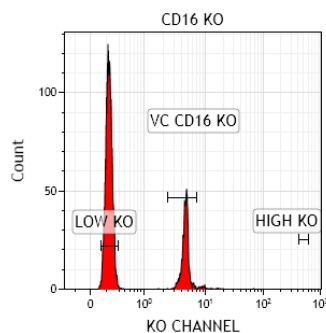
a. Label your **LMD** File as follows:

QC8PeaksBeads_Capture Fluorochrome_ INSTITUTION_INSTRUMENT_ddMMMyyyy of reception_ddMMMyyyy of acquisition

- i. With the name of the protocol = QC8PeaksBeads
- ii. Sample ID1 = Capture Fluorochrome
- iii. Sample ID2 = INSTITUTION_INSTRUMENT
- iv. Sample ID3 = ddMMMyyyy of reception
- v. Sample ID4 = ddMMMyyyy of acquisition

Example: QC8PeaksBeads_Capture KO_UBO_NAVIOS_22OCT2014_22OCT2014

- a. Run the tube 1 containing the anti-CD16-stained Capture beads.
- b. Adjust the **8 PEAKS BEADS** region on FSC-A/SSC-A plot to capture the Beads and exclude any doublets.
- c. Adjust fluorescence's PMT voltage of P14 (KO-A) to insert the peak of the positive beads in the target "VC CD16 KO", if needed



9. Repeat step 8 and 9 for each tube.

10. Save the **QC 8 Peaks protocol** for your laboratory with the newly determined voltages.



The following table shows no changes of the PMT voltages after acquisition of the 8 tubes

	PARAMETERS	SPECIFIC NAMES	PMT voltages	Gain
P1	FSC AREA Lin	FSC-A	1000	2.0
P2	SSC AREA Lin	SSC-A	473	5.0
P7	FITC A Log	FITC-A	421	1.0
P8	PE A Log	PE-A	467	1.0
P9	PC5.5 A Log	PC5-5-A	484	1.0
P10	PC7 A Log	PC7-A	562	1.0
P11	APC A Log	APC-A	524	1.0
P12	APC-Alexa750 A Log	APC-AF750-A	631	1.0
P13	PB A Log	PB-A	414	1.0
P14	KO A Log	KO-A	382	1.0

Note: These PMT voltages are only indications determined on Navios. Depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

If you have change some PMT voltages, the new values will induce a modification into the fluorescence intensities of the 8 Peaks Beads.

The target position of “LOW fluorescence” and “HIGH fluorescence” might thus be changed. Therefore, repeat the acquisition of the 8 Peaks Beads (Part 3) to re-position your “LOW fluorescence” and/or “HIGH fluorescence” target.

These new positions will now correspond to your own internal QC control.



Part 3: Acquisition of the QC 8 Peaks beads

Important Note

Following the adjustments of the PMT voltages performed with the VersaComp Antibody Capture Bead in Part#2, the position of the targets “LOW fluorescence” and “HIGH fluorescence” may have changed.

The position of the targets must be definitely adjusted with a supplementary acquisition of the 8 Peaks Beads.

1. Prepare the 8 Peaks beads

- a. In a cytometer tube, put 500µl of purified water (ref# BO0209C)
- b. Add 1 droplet of 8 Peaks Beads (ref# A79016, lot AF01)

2. Load the QC 8 Peaks protocol

- a. Label your LMD File as follows:

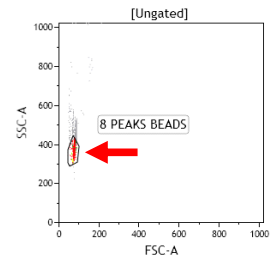
QC8PeaksBeads_After Capture Beads_ INSTITUTION_INSTRUMENT_ddMMMyyyy of reception_ddMMMyyy of acquisition

- i. With the name of the protocol = QC8PeaksBeads
- ii. Sample ID1 = After Capture Beads
- iii. Sample ID2 = INSTITUTION_INSTRUMENT
- iv. Sample ID3 = ddmMMMyyyy of reception
- v. Sample ID4 = ddMMMyyyy of acquisition

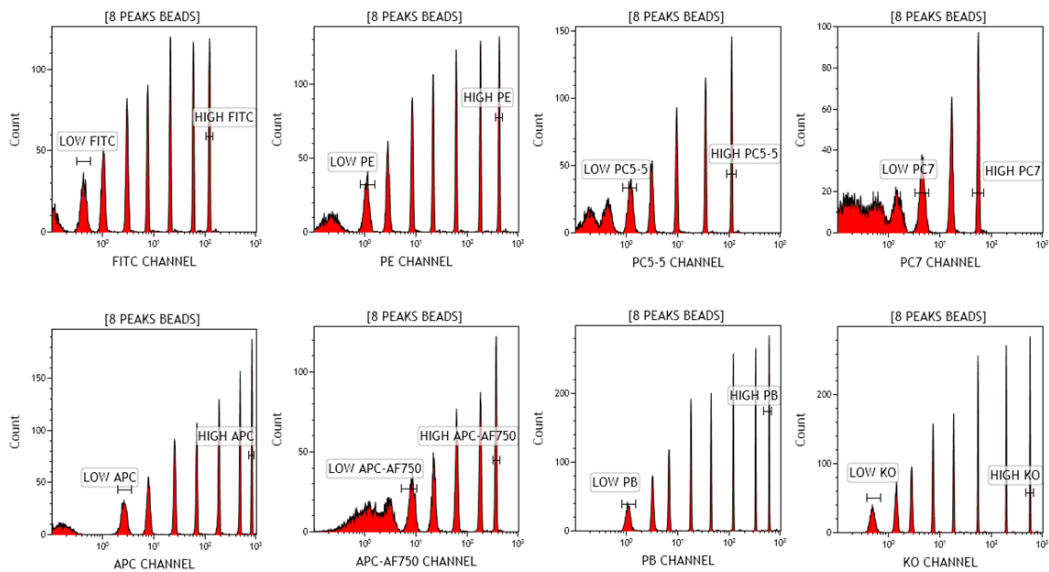
Example: QC8PeaksBeads_After Capture Beads_UBO_NAVIOS_22OCT014_22OCT2014

- b. Run the tube containing the diluted 8 peaks beads at low speed

c. Adjust the **8 PEAKS BEADS** region on FSC-A/SSC-A plot to capture the Beads and exclude any doublets (red arrow):



d. Adjust the position of the targets to encompass the corresponding peaks, if needed





3. Save the protocol for your laboratory with the newly determined voltages and the associated new position of the “LOW fluorescence” and “HIGH fluorescence” targets.

This protocol must be specifically used for the PRECISESADS project; otherwise the data cannot be compared between sites.

Indicate the new values that have been determined in the following table:

	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages	Gain
P1	FSC INT Lin	FSC-A		2.0
P2	SSC INT Lin	SSC-A		5.0
P7	FL1 INT Log (FITC)	FITC-A		1.0
P8	FL2 INT Log (PE)	PE-A		1.0
P9	FL4 INT Log (PC5.5)	PC5-5-A		1.0
P10	FL5 INT Log (PC7)	PC7-A		1.0
P11	FL6 INT Log (APC)	APC-A		1.0
P12	FL8 INT Log (APC-Alexa750)	APC-AF750-A		1.0
P13	FL9 INT Log (PB)	PB-A		1.0
P14	FL10 INT Log (Krome Orange)	KO-A		1.0

These voltages will be used in the panel’s sheets for the analyses of the blood samples (see SOP#3).

4. You are now ready to prepare the protocols for the blood sample analyses (see SOP#2)

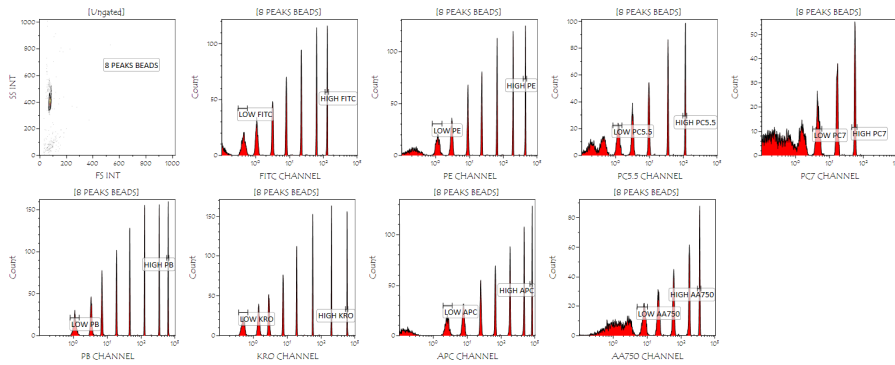
AT THE END: QUIT YOUR SPECIFIC PRECISESADS CONFIGURATION



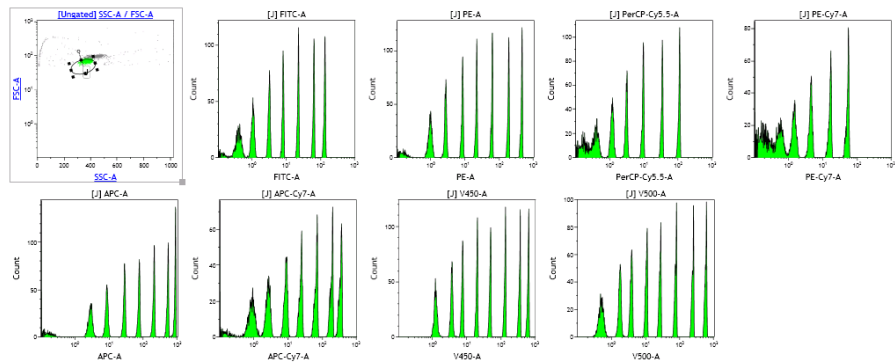
Annex 1: Examples of beads acquisition

EXAMPLES OF 8 PEAK BEADS ACQUISITION before capture beads acquisition

Navios cytometer

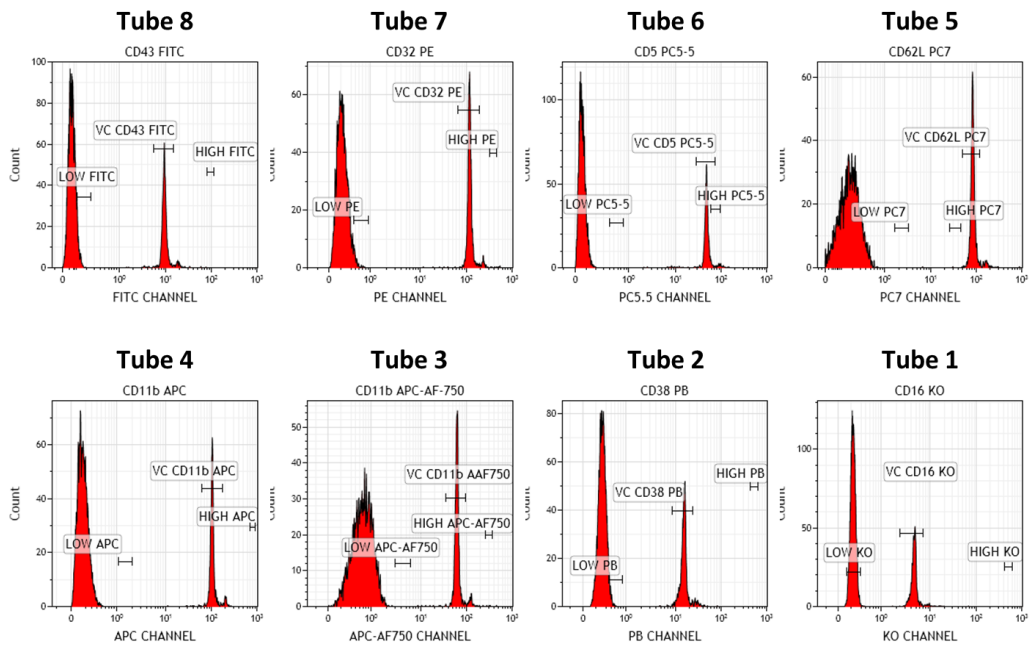


Facs VERSE cytometer



EXAMPLES OF THE 8 TUBES FOR THE CAPTURE BEADS ACQUISITION

Navios cytometer





If you have any issue with this procedure you can contact:

UBO team: Christophe JAMIN christophe.jamin@univ-brest.fr

or Jacques-Olivier PERS pers@univ-brest.fr

and

Beckman Coulter reference team: Olivier JAEN ojaen@beckman.com

or Nicole WEIT nweit@beckman.com



Annex 2: Full addresses of the participants

Note:

Participants 1 to 8: Task 5.1 and 5.2: panel 1 to panel 9
 Participants 9 to 11: Task 5.1 only: panel 1 and panel 2

Task 5.1 and 5.2

1	UBO	JO Pers, C Jamin	Navios	Beckman Coulter
2	FPS	C Marañon	FACS Verse	BD
3	SAS CORDOBA	R Lopez, N Barbarroja	FACS Aria III	BD
4	IDIBELL	E Ballestar, L Ciudad, D Alvarez, N Ramirez	Gallios	Beckman Coulter
5	IRCCS	E Trombetta	FACS Canto II	BD
6	DRFZ	F Hiepe, R Biesen V Gerl	FACS Canto II	BD
7	UNIGE	C Chizzolini, A Zuber	Navios	Beckman Coulter
8	K.U.LEUVEN	R Lories, E De Langhe, T Cantaert	LSR Fortessa	BD
9	UCL	B Lauwerys, J Ducreux	FACS Canto II	BD

Task 5.1

10	CHPORTO	C Vasconcelos, E Neves	Navios	Beckman Coulter
11	MHH	T Witte	FACS Canto II	BD



1- UBO (Brest): Navios (BCoulter)

Mr Christophe JAMIN
Laboratoire d'Immunologie
CHRU Morvan
5 av Foch
BP824
29609 Brest
FRANCE

christophe.jamin@univ-brest.fr
pers@univ-brest.fr

2- FPS (Granada): FACS Verse (BD)

Miss Concepción MARAÑÓN
PTS Granada
Parque Tecnológico de Ciencias de la Salud
Avenida de la Ilustración 114
18016
Granada
SPAIN

concepcion.maranon@genyo.es
marta.alarcon@genyo.es

3- SAS (Cordoba): FACS Aria III (BD)

Rosario Lopez-Pedrerá
IMIBIC-Unidad de Investigación
Hospital Universitario Reina Sofía
Avenida Menéndez Pidal s/n
14004
Córdoba
SPAIN

Tels. +34957736542 / +34651319413

rosario.lopez.exts@juntadeandalucia.es
nuria.barbarroja.exts@juntadeandalucia.es



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

4- IDIBELL (Barcelona): Gallios (BCoulter)

Prof Esteban BALLESTAR and Dr Laura CIUDAD
Cancer Epigenetics and Biology Programme (PEBC)
Bellvitge Biomedical Research Institute
Avda. Gran Via 199-203
08908 L'Hospitalet de Llobregat
Barcelona
SPAIN

eballestar@idibell.cat

lciedad@idibell.cat

dalvarez@idibell.cat

nramirez@idibell.cat

5- IRCCS (Milan): FACS Canto II (BD)

Elena Trombetta
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico
via F. Sforza n° 35
Padiglione Invernizzi 5° piano - Servizio di Citometria
20122 - Milano (MI)
Italy

trombetta.elena@gmail.com

6- DRFZ (Berlin): FACS Canto II (BD)

Prof Falk HIEPE
Dr Robert BIESEN
Dr Velia GERL
Department of Rheumatology and Clinical Immunology, Charité
Universitätsmedizin Berlin
Charitéplatz 1
D-10117 Berlin
GERMANY

falk.hiepe@charite.de

Robert.Biesen@charite.de

gerl@drfz.de



7- UNIGE (Geneva): Navios (BCoulter)

Prof Carlo CHIZZOLINI
Immunology and Allergy
University Hospital and School of Medicine
Rue Gabrielle Perret-Gentil 4
Geneva 14
1211
SWITZERLAND

Carlo.Chizzolini@unige.ch
Aleksandra.Zuber@unige.ch

8- KUL (Leuven): Fortessa (BD)

Prof Rik Lories
Faculty of Medicine
Laboratory of Tissue Homeostasis and Disease
Skeletal Biology and Engineering Research Center (SBE)
Department of Development and Regeneration
O&N - Box 813
Herestraat 49
B3000 Leuven
Belgium

Rik.Lories@med.kuleuven.be
ellen.delanghe@uzleuven.be
tineke.cantaert@med.kuleuven.be

9- UCL (Louvain) : FACS Canto II (BD)

Julie Ducreux
Service de Rhumatologie
Cliniques Universitaires Saint-Luc & Institut de Recherche Expérimentale et Clinique
Université catholique de Louvain
Avenue Hippocrate 10
bte B2.5390
1200 Bruxelles
BELGIQUE

Bernard.Lauwerys@uclouvain.be
julie.ducreux@uclouvain.be



10- CHP (Porto): Navios (BCoulter)

Dr Esmeralda NEVES
Centro Hospitalar do Porto
Serviço de Imunologia EX-CICAP
Rua D. Manuel II
4050-345
Porto
PORTUGAL

cvcarlosvasconcelos@gmail.com
m.esmeraldaneves@gmail.com

11- MHH (Hannover): FACS Canto II (BD)

Prof Torsten WITTE
Klinik für Immunologie und Rheumatologie
Medizinische Hochschule Hannover
Carl-Neuberg-Str. 1
30625
Hannover
GERMANY

Witte.Torsten@mh-hannover.de

SOP préparation des protocoles d'acquisition



innovative
medicines
initiative



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

Work Package 5

Standard Operating Procedure #2

Flow Cytometry protocol preparation

for BC users



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases



Protocol creation for Multi-Color Duraclone Precisesads project

PURPOSE

The purpose of this document is to provide instructions and procedures required for the creation of the different protocols dedicated to the PRECISESADS project on Navios and Gallios Beckman Coulter flow cytometers.

PREREQUISITE

It is absolutely required to create a **specific configuration** in your flow cytometer that will be dedicated to the PRECISESADS project.

Thus, new names should be assigned to the parameters:

	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	FITC-A
P8	FL2 INT Log (PE)	PE-A
P9	FL4 INT Log (PC5.5)	PC5-5-A
P10	FL5 INT Log (PC7)	PC7-A
P11	FL6 INT Log (APC)	APC-A
P12	FL8 INT Log (APC-Alexa 750)	APC-AF750-A
P13	FL9 INT Log (PB)	PB-A
P14	FL10 INT Log (Krome Orange)	KO-A
P15	TIME	TIME

These names will be the same in all centers and all cytometers (BD and BC flow cytometers) to enable the comparisons of the data during the whole project.



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

SUMMARY

Part 1: PMT calibration: page 4

Validation with the QC 8 Peaks protocol

Part 2: Preparation of the Panel's protocols: page 6

A- PANEL 1 protocol page 8

B- PANEL 2 protocol page 14

C- PANEL 3 protocol page 19

D- PANEL 4 protocol page 23

E- PANEL 5 protocol page 27

F- PANEL 6 protocol page 31

G- PANEL 7 protocol page 35

H- PANEL 8 protocol page 39

I- PANEL 9 protocol page 43



TO START: OPEN YOUR SPECIFIC PRECISESADS CONFIGURATION

Part 1: PMT Calibration: Validation using the QC 8 Peaks Beads protocol

Please refer to the SOP#1 for the preparation of 8 Peaks Beads protocol

1. Save the protocol for your laboratory with the determined voltages

This protocol must be specifically used for the PRECISESADS project; otherwise the data can not be compared between sites.

Indicate the values that have been determined with the SOP#1 in the following table:

	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	FITC-A	
P8	FL2 INT Log (PE)	PE-A	
P9	FL4 INT Log (PC5.5)	PC5-5-A	
P10	FL5 INT Log (PC7)	PC7-A	
P11	FL6 INT Log (APC)	APC-A	
P12	FL8 INT Log (APC-Alexa750)	APC-AF750-A	
P13	FL9 INT Log (PB)	PB-A	
P14	FL10 INT Log (Krome Orange)	KO-A	

These voltages will be used in the panel's protocols for the analyses of the blood samples (see Part 2, step 2 of all chapters A to I).

2. You are now ready to prepare the protocols for the blood sample analyses (Part 2)



DO NOT FORGET: OPEN YOUR SPECIFIC PRECISESADS CONFIGURATION

Part 2: Preparation of the Panel's protocols Compensation calibration

General information

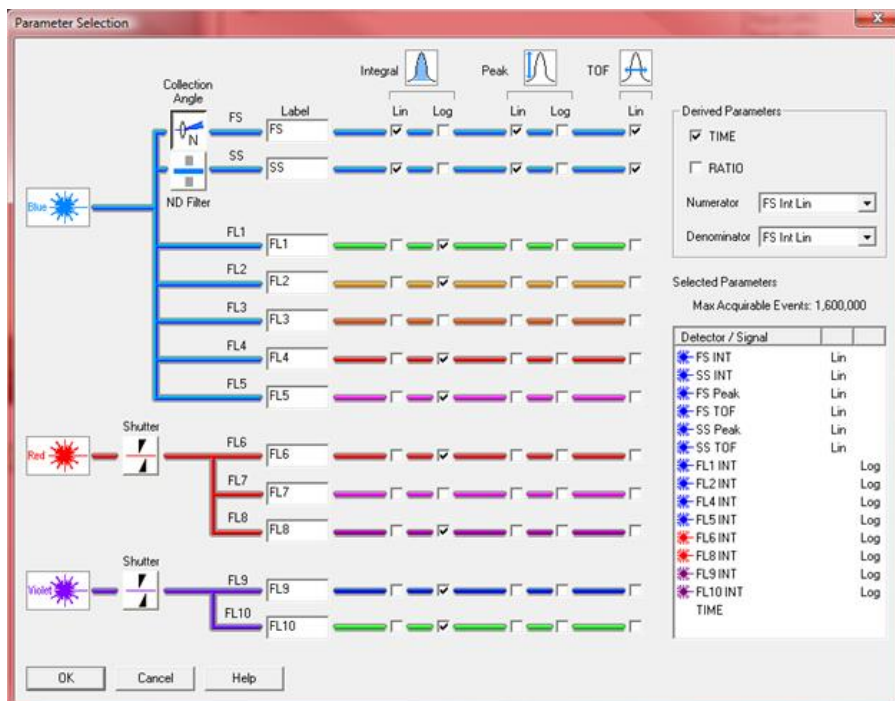
- The following table shows the parameters that have to be selected in the indicated order for the PRECISESADS project:

	PARAMETERS
P1	FSC INT Lin
P2	SSC INT Lin
P3	FS Peak Lin
P4	FS TOF Lin
P5	SS Peak Lin
P6	SS TOF Lin
P7	FL1 INT Log (FITC)
P8	FL2 INT Log (PE)
P9	FL4 INT Log (PC5.5)
P10	FL5 INT Log (PC7)
P11	FL6 INT Log (APC)
P12	FL8 INT Log (APC-Alexa 750)
P13	FL9 INT Log (PB)
P14	FL10 INT Log (Krome Orange)
P15	TIME

- The corresponding dyes are indicated in the following table

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
FITC	PE	PC5.5 PerCP Cy5,5	PC7	APC	APC AF750 AF750 APC H7	PB V450	KO V500

3. To create a new protocol, clic on "Cytometer":
 - a. select "Cyometer Controls"
 - b. clic on "Parameters"
 - c. select the "N" Collection Angle
 - d. select all the parameters, as described below, in the following order:



Parameter Selection

Collection Angle: N

Integral: Lin Log

Peak: Lin Log

TOF: Lin

Label	Lin	Log	Lin	Log	Lin
FS	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
SS	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
FL1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL4	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL5	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL6	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL7	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL8	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL9	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FL10	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Derived Parameters:

TIME

RATIO

Numerator: FS Int Lin

Denominator: FS Int Lin

Selected Parameters:

Max Acquirable Events: 1,600,000

Detector / Signal	Scale
FS INT	Lin
SS INT	Lin
FS Peak	Lin
FS TDF	Lin
SS Peak	Lin
SS TDF	Lin
FL1 INT	Log
FL2 INT	Log
FL4 INT	Log
FL5 INT	Log
FL6 INT	Log
FL8 INT	Log
FL9 INT	Log
FL10 INT	Log
TIME	

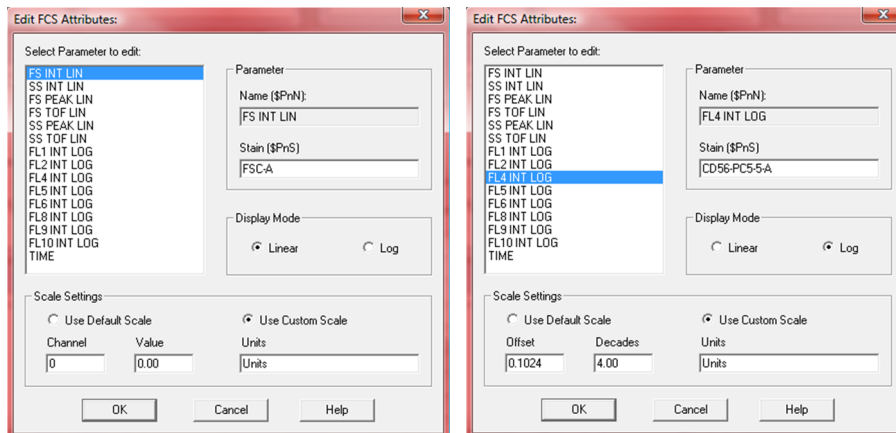
OK Cancel Help

This ordered selection of the parameters will facilitate the comparison of the data with the BD flow cytometers.

4. To change the description of the different parameters:

- a. Open "File"
- b. Select "Edit FCS Header Attributes"
- c. Select the Parameter to Edit
- d. In the "Stain(\$PnS)" box, change the name as required (see §1 of each chapter).

Example for 2 parameters:



5. For the labeling of your LMD File, use the following items:

- a. The name of the protocol = Panel x
- b. Sample ID1 = patient ID
- c. Sample ID2 = INSTITUTION_INSTRUMENT
- d. Sample ID3 = ddMMMyyy of reception
- e. Sample ID4 = ddMMMyyy of acquisition

Panel x_patient ID_INSTITUTION_INSTRUMENT_ddMMMyyy of reception_ddMMMyyy of acquisition

Example: Panel 1_32140001_UBO_NAVIOS_09JUL2014_09JUL2014

A. Create a new protocol named PANEL 1

- The following parameters have to be selected in the indicated order, with the specific names:

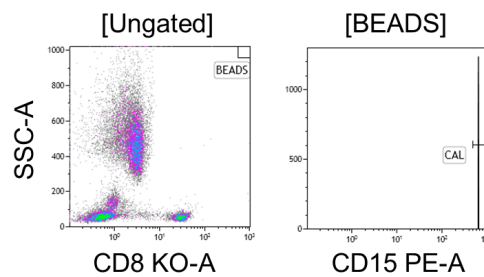
PANEL 1	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	CD16 FITC-A
P8	FL2 INT Log (PE)	CD15 PE-A
P9	FL4 INT Log (PC5.5)	CD56 PC5-5-A
P10	FL5 INT Log (PC7)	CD14 PC7-A
P11	FL6 INT Log (APC)	CD19 APC-A
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A
P13	FL9 INT Log (PB)	CD4 PB-A
P14	FL10 INT Log (Krome Orange)	CD8 KO-A
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
CD16 FITC	CD15 PE	CD56 PC5.5	CD14 PC7	CD19 APC	CD3 APC-AF750	CD4 PB	CD8 KO

- verify for P9 that the mirror is set for 695nm, not 675nm
 - note for P14 which filter is installed (500nm BP, 520 or 540nm BP)
 - set the threshold that allow the detection of the Flow-Count Fluorospheres (# 7547053)
- Report the PMT voltages that have been saved at the step 1 of the Part 1 (table page 4).

PANEL 1	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	CD16 FITC-A	
P8	FL2 INT Log (PE)	CD15 PE-A	
P9	FL4 INT Log (PC5.5)	CD56 PC5-5-A	
P10	FL5 INT Log (PC7)	CD14 PC7-A	
P11	FL6 INT Log (APC)	CD19 APC-A	
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A	
P13	FL9 INT Log (PB)	CD4 PB-A	
P14	FL10 INT Log (Krome Orange)	CD8 KO-A	

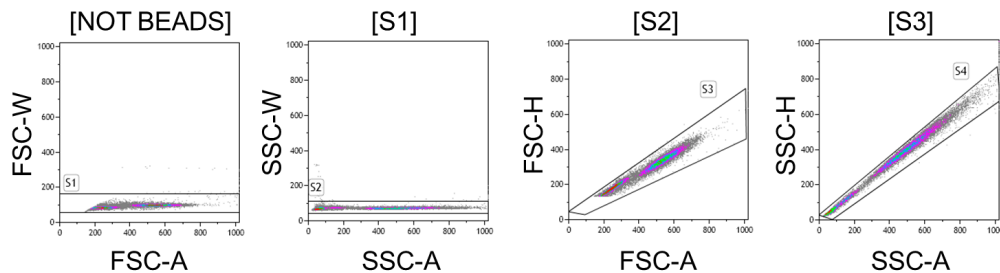
3. Create the following plots to determine the absolute number of cells



- The CD8 KO-A/SSC-A plot is **ungated**. It contains a region called “**BEADS**” corresponding to the Flow-Count Fluorospheres (# 7547053).
- The CD15 plot is **gated on BEADS**. It contains a region called “**CAL**” corresponding to the Flow-Count Fluorospheres (# 7547053)

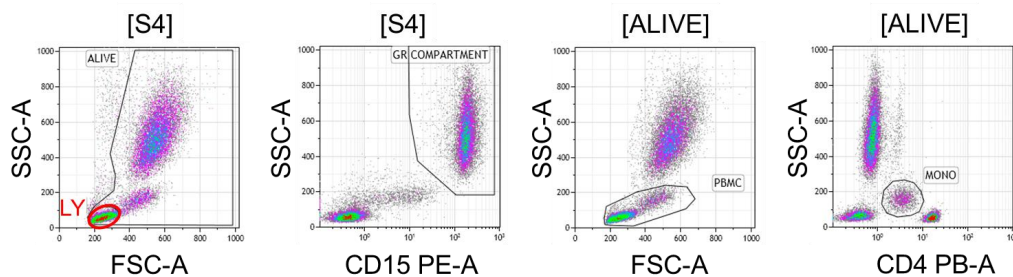
4. Create the following plots for the gating strategy of the PANEL 1 protocol:

a. Create 4 bi-parametric plots to exclude the doublets



- The FSC-A/FSC-W plot is **gated on NOT BEADS** to exclude the Flow Count beads. It contains a region called “**S1**”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “**S2**”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “**S3**”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “**S4**”.

b. Create 4 bi-parametric plots to select the populations of cells.



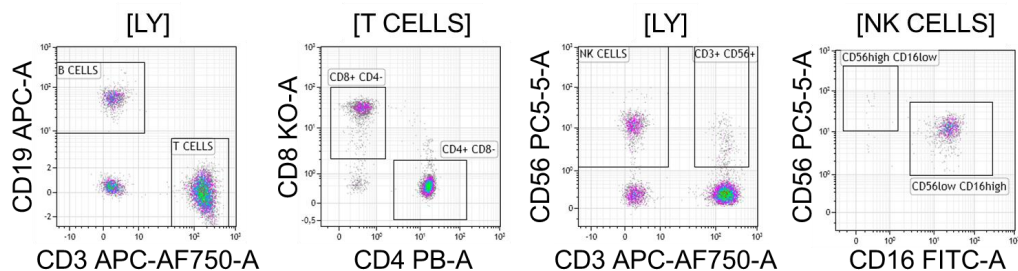
- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called “**ALIVE**” corresponding to the live cells and a region called “**LY**” corresponding to the lymphocytes.
- The CD15 PE-A/SSC-A plot is **gated on S4**. It contains a region called “**GR COMPARTMENT**” corresponding to the granulocytes.
- The FSC-A/SSC-A plot is **gated on ALIVE**. It contains a region called “**PBMC**” corresponding to the lymphocytes and monocytes.
- The CD4 PB-A/SSC-A plot is **gated on ALIVE**. It contains a region called “**MONO**” corresponding to the monocytes.

5. The PANEL 1 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
CD3+ T cells	LY CD3+ CD19-
CD4+ T cells	LY CD3+CD4+CD8-
CD8+ T cells	LY CD3+CD4-CD8+
CD19+ B cells	LY CD3- CD19+
CD3-CD56+ NK cells	LY CD3- CD56+
CD3+CD56+ NK-like T cells	LY CD3+CD56+
CD14+ monocytes	ALIVE CD4low/SS-A-INTlow
CD14highCD16- classical Monocytes	MONO CD4low/SS-A-INTlow CD14highCD16-
CD14+CD16 intermediate Monocytes	MONO CD4low/SS-A-INTlow CD14+CD16int
CD14lowCD16+ non classic Monocytes	MONO CD4low/SS-A-INTlow CD14lowCD16+
CD15+CD14low LDGs	PBMC CD15+CD14low
CD15lowCD16high Neutrophils	GR COMPARTMENT CD15lowCD16high
CD15highCD16- Eosinophils	GR COMPARTMENT CD15highCD16-

6. Create the following plots to identify those populations

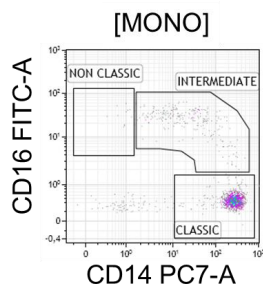
a. For the identification of the lymphocytes populations



- The CD3 APC-AF750-A/CD19 APC-A plot is **gated on LY**. It contains a region called “**B CELLS**” corresponding to the B lymphocytes and a region called “**T CELLS**” corresponding to the T lymphocytes.
- The CD4 PB-A/CD8 KO-A plot is **gated on T CELLS**. It contains a region called “**CD8+ CD4+**” corresponding to the CD8+ T lymphocytes and a region called “**CD4+ CD8-**” corresponding to the CD4+ T lymphocytes.
- The CD3 APC-AF750-A/CD56 PC5-5-A plot is **gated on LY**. It contains a region called “**NK CELLS**” corresponding to the NK cells and a region called “**CD3+ CD56+**” corresponding to the CD3+ NK-like T cells
- The CD16 FITC-A/CD56 PC5-5-A plot is **gated on NK CELLS**. It contains a region called “**CD56high CD16low**” corresponding to the naïve NK cells and a region called “**CD56low CD16high**” corresponding to the active NK cells.

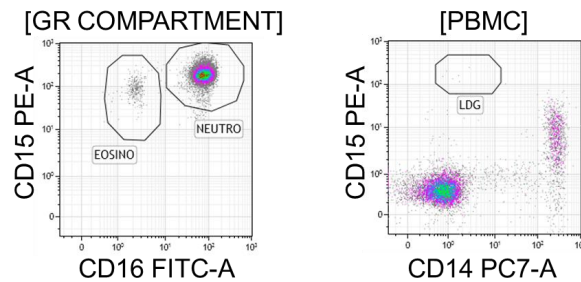
IMPORTANT NOTE FOR PANEL 1: insert a STOP count at 1,000 B cells or at 10 min of acquisition.

b. For the identification of the monocytes populations



- The CD14 PC7-A/CD16 FITC-A plot is **gated on MONO**. It contains 3 regions called “**NON CLASSIC**”, “**INTERMEDIATE**” and “**CLASSIC**” corresponding to the monocytes populations.

c. For the identification of the granulocytes populations



- The CD16 FITC-A/CD15 PE-A plot is **gated on GR COMPARTMENT**. It contains a region called “**EOSINO**”, corresponding to the eosinophils and a region called “**NEUTRO**”, corresponding to the neutrophils
- The CD14 PC7-A/CD15 PE-A plot is **gated on PBMC**. It contains a region called “**LDG**”, corresponding to the Large Density Granulocytes.



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

7. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 1							
	P7	P8	P9	P10	P11	P12	P13	P14	
P7		0,40	0,00	0,10	0,00	0,00	0,00	0,00	
P8	33,00		3,00	2,00	0,00	0,00	0,00	0,00	
P9	1,00	3,00		0,20	0,10	0,00	0,00	0,00	
P10	0,00	1,00	58,04		0,00	0,40	0,00	0,00	
P11	0,00	0,00	5,00	0,00		5,00	0,00	0,00	
P12	0,00	0,00	45,78	26,12	11,00		0,00	0,00	
P13	0,00	0,00	0,00	0,00	0,10	0,00		3,30	
P14	4,00	4,00	1,00	0,10	0,10	0,00	12,05		

As an example: 33% correspond to the FITC (P7) compensation into the PE channel (P8)

$$\text{PE pure} = \text{PE (P8)} - 33\% \text{ of FITC (P7)} - 3\% \text{ of PC5.5 (P9)} - 2\% \text{ of PC7 (P10)}$$

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

B. Create a new protocol named PANEL 2

1. The following parameters have to be selected in the indicated order, with the specific names:

PANEL 2	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	CD1c FITC-A
P8	FL2 INT Log (PE)	Lin PE-A
P9	FL4 INT Log (PC5.5)	CD141 PC5-5-A
P10	FL5 INT Log (PC7)	CD11c PC7-A
P11	FL6 INT Log (APC)	CD123 APC-A
P12	FL8 INT Log (APC-Alexa750)	DRAQ7-A
P13	FL9 INT Log (PB)	HLA DR PB-A
P14	FL10 INT Log (Krome Orange)	
P15	TIME	TIME

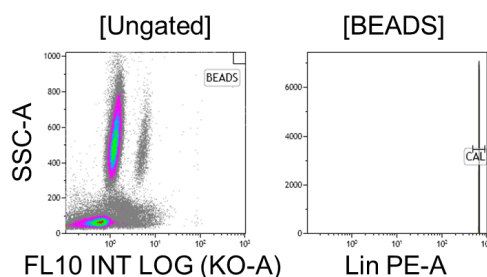
Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
CD1c FITC	Lin PE	CD141 PC5.5	CD11c PC7	CD123 APC	DRAQ7	HLA DR PB	

- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)
- set the threshold that allow the detection of the Flow-Count Fluorospheres (# 7547053)

2. Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 2	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	CD1c FITC-A	
P8	FL2 INT Log (PE)	Lin PE-A	
P9	FL4 INT Log (PC5.5)	CD141 PC5-5-A	
P10	FL5 INT Log (PC7)	CD11c PC7-A	
P11	FL6 INT Log (APC)	CD123 APC-A	
P12	FL8 INT Log (APC-Alexa750)	DRAQ7-A	
P13	FL9 INT Log (PB)	HLA DR PB-A	
P14	FL10 INT Log (Krome Orange)		

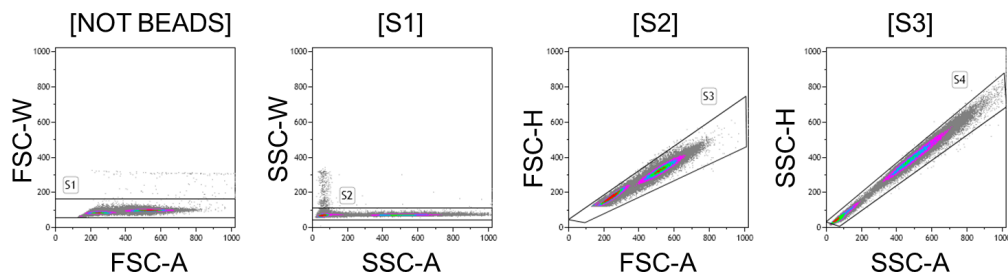
3. Create the following plots to determine the absolute number of cells



- The FL10 INT LOG (KO-A)/SSC-A plot is **ungated**. It contains a region called “**BEADS**” corresponding to the Flow-Count Fluorospheres (# 7547053).
- The Lin PE-A plot is **gated on BEADS**. It contains a region called “**CAL**” corresponding to the Flow-Count Fluorospheres (# 7547053)

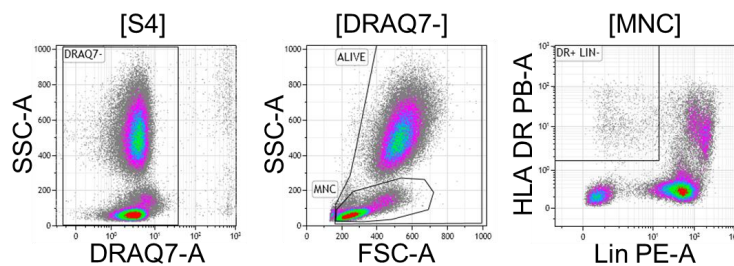
4. Create the following plots for the gating strategy of the PANEL 2 protocol:

a. Create 4 bi-parametric plots to exclude the doublets and the Flow-Count Beads



- The FSC-A/FSC-W plot is **gated on not BEADS to exclude the Flow Count beads**. It contains a region called “**S1**”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “**S2**”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “**S3**”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “**S4**”.

b. Create 3 bi-parametric plots to select the populations of cells



- The DRAQ7-A/SSC-A plot is **gated on S4**. It contains a region called “**DRAQ7-**” corresponding to the live cells.
- The FSC-A/SSC-A plot is **gated on DRAQ7-**. It contains a region called “**ALIVE**” and a region called “**MNC**” containing the basophils.
- The Lin PE-A/HLA DR PB-A plot is **gated on MNC**. It contains a region called “**DR+ LIN-**” containing the dendritic cells.

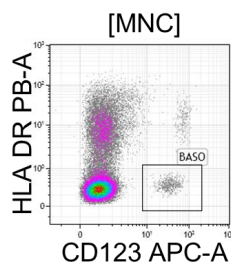
IMPORTANT NOTE FOR PANEL 2: insert a STOP count at 2,000 Lin-DR+ cells or at 17 min of acquisition.

5. The PANEL 2 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
Lin-DR+CD11c-CD123+ pDC	MNC DR+Lin-CD11c-CD123+
Lin-DR+CD11c+CD123-CD1c+ mDC1	MNC DR+Lin-CD11c+CD123-CD1c+
Lin-DR+CD11c+CD123-CD141+ mDC2	MNC DR+Lin-CD11c+CD123-CD141+
Lin-DR-CD123+ Basophils	MNC DR- CD123+

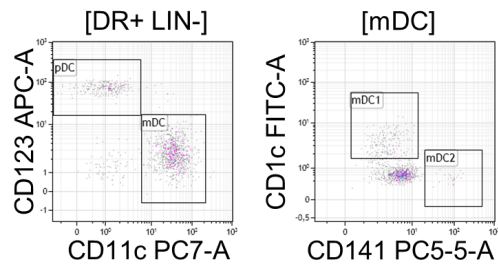
6. Create the following plots to identify those populations

a. For the identification of the basophils



- The CD123 APC-A/HLA DR PB-A plot is **gated on MNC**. It contains a region called “**BASO**” corresponding to the basophils.

b. For the identifications of the dendritic cell populations



- The CD11c PC7-A/CD123 APC-A plot is **gated on DR+ LIN-**. It contains a region called “pDC” and a region called “mDC”.
- The CD141 PC5-5-A/CD1c FITC-A plot is **gated on mDC**. It contains a region called “mDC1” and a region called “mDC2”.



7. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 2							
	P7	P8	P9	P10	P11	P12	P13	P14	
P7		0,50	0,00	0,00	0,00	0,00	0,00	0,00	
P8	30,00		4,00	2,00	0,00	0,00	0,00	0,00	
P9	0,00	3,00		1,00	0,00	0,00	0,00	0,00	
P10	0,00	1,00	56,00		0,00	1,00	0,00	0,00	
P11	0,00	0,00	5,00	0,00		3,00	0,00	0,00	
P12	0,00	0,00	28,00	21,00	10,00		0,00	0,00	
P13	0,00	0,00	0,00	0,00	0,00	0,00		0,00	
P14	0,00	4,00	0,00	0,00	0,00	0,00	0,00		

As an example: 30% correspond to the FITC (P7) compensation into the PE channel (P8)

$$\text{PE pure} = \text{PE (P8)} - 30\% \text{ of FITC (P7)} - 4\% \text{ of PC5.5 (P9)} - 2\% \text{ of PC7 (P10)}$$

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

C. Create a new protocol named PANEL 3

- The following parameters have to be selected in the indicated order, with the specific names:

PANEL 3	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	CD57 FITC-A
P8	FL2 INT Log (PE)	CD45RA PE-A
P9	FL4 INT Log (PC5.5)	CD27 PC5-5-A
P10	FL5 INT Log (PC7)	CD62L PC7-A
P11	FL6 INT Log (APC)	CD38 APC-A
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A
P13	FL9 INT Log (PB)	CD4 PB-A
P14	FL10 INT Log (Krome Orange)	CD8 KO-A
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
CD57 FITC	CD45RA PE	CD27 PC5.5	CD62L PC7	CD38 APC	CD3 APC-AF750	CD4 PB	CD8 KO

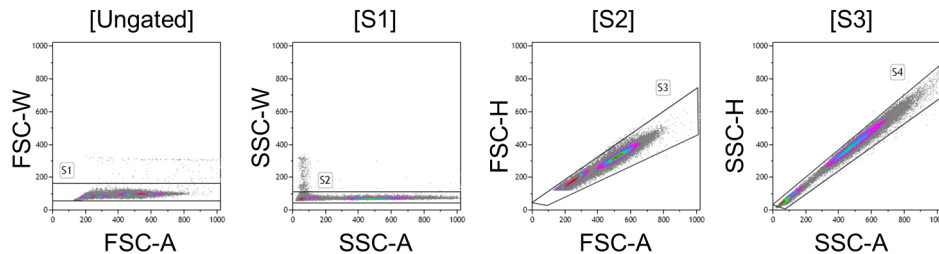
- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)

- Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 3	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	CD57 FITC-A	
P8	FL2 INT Log (PE)	CD45RA PE-A	
P9	FL4 INT Log (PC5.5)	CD27 PC5-5-A	
P10	FL5 INT Log (PC7)	CD62L PC7-A	
P11	FL6 INT Log (APC)	CD38 APC-A	
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A	
P13	FL9 INT Log (PB)	CD4 PB-A	
P14	FL10 INT Log (Krome Orange)	CD8 KO-A	

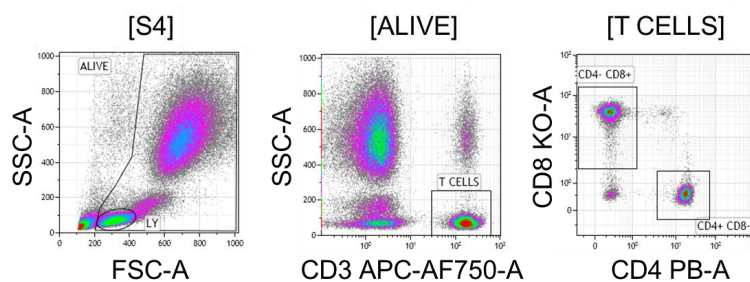
3. Create the following plots for the gating strategy of the PANEL 3 protocol:

a. Create 4 bi-parametric plots to exclude the doublets:



- The FSC-A/FSC-W plot is ungated. It contains a region called "S1".
- The SSC-A/SSC-W plot is gated on S1. It contains a region called "S2".
- The FSC-A/FSC-H plot is gated on S2. It contains a region called "S3".
- The SSC-A/SSC-H plot is gated on S3. It contains a region called "S4".

b. Create 3 bi-parametric plots to select the T lymphocytes:



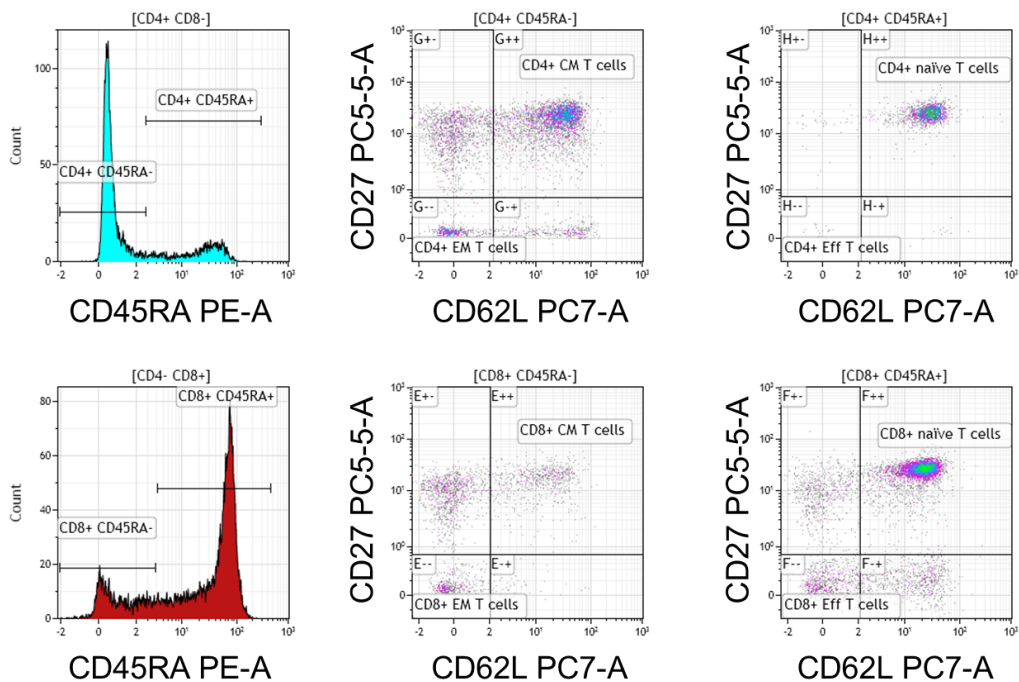
- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called "ALIVE" and a region called "LY" containing the lymphocytes.
- The CD3 APC-AF750-A/SSC-A plot is **gated on ALIVE**. It contains a region called "T CELLS" containing the T lymphocytes.
- The CD4 KO-A/CD8 PB-A plot is **gated on T CELLS**. It contains a region called "CD4- CD8+" a region called "CD4+ CD8-" containing the CD8+ T cells and the CD4+ T cells.

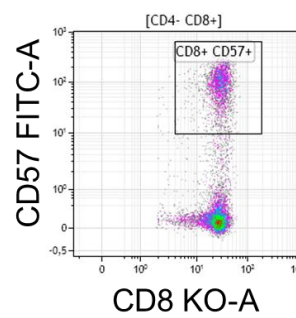
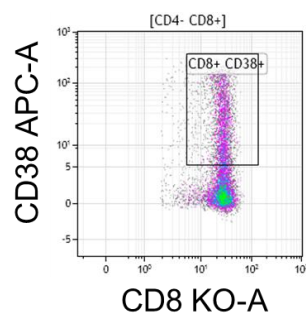
IMPORTANT NOTE FOR PANEL 3: insert a STOP count at 20,000 T cells or at 10 min of acquisition.

4. The PANEL 3 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
Naïve CD45RA+CD62LhighCD27+ CD4+ T cells	LY CD3+ CD4+ CD8- CD45RA+CD62LhighCD27+
Effector CD45RA+CD62LlowCD27- CD4+ T cells	LY CD3+ CD4+ CD8- CD45RA+CD62LlowCD27-
Central Memory CD45RA-CD62LhighCD27+ CD4+ T cells	LY CD3+ CD4+ CD8- CD45RA-CD62LhighCD27+
Effector Memory CD45RA-CD62LlowCD27- CD4+ T cells	LY CD3+ CD4+ CD8- CD45RA-CD62LlowCD27-
Naïve CD45RA+CD62LhighCD27+ CD8+ T cells	LY CD3+ CD4- CD8+ CD45RA+CD62LhighCD27+
Effector CD45RA+CD62LlowCD27- CD8+ T cells	LY CD3+ CD4- CD8+ CD45RA+CD62LlowCD27-
Central Memory CD45RA-CD62LhighCD27+ CD8+ T cells	LY CD3+ CD4- CD8+ CD45RA-CD62LhighCD27+
Effector Memory CD45RA-CD62LlowCD27- CD8+ T cells	LY CD3+ CD4- CD8+ CD45RA-CD62LlowCD27-
CD8+CD38+ Activated T cells	LY CD3+ CD4- CD8+ CD38+
CD8+CD57+ Cytotoxic T cells	LY CD3+ CD4- CD8+ CD57+

5. Create the following plots to identify those populations





6. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 3							
	P7	P8	P9	P10	P11	P12	P13	P14	
P7		0,50	0,00	0,10	0,00	0,00	0,00	0,20	
P8	34,00		4,00	1,40	0,00	0,00	0,00	1,00	
P9	0,70	3,00		0,20	0,10	0,00	0,00	0,00	
P10	0,90	1,00	58,00		0,00	1,00	0,00	0,00	
P11	0,00	0,00	5,00	0,00		4,00	1,00	0,00	
P12	0,00	0,00	28,00	33,69	10,00		0,00	0,00	
P13	0,00	0,00	0,00	0,00	0,00	0,00		2,60	
P14	4,00	4,00	0,00	0,00	0,00	0,00	12,00		

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

D. Create a new protocol named PANEL 4

- The following parameters have to be selected in the indicated order, with the specific names:

PANEL 4	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	gd TCR FITC-A
P8	FL2 INT Log (PE)	ab TCR PE-A
P9	FL4 INT Log (PC5.5)	CD25 PC5-5-A
P10	FL5 INT Log (PC7)	CD127 PC7-A
P11	FL6 INT Log (APC)	CD69 APC-A
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A
P13	FL9 INT Log (PB)	CD4 PB-A
P14	FL10 INT Log (Krome Orange)	CD8 KO-A
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
gd TCR FITC	ab TCR PE	CD25 PC5.5	CD127 PC7	CD69 APC	CD3 APC-AF750	CD4 PB	CD8 KO

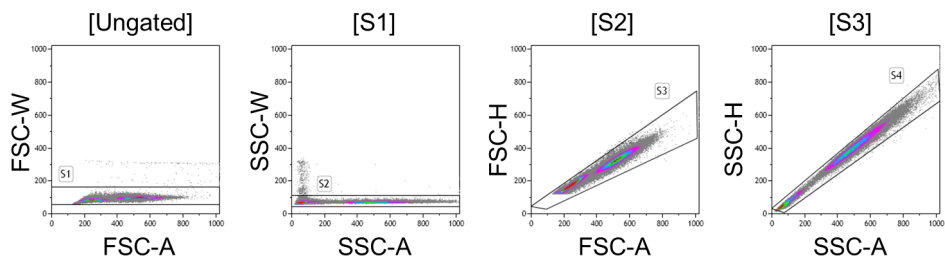
- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)

- Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 4	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	gd TCR FITC-A	
P8	FL2 INT Log (PE)	ab TCR PE-A	
P9	FL4 INT Log (PC5.5)	CD25 PC5-5-A	
P10	FL5 INT Log (PC7)	CD127 PC7-A	
P11	FL6 INT Log (APC)	CD69 APC-A	
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A	
P13	FL9 INT Log (PB)	CD4 PB-A	
P14	FL10 INT Log (Krome Orange)	CD8 KO-A	

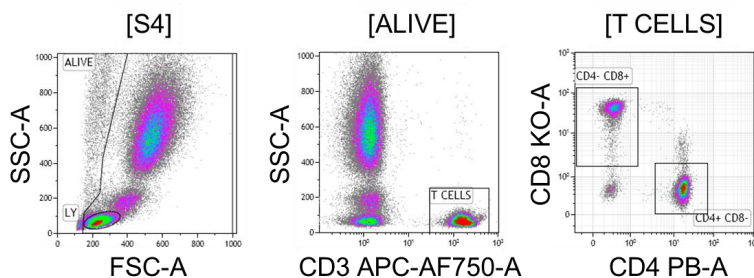
3. Create the following plots for the gating strategy of the PANEL 4 protocol:

a. Create 4 bi-parametric plots to exclude the doublets:



- The FSC-A/FSC-W plot is **ungated**. It contains a region called “S1”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “S2”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “S3”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “S4”.

b. Create 3 bi-parametric plots to select the T lymphocytes:



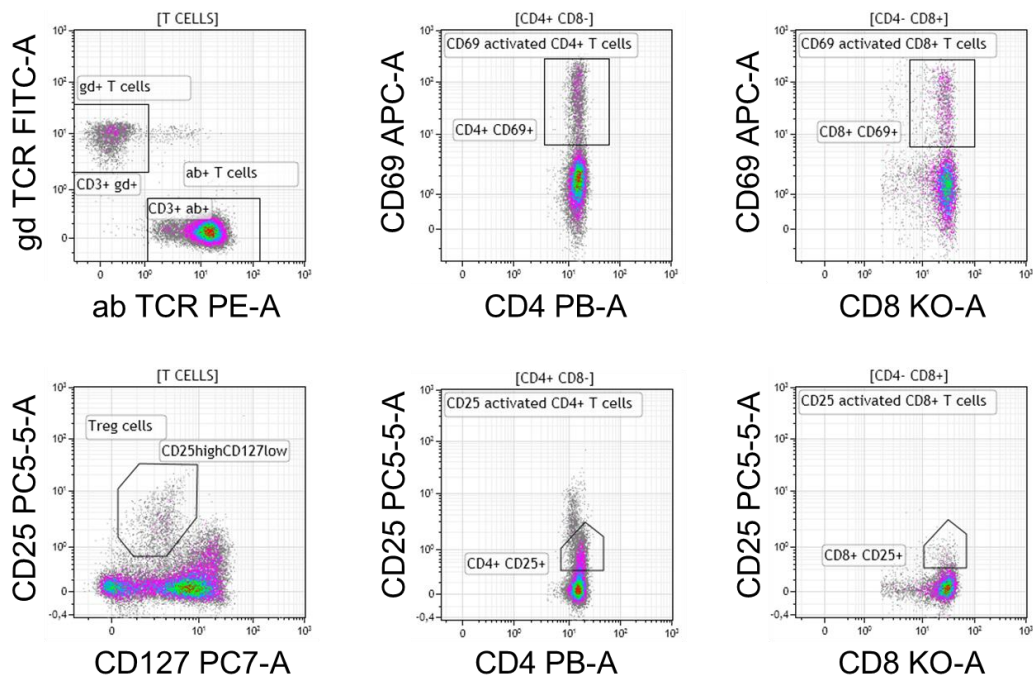
- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called “ALIVE” and a region called “LY” containing the lymphocytes.
- The CD3 APC-AF750-A/SSC-A plot is **gated on ALIVE**. It contains a region called “T CELLS” containing the T lymphocytes.
- The CD4 KO-A/CD8 PB-A plot is **gated on T CELLS**. It contains a region called “CD4- CD8+” a region called “CD4+ CD8-” containing the CD8+ T cells and the CD4+ T cells.

IMPORTANT NOTE FOR PANEL 4: insert a STOP count at 20,000 T cells or at 10 min of acquisition.

4. The PANEL 4 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
ab+ T cells	LY CD3+ ab TCR+
gd+ T cells	LY CD3+ gd TCR+
CD69+ activated CD4+ T cells	LY CD3+ CD4+ CD8- CD69+
CD25+ activated CD4+ T cells	LY CD3+ CD4+ CD8- CD25+
CD69+ activated CD8+ T cells	LY CD3+ CD4- CD8+ CD69+
CD25+ activated CD8+ T cells	LY CD3+ CD4- CD8+ CD25+
CD25 ^{high} CD127 ^{low} CD4+ Regulatory T cells	LY CD3+ CD25 ^{high} CD127 ^{low}

5. Create the following plots to identify those populations





6. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)								PANEL 4
	P7	P8	P9	P10	P11	P12	P13	P14
P7		0,50	0,00	0,00	0,00	0,00	0,00	0,20
P8	33,00		4,00	1,10	0,00	0,00	0,00	0,00
P9	1,00	3,00		0,20	0,10	0,00	0,00	0,00
P10	0,00	1,00	39,93		0,00	1,00	0,00	0,00
P11	0,00	0,00	5,00	0,00		4,70	1,00	0,00
P12	0,00	0,00	28,00	18,00	8,00		0,00	0,00
P13	0,00	0,00	0,00	0,00	0,00	0,00		2,40
P14	4,00	3,00	1,00	0,00	0,00	0,00	13,00	

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

E. Create a new protocol named PANEL 5

1. The following parameters have to be selected in the indicated order, with the specific names:

PANEL 5	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	IgD FITC-A
P8	FL2 INT Log (PE)	CD267 PE-A
P9	FL4 INT Log (PC5.5)	CD5 PC5-5-A
P10	FL5 INT Log (PC7)	CD27 PC7-A
P11	FL6 INT Log (APC)	CD19 APC-A
P12	FL8 INT Log (APC-Alexa750)	CD24 APC-AF750-A
P13	FL9 INT Log (PB)	CD38 PB-A
P14	FL10 INT Log (Krome Orange)	
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
IgD FITC	CD267 PE	CD5 PC5.5	CD27 PC7	CD19 APC	CD24 APC-AF750	CD38 PB	

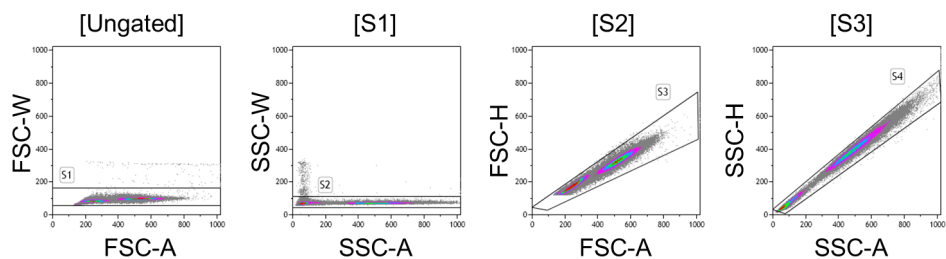
- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)

2. Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 5	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	IgD FITC-A	
P8	FL2 INT Log (PE)	CD267 PE-A	
P9	FL4 INT Log (PC5.5)	CD5 PC5-5-A	
P10	FL5 INT Log (PC7)	CD27 PC7-A	
P11	FL6 INT Log (APC)	CD19 APC-A	
P12	FL8 INT Log (APC-Alexa750)	CD24 APC-AF750-A	
P13	FL9 INT Log (PB)	CD38 PB-A	
P14	FL10 INT Log (Krome Orange)		

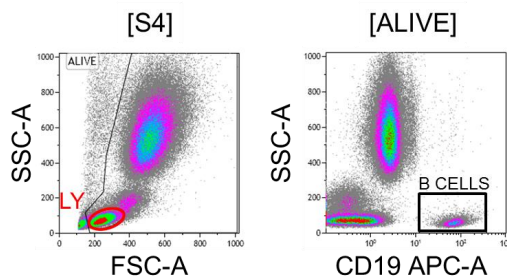
3. Create the following plots for the gating strategy of the PANEL 5 protocol:

a. Create 4 bi-parametric plots to exclude the doublets:



- The FSC-A/FSC-W plot is **ungated**. It contains a region called “S1”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “S2”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “S3”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “S4”.

b. Create 2 bi-parametric plots to select the B lymphocytes:



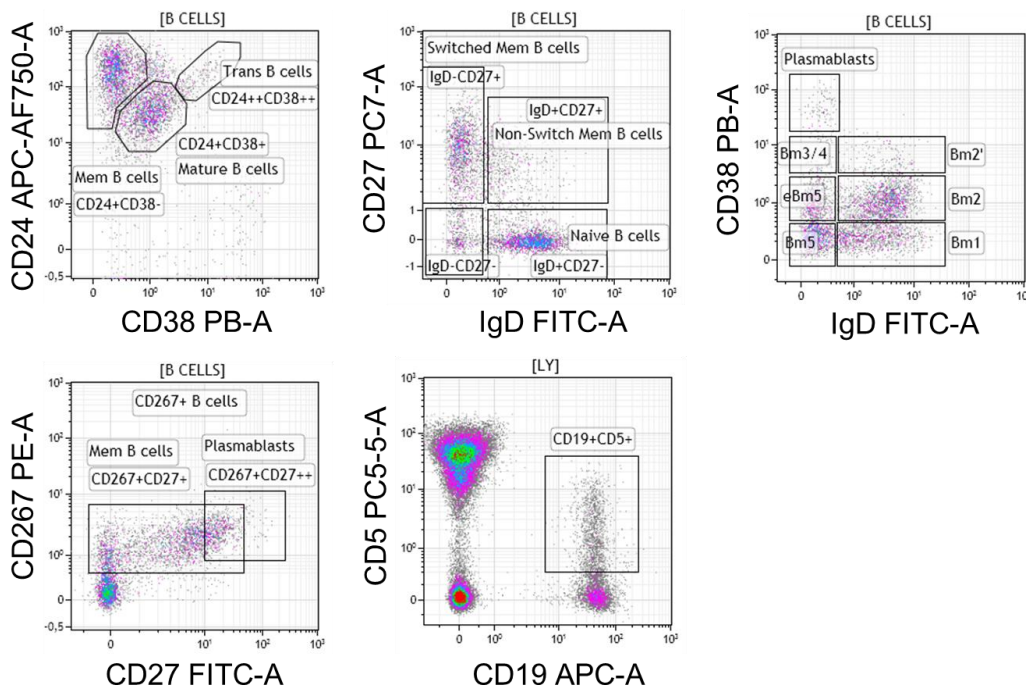
- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called “ALIVE” and a region called “LY” containing the lymphocytes.
- The CD19 APC-A/SSC-A plot is **gated on ALIVE**. It contains a region called “B CELLS” containing the B lymphocytes.

IMPORTANT NOTE FOR PANEL 5: insert a STOP count at 20,000 B cells or at 10 min of acquisition

4. The PANEL 5 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
CD24 ^{high} CD38 ^{high} Transitional B cells	LY CD19+ CD24 ⁺⁺ CD38 ⁺⁺
CD24+CD38+ Mature B cells	LY CD19+ CD24+ CD38+
CD24+CD38- Memory B cells	LY CD19+ CD24+ CD38-
CD27+IgD- switched memory B cells	LY CD9+ IgD- CD27+
CD27+IgD+ Non-switched memory B cells	LY CD19+ IgD+ CD27+
CD27-IgD+ Naive B cells	LY CD19+ IgD+ CD27-
CD27-IgD- Double negative B cells	LY CD19+ IgD- CD27-
IgD+CD38- Bm1	LY CD19+ IgD+ CD38-
IgD+CD38+ Bm2	LY CD19+ IgD+ CD38+
IgD+CD38++ Bm2'	LY CD19+ IgD+ CD38 ⁺⁺
IgD-CD38++ Bm3/4	LY CD19+ IgD- CD38 ⁺⁺
IgD-CD38+ eBm5	LY CD19+ IgD- CD38+
IgD-CD38- Bm5	LY CD19+ IgD- CD38-
IgD-CD38 ^{high} CD267+ Plasmablasts	LY CD19+ IgD-CD38 ⁺⁺⁺ CD267+
CD5+ B cells	LY CD19+ CD5+
IgD+CD267+CD5+CD24+CD27+CD38+ Breg cells	LY CD19+

5. Create the following plots to identify those populations





6. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 5							
	P7	P8	P9	P10	P11	P12	P13	P14	
P7		0,30	0,10	0,10	0,00	0,00	0,00	0,00	
P8	31,00		4,40	2,20	0,00	0,00	0,00	0,00	
P9	0,80	3,70		0,00	0,00	0,00	0,00	0,00	
P10	0,00	1,00	59,00		0,00	0,40	0,00	0,00	
P11	0,00	0,00	5,50	0,00		3,00	0,00	0,00	
P12	0,00	0,00	48,66	18,00	16,00		0,00	0,00	
P13	0,00	0,00	0,00	0,00	0,00	0,00		0,00	
P14	3,10	3,10	0,20	0,00	0,00	0,00	11,83		

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

F. Create a new protocol named PANEL 6

- The following parameters have to be selected in the indicated order, with the specific names:

PANEL 6	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	CD43 FITC-A
P8	FL2 INT Log (PE)	CD69 PE-A
P9	FL4 INT Log (PC5.5)	CD5 PC5-5-A
P10	FL5 INT Log (PC7)	CD27 PC7-A
P11	FL6 INT Log (APC)	CD11b APC-A
P12	FL8 INT Log (APC-Alexa750)	
P13	FL9 INT Log (PB)	CD20 PB-A
P14	FL10 INT Log (Krome Orange)	CD3 KO-A
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
CD43 FITC	CD69 PE	CD5 PC5.5	CD27 PC7	CD11b APC		CD20 PB	CD3 KO

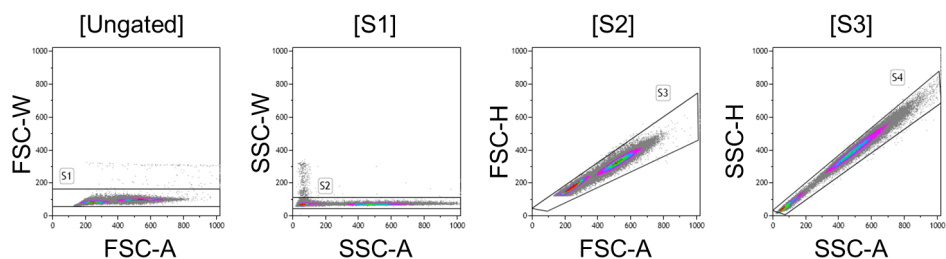
- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)

- Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 6	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	CD43 FITC-A	
P8	FL2 INT Log (PE)	CD69 PE-A	
P9	FL4 INT Log (PC5.5)	CD5 PC5-5-A	
P10	FL5 INT Log (PC7)	CD27 PC7-A	
P11	FL6 INT Log (APC)	CD11b APC-A	
P12	FL8 INT Log (APC-Alexa750)		
P13	FL9 INT Log (PB)	CD20 PB-A	
P14	FL10 INT Log (Krome Orange)	CD3 KO-A	

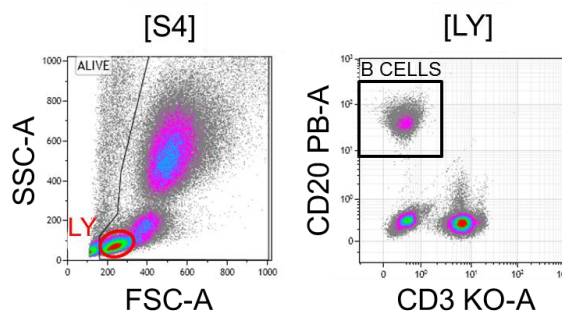
3. Create the following plots for the gating strategy of the PANEL 6 protocol:

c. Create 4 bi-parametric plots to exclude the doublets:



- The FSC-A/FSC-W plot is **ungated**. It contains a region called “S1”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “S2”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “S3”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “S4”.

d. Create 2 bi-parametric plots to select the B lymphocytes:



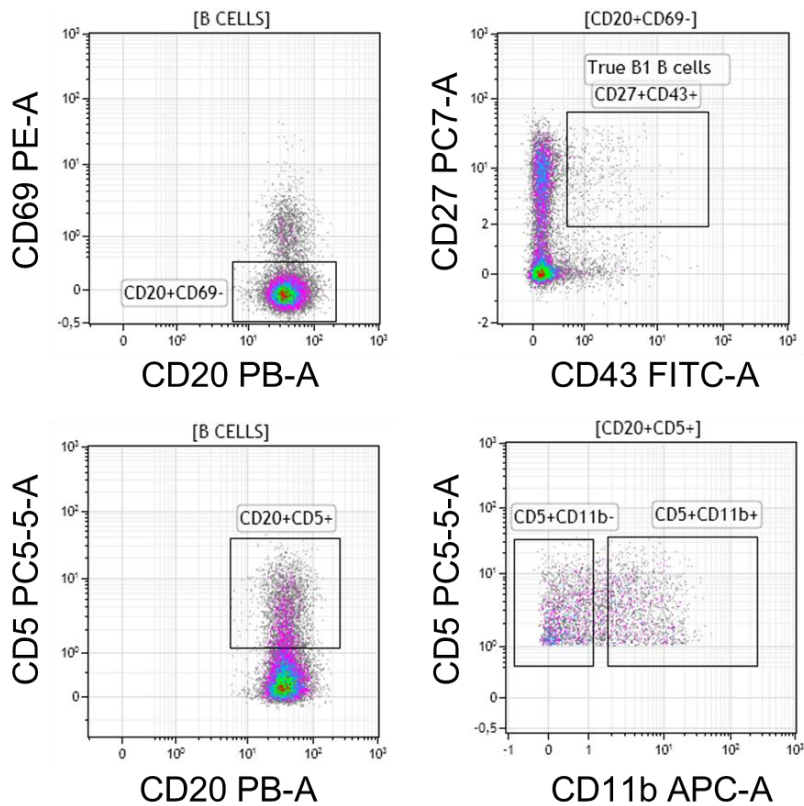
- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called “ALIVE” and a region called “LY” containing the lymphocytes.
- The CD3 KO-A/CD20-A PB plot is **gated on LY**. It contains a region called “B CELLS” containing the B lymphocytes.

IMPORTANT NOTE FOR PANEL 6: insert a STOP count at 20,000 B cells or at 10 min of acquisition

4. The PANEL 6 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
CD5+ B cells	LY CD20+ CD5+
B1 B cells	LY CD20+ CD3- CD69- CD27+ CD43+
CD5+CD11b- B cells	LY CD20+ CD5+ CD11b-
CD5+CD11b+ B cells	LY CD20+ CD5+ CD11b+

5. Create the following plots to identify those populations





6. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 6							
	P7	P8	P9	P10	P11	P12	P13	P14	
P7		1,00	0,00	0,10	0,00	0,00	0,10	0,00	
P8	33,50		4,40	2,60	0,10	0,00	0,00	0,00	
P9	0,70	2,80		0,00	0,00	0,40	0,00	0,00	
P10	0,00	0,60	58,18		0,00	0,00	0,00	0,00	
P11	0,00	0,00	4,00	0,00		0,00	0,00	0,00	
P12	0,00	0,00	47,96	22,16	9,03		0,10	0,00	
P13	0,00	0,00	0,10	0,00	0,00	0,00		2,00	
P14	3,50	4,00	0,00	0,00	0,10	0,00	11,60		

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

G. Create a new protocol named PANEL 7

1. The following parameters have to be selected in the indicated order, with the specific names:

PANEL 7	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	CD64 FITC-A
P8	FL2 INT Log (PE)	CD32 PE-A
P9	FL4 INT Log (PC5.5)	CD18 PC5-5-A
P10	FL5 INT Log (PC7)	CD14 PC7-A
P11	FL6 INT Log (APC)	CD19 APC-A
P12	FL8 INT Log (APC-Alexa750)	CD11b APC-AF750-A
P13	FL9 INT Log (PB)	
P14	FL10 INT Log (Krome Orange)	CD16 KO-A
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
CD64 FITC	CD32 PE	CD18 PC5.5	CD14 PC7	CD19 APC	CD11b APC-AF750		CD16 KO

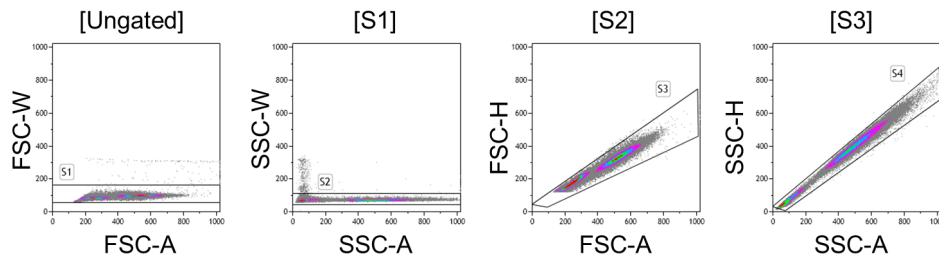
- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)

2. Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 7	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	CD64 FITC-A	
P8	FL2 INT Log (PE)	CD32 PE-A	
P9	FL4 INT Log (PC5.5)	CD18 PC5-5-A	
P10	FL5 INT Log (PC7)	CD14 PC7-A	
P11	FL6 INT Log (APC)	CD19 APC-A	
P12	FL8 INT Log (APC-Alexa750)	CD11b APC-AF750-A	
P13	FL9 INT Log (PB)		
P14	FL10 INT Log (Krome Orange)	CD16 KO-A	

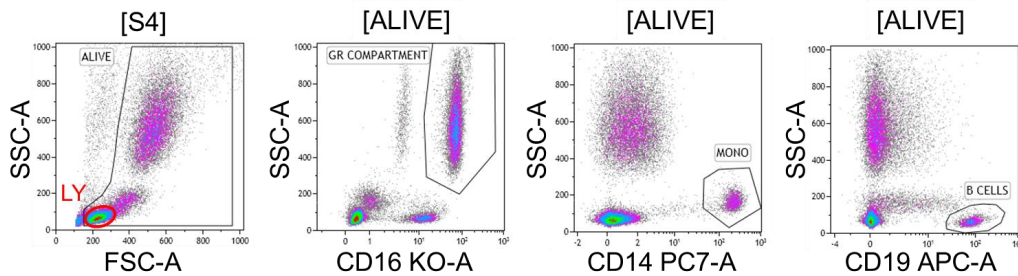
3. Create the following plots for the gating strategy of the PANEL 7 protocol:

a. Create 4 bi-parametric plots to exclude the doublets:



- The FSC-A/FSC-W plot is **ungated**. It contains a region called “S1”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “S2”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “S3”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “S4”.

b. Create 4 bi-parametric plots to select the different populations:



- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called “ALIVE” and a region called “LY” containing the lymphocytes.
- The CD16 KO-A/SSC-A plot is **gated on ALIVE**. It contains a region called “GR COMPARTMENT” containing the PMN.
- The CD14 PC7-A/SSC-A plot is **gated on ALIVE**. It contains a region called “MONO” containing the monocytes.

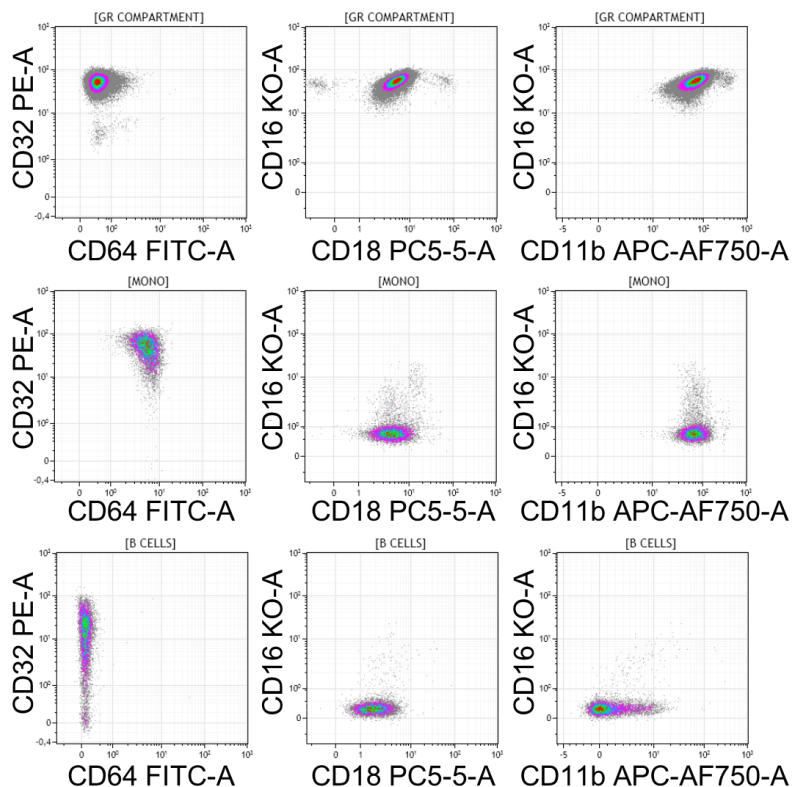
- CD19 APC-A/SSC-A plot is **gated on ALIVE**. It contains a region called "**B CELLS**" containing the B lymphocytes.

IMPORTANT NOTE FOR PANEL 7: insert a STOP count at 10,000 B cells or at 5 min of acquisition

4. The PANEL 7 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
CD16, CD32, CD64, CD11b, CD18 positivities on Granulocytes	GR COMPARTMENT
CD16, CD32, CD64, CD11b, CD18 positivities on Monocytes	MONO
CD16, CD32, CD64, CD11b, CD18 positivities on B cells	B CELLS

5. Create the following plots to identify those populations





6. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 7							
	P7	P8	P9	P10	P11	P12	P13	P14	
P7		0,60	0,00	0,00	0,00	0,00	0,00	0,00	
P8	30,00		4,00	5,00	0,00	0,10	0,00	0,10	
P9	0,00	2,90		2,00	0,00	3,00	0,00	0,00	
P10	0,00	1,20	52,00		0,00	2,60	0,00	0,00	
P11	0,00	0,00	5,00	0,00		7,00	0,00	0,00	
P12	0,00	0,00	28,00	25,00	10,00		0,00	0,00	
P13	0,00	0,20	0,00	0,00	0,00	0,00		2,00	
P14	4,00	4,60	1,00	0,00	0,00	0,00	0,00		

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

H. Create a new protocol named PANEL 8

1. The following parameters have to be selected in the indicated order, with the specific names:

PANEL 8	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	CD35 FITC-A
P8	FL2 INT Log (PE)	CD59 PE-A
P9	FL4 INT Log (PC5.5)	
P10	FL5 INT Log (PC7)	CD14 PC7-A
P11	FL6 INT Log (APC)	CD19 APC-A
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A
P13	FL9 INT Log (PB)	CD21 PB-A
P14	FL10 INT Log (Krome Orange)	CD16 KO-A
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
CD35 FITC	CD59 PE		CD14 PC7	CD19 APC	CD3 APC-AF750	CD21 PB	CD16 KO

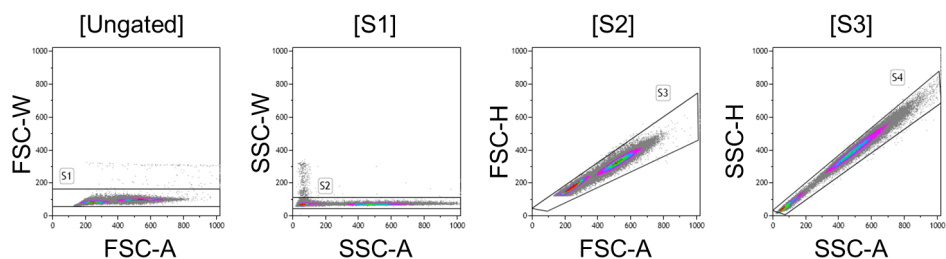
- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)

2. Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 8	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	CD35 FITC-A	
P8	FL2 INT Log (PE)	CD59 PE-A	
P9	FL4 INT Log (PC5.5)		
P10	FL5 INT Log (PC7)	CD14 PC7-A	
P11	FL6 INT Log (APC)	CD19 APC-A	
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A	
P13	FL9 INT Log (PB)	CD21 PB-A	
P14	FL10 INT Log (Krome Orange)	CD16 KO-A	

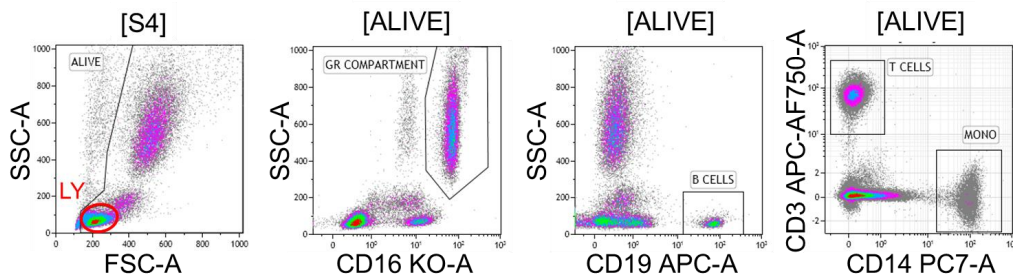
3. Create the following plots for the gating strategy of the PANEL 8 protocol:

a. Create 4 bi-parametric plots to exclude the doublets:



- The FSC-A/FSC-W plot is **ungated**. It contains a region called “S1”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “S2”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “S3”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “S4”.

b. Create 4 bi-parametric plots to select the B lymphocytes:



- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called “ALIVE” and a region called “LY” containing the lymphocytes.
- The CD16 KO-A/SSC-A plot is **gated on ALIVE**. It contains a region called “GR COMPARTMENT” containing the PMN.

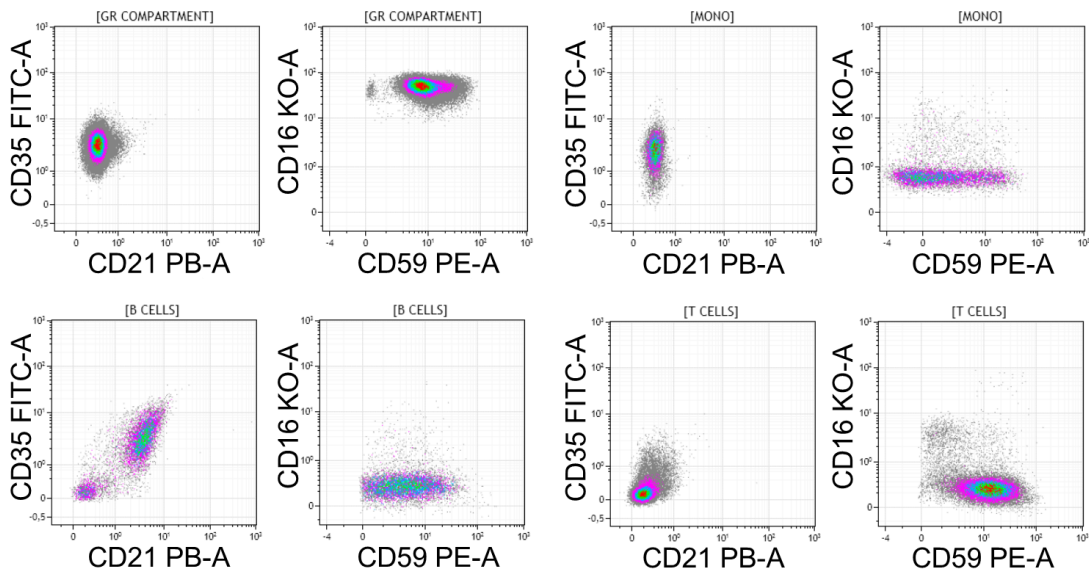
- The CD19 APC-A/SSC-A plot is **gated on ALIVE**. It contains a region called **“B CELLS”** containing the B lymphocytes.
- CD14 PC7-A/CD3 APC-AF750-A plot is **gated on ALIVE**. It contains a region called **“T CELLS”** containing the T lymphocytes and a region called **“MONO”** containing the monocytes.

IMPORTANT NOTE FOR PANEL 8: insert a STOP count at 10,000 B cells or at 5 min of acquisition

4. The PANEL 8 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
CD21, CD35, CD59 positivities on Granulocytes	GR COMPARTMENT
CD21, CD35, CD59 positivities on Monocytes	MONO
CD21, CD35, CD59 positivities on B cells	B CELLS
CD21, CD35, CD59 positivities on T cells	T CELLS

5. Create the following plots to identify those populations





6. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 8							
	P7	P8	P9	P10	P11	P12	P13	P14	
P7		0,60	0,00	0,00	0,00	0,00	0,00	0,00	
P8	31,00		0,00	1,00	0,00	0,00	0,00	1,00	
P9	0,50	2,90		0,10	0,10	0,00	0,00	0,10	
P10	0,00	1,00	0,00		0,00	0,40	0,00	0,00	
P11	0,00	0,00	0,00	0,00		4,76	1,00	0,00	
P12	0,00	0,00	0,00	28,67	9,80		0,00	0,00	
P13	0,00	0,00	0,00	0,00	0,00	0,00		2,20	
P14	3,60	4,10	0,00	0,00	0,00	0,00	13,10		

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

I. Create a new protocol named PANEL 9

1. The following parameters have to be selected in the indicated order, with the specific names:

PANEL 9	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	CD46 FITC-A
P8	FL2 INT Log (PE)	CD55 PE-A
P9	FL4 INT Log (PC5.5)	
P10	FL5 INT Log (PC7)	CD14 PC7-A
P11	FL6 INT Log (APC)	CD19 APC-A
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A
P13	FL9 INT Log (PB)	
P14	FL10 INT Log (Krome Orange)	CD16 KO-A
P15	TIME	TIME

Blue laser : 488 nm				Red laser : 638 nm		Violet laser : 405 nm	
P7	P8	P9	P10	P11	P12	P13	P14
CD46 FITC	CD55 PE		CD14 PC7	CD19 APC	CD3 APC-AF750		CD16 KO

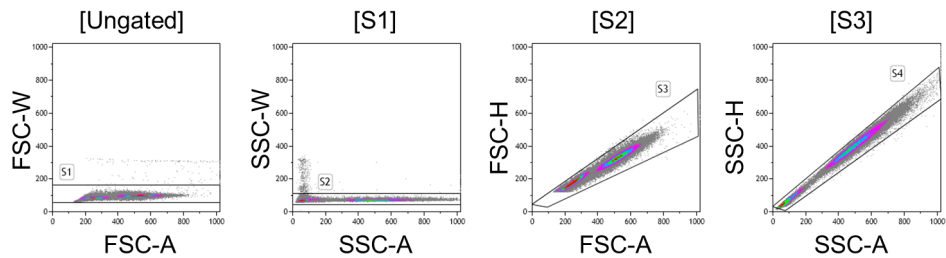
- verify for P9 that the mirror is set for 695nm, not 675nm
- note for P14 which filter is installed (500nm BP, 520 or 540nm BP)

2. Report the PMT voltages that have been saved at the step 1 of the Part 1 (page 4).

PANEL 9	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	CD46 FITC-A	
P8	FL2 INT Log (PE)	CD55 PE-A	
P9	FL4 INT Log (PC5.5)		
P10	FL5 INT Log (PC7)	CD14 PC7-A	
P11	FL6 INT Log (APC)	CD19 APC-A	
P12	FL8 INT Log (APC-Alexa750)	CD3 APC-AF750-A	
P13	FL9 INT Log (PB)		
P14	FL10 INT Log (Krome Orange)	CD16 KO-A	

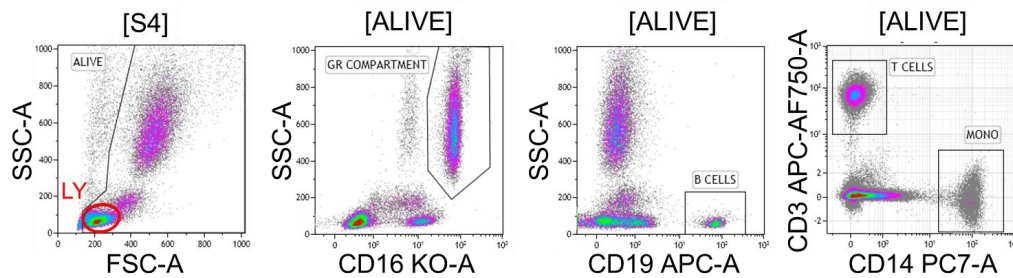
3. Create the following plots for the gating strategy of the PANEL 9 protocol:

a. Create 4 bi-parametric plots to exclude the doublets:



- The FSC-A/FSC-W plot is **ungated**. It contains a region called “**S1**”.
- The SSC-A/SSC-W plot is **gated on S1**. It contains a region called “**S2**”.
- The FSC-A/FSC-H plot is **gated on S2**. It contains a region called “**S3**”.
- The SSC-A/SSC-H plot is **gated on S3**. It contains a region called “**S4**”.

b. Create 4 bi-parametric plots to select the B lymphocytes:



- The FSC-A/SSC-A plot is **gated on S4**. It contains a region called “**ALIVE**” and a region called “**LY**” containing the lymphocytes.
- The CD16 KO-A/SSC-A plot is **gated on ALIVE**. It contains a region called “**GR COMPARTMENT**” containing the PMN.



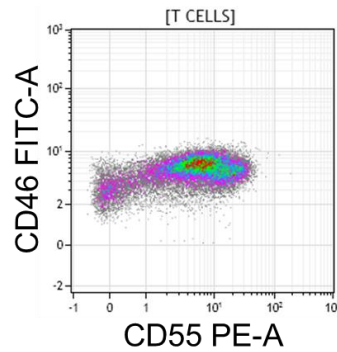
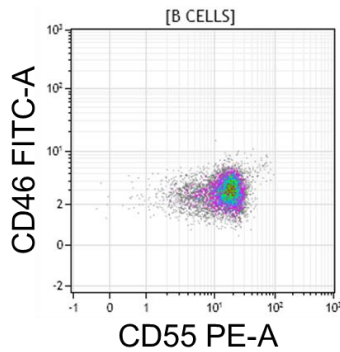
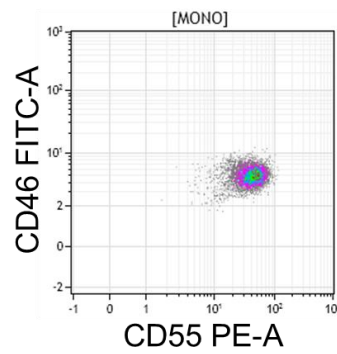
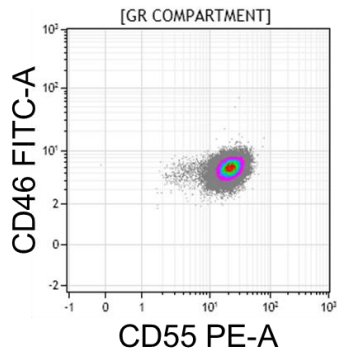
- The CD19 APC-A/SSC-A plot is **gated on ALIVE**. It contains a region called **“B CELLS”** containing the B lymphocytes.
- CD14 PC7-A/CD3 APC-AF750-A plot is **gated on ALIVE**. It contains a region called **“T CELLS”** containing the T lymphocytes and a region called **“MONO”** containing the monocytes.

IMPORTANT NOTE FOR PANEL 9: insert a STOP count at 10,000 B cells or at 5 min of acquisition

4. The PANEL 9 is dedicated to the analyses of the following populations (but not limited to):

POPULATION	GATING STRATEGY
CD21, CD35, CD59 positivities on Granulocytes	GR COMPARTMENT
CD21, CD35, CD59 positivities on Monocytes	MONO
CD21, CD35, CD59 positivities on B cells	B CELLS
CD21, CD35, CD59 positivities on T cells	T CELLS

5. Create the following plots to identify those populations





6. The following compensation matrix can be used as an indication for appropriate compensation

Spillover (%)		PANEL 9						
	P7	P8	P9	P10	P11	P12	P13	P14
P7		1,00	0,00	0,00	0,00	0,00	0,00	0,00
P8	31,00		0,00	1,00	0,00	0,00	0,00	1,00
P9	1,00	2,80		0,10	0,00	0,00	0,00	0,00
P10	0,00	1,00	0,00		0,00	0,40	0,00	0,00
P11	0,00	0,00	0,00	0,00		4,70	0,00	0,00
P12	0,00	0,00	0,00	27,85	9,70		0,00	0,00
P13	0,00	0,00	0,00	0,00	0,00	0,00		2,00
P14	4,00	4,00	0,00	0,00	0,10	0,00	0,00	

Note: these values are given as indications. They have been determined on Navios. However, depending on your BC flow cytometer (Navios or Gallios), but also between Navios cytometers, the values may vary.

AT THE END: QUIT YOUR SPECIFIC PRECISESADS CONFIGURATION



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

If you have any issue with this procedure you can contact:

UBO team: Christophe JAMIN christophe.jamin@univ-brest.fr

or Jacques-Olivier PERS pers@univ-brest.fr

and

Beckman Coulter reference team: Olivier JAEN ojaen@beckman.com

or Nicole WEIT nweit@beckman.com

SOP Acquisition des panels



innovative
medicines
initiative



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

Work Package 5

Standard Operating Procedure #3

Daily Flow Cytometry acquisition

for Beckman Coulter users



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases



Daily Procedure for Multi-Color Duraclone Precisesads project

PURPOSE

The purpose of this document is to provide instructions and procedures required for the generation of data during the acquisition of Precisesads cytometry panels in whole blood samples using a Navios and Gallios Beckman Coulter Flow Cytometers.

PREREQUISITES

- Use the **specific configuration** in your flow cytometer that has been specifically created for and dedicated to the PRECISESADS project.

Thus, specific denominations are assigned to the parameters which are different from the usual configuration (see SOP#1 and SOP#2):

	PARAMETERS	SPECIFIC NAME TO ASSIGN
P1	FSC INT Lin	FSC-A
P2	SSC INT Lin	SSC-A
P3	FS Peak Lin	FSC-H
P4	FS TOF Lin	FSC-W
P5	SS Peak Lin	SSC-H
P6	SS TOF Lin	SSC-W
P7	FL1 INT Log (FITC)	FITC-A
P8	FL2 INT Log (PE)	PE-A
P9	FL4 INT Log (PC5.5)	PC5-5-A
P10	FL5 INT Log (PC7)	PC7-A
P11	FL6 INT Log (APC)	APC-A
P12	FL8 INT Log (APC-Alexa750)	APC-AF750-A
P13	FL9 INT Log (PB)	PB-A
P14	FL10 INT Log (Krome Orange)	KO-A
P15	TIME	TIME

These names will be the same in all centers and all cytometers (BD and BC flow cytometers) to enable the comparisons of the data during the whole project.

- Copy/paste the QC 8 Peaks beads and the 9 protocols to use the proper settings obtained with 8 Peaks beads (refer to SOP#1).
- Get the proper compensation matrices (refer to SOP#2)



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

SUMMARY

- **Reagents required** **page 4**

- **Part 1: Validation using the QC 8 Peaks Beads protocol** **page 5**

- **Part 2: Preparation of Whole Blood samples** **page 7**

- **Part 3: Acquisition of Whole Blood samples** **page 10**

- **Annex 1: Full addresses of the participants** **page 13**



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

TO START: OPEN YOUR SPECIFIC PRECISESADS CONFIGURATION

Part 1: Validation using the QC 8 Peaks Beads protocol

Please refer to the SOP#1 for the preparation of 8 Peaks Beads protocol

1. Prepare the 8 Peaks beads

- a. In a cytometer tube, put 500µl of purified water (ref# BO0209C)
- b. Add 1 droplet of 8 Peaks Beads (ref# A79016, lot AF01)
- c. Vortex

Note: The diluted beads should be prepared immediately before run, and can be kept no more than 1 hour at room temperature in the dark.

2. Load the QC 8 Peaks protocol

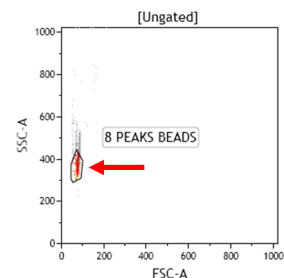
- a. Label your LMD File as follows:

QC8PeaksBeads_patient ID_INSTITUTION_INSTRUMENT_ddMMMyyyy of reception_ddMMMyyyy of acquisition

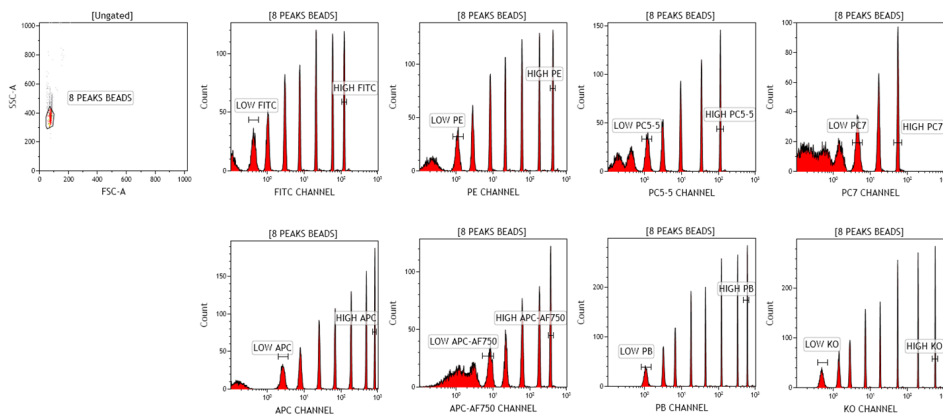
- i. With the name of the protocol = QC8PeaksBeads
- ii. Sample ID1 = patient ID
- iii. Sample ID2 = INSTITUTION_INSTRUMENT
- iv. Sample ID3 = ddMMMyyyy of reception
- v. Sample ID4 = ddMMMyyyy of acquisition

Example: QC8PeaksBeads_32140001_UBO_Navios_22SEP2014_22SEP2014

- b. Run the tube containing the diluted 8 peaks beads at **low speed**
- c. If necessary, adjust the threshold to detect the beads.
- d. Adjust the **8 PEAKS BEADS** region in the FSC-A/SSC-A plot to capture the Beads and exclude any doublets (red arrow):



e. Adjust Fluorescence's PMT voltages to insert the peaks in the targets



f. Acquire 5.000 beads.

3. Save the protocol for your laboratory with the newly determined voltages, if any.

These PMT must be specifically used for the PRECISESADS project; otherwise the data cannot be compared between sites.

Indicate the new values, if any, which have been determined in the following table:

	PARAMETERS	SPECIFIC NAMES	New values of PMT voltages
P1	FSC INT Lin	FSC-A	
P2	SSC INT Lin	SSC-A	
P7	FL1 INT Log (FITC)	FITC-A	
P8	FL2 INT Log (PE)	PE-A	
P9	FL4 INT Log (PC5.5)	PC5-5-A	
P10	FL5 INT Log (PC7)	PC7-A	
P11	FL6 INT Log (APC)	APC-A	
P12	FL8 INT Log (APC-Alexa750)	APC-AF750-A	
P13	FL9 INT Log (PB)	PB-A	
P14	FL10 INT Log (Krome Orange)	KO-A	

These voltages will be loaded in the panel's protocols for the analyses of the blood samples (see Part 3, the step 1, page 10).

4. You are now ready to perform the blood sample preparation (Part 2) and analyses (Part 3).

Note: Step 1 to 4 should be repeated one time per day before the acquisition of blood samples.

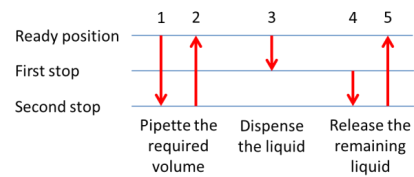
Part 2: Preparation of Whole Blood samples

PREREQUISITE

- The temperature stability should be 22-23°C, and never < 20°C nor > 25°C.
- The compensation matrices are already set inside the 9 protocols (refer to SOP#2).
- However, readjustments could be done after the first acquisition.

NOTE

- **Reverse pipetting means:**
- **Keep the blood samples at Room Temperature before staining.**
- Maximum time required from the preparation to the acquisition:
 - Panel 1 and Panel 2: 1 hour and 30 min
 - Panel 1 to Panel 9: 2 hours and 15 min
- Description of the 9 panels:



	Blood (µL)	LASER 1 488 nm	LASER 2 638 nm	LASER 3 405 nm	IOTest 3 Lyse + Fix (mL)	Flow-Count fluorospheres (µL)					
		FITC	PE	PC5.5	PC7	APC	APC AF750	PB	KO		
Panel 1 whole peripheral blood sub-populations	50	CD16	CD15	CD56	CD14	CD19	CD3	CD4	CD8	2	50
Panel 2 whole peripheral blood sub-populations	100	CD1c	Lin	CD141	CD11c	CD123	DRAQ7	HLA DR		2	100
Panel 3 T cell subsets #1	50	CD57	CD45RA	CD27	CD62L	CD38	CD3	CD4	CD8	2	0
Panel 4 T cell subsets #2	50	TCR g-d	TCR a-b	CD25	CD127	CD69	CD3	CD4	CD8	2	0
Panel 5 B cell subsets #1	100	IgD	CD267	CD5	CD27	CD19	CD24	CD38		2	0
Panel 6 B cell subset #2	100	CD43	CD69	CD5	CD27	CD11b		CD20	CD3	2	0
Panel 7 Neutrophils and Fc gamma Receptors	50	CD64	CD32	CD18	CD14	CD19	CD11b		CD16	2	0
Panel 8 Complement family proteins #1	50	CD35	CD59		CD14	CD19	CD3	CD21	CD16	2	0
Panel 9 Complement family proteins #2	50	CD46	CD55		CD14	CD19	CD3		CD16	2	0



innovative
medicines
initiative



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

1. Prepare the IOTest 3 lysing 1X from IOTest 3 lysing 10X :
 - a. For the preparation of Panel 1 and Panel 2
 - i. In a 15mL Falcon tube, put 400 μ L of IOTest 3 Lysing Solution 10X (ref# A07799)
 - ii. Add 3600 μ L of Distilled water to complete to 4 mL
 - iii. Vortex
 - b. For the preparation of full Panels (Panel 1 to Panel 9)
 - i. In a 50mL Falcon tube, put 2 mL of IOTest 3 Lysing Solution 10X (ref# A07799)
 - ii. Add 18 mL of Distilled water to complete to 20 mL
 - iii. Vortex
2. Include the IOTest 3 fixative solution into the IOTest3 lysing 1X (IOTest 3 lyse 1X + fix)
 - a. For the preparation of Panel 1 and Panel 2
 - i. Add 100 μ L of IOTest 3 Fixative Solution (ref# A07800) to the 4 mL of the IOTest3 lysing 1X (IOTest 3 lyse 1X + fix)
 - ii. Vortex
 - b. For the preparation of full Panels (Panel 1 to Panel 9)
 - i. Add 500 μ L of IOTest 3 Fixative Solution (ref# A07800) to the 20 mL of the IOTest3 lysing 1X (IOTest 3 lyse 1X + fix)
 - ii. Vortex
3. **in Panel 2:** Add 2.5 μ L of DRAQ7 (ref# B25595)
4. Drop **by reverse pipetting** the blood samples (either 50 μ L or 100 μ L) as indicated in the above table into the corresponding Duraclone tubes
5. **Vortex 10 sec** to rehydrate the antibodies.
6. Incubate all tubes for 20 min at room temperature in the dark.



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

7. Add 2 mL of IOTest 3 lysing 1X + fix at room temperature.
8. Vortex 2 sec.
9. Incubate all tubes for 20 min at room temperature in the dark.
10. For panel 1 and 2 : add the FlowCount fluorospheres (ref# 7547053)
 - a. **Vortex the vial 5 sec and shake it 5 times by inversion.**
 - b. Add **by reverse pipetting** 50 μ L into panel 1 and 100 μ L into panel 2 (same volume as blood with the same pipet).
 - c. **Vortex 2 sec.**
 - d. Panels are ready for acquisition.
11. For panel 3 to 9 :
 - a. Centrifuge at 400g for 5 min at room temperature.
 - b. Discard the supernatant by returning the tube, and keep the tube upside down.
 - c. Carefully absorb the liquid on paper towels to eliminate all the droplets.
 - d. Add 4 mL of PBS, put the cap, and mix gently by reversing the tubes 4 times.
 - e. Centrifuge at 400g for 5 min at room temperature.
 - f. Discard the supernatant by returning the tube, and keep the tube upside down.
 - g. Carefully absorb the liquid on paper towels to eliminate all the droplets.
 - h. Dispense 500 μ L of PBS.
 - i. Vortex 2 sec.
 - j. Panels are ready for acquisition

Note: the acquisitions should be performed **within 2 hours** after preparation. If the flow cytometer is not immediately available, the tubes should be kept **at 4°C in the dark** until acquisition.



PRECISESADS

Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases

Part 3: Acquisition of Whole Blood samples

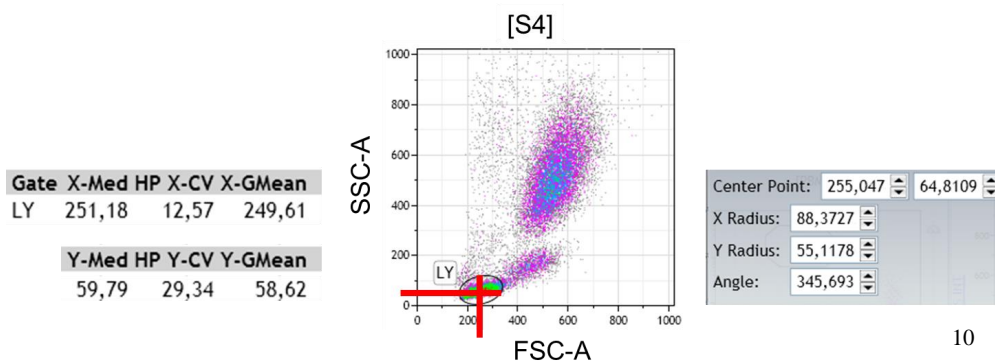
1. Load all the protocols for the corresponding panels.
 - a. Please refer to the SOP#2 for the preparation of the different protocols.
 - b. Set the correct PMT values if they have been modified after the 8 Peaks Beads acquisition.
 - c. Label your LMD File as follows:

Panel x_patient ID_INSTITUTION_INSTRUMENT_ddMMMyyyy of reception_ddMMMyyy of acquisition

- i. With the name of the protocol = Panel x
- ii. Sample ID1 = patient ID
- iii. Sample ID2 = INSTITUTION_INSTRUMENT
- iv. Sample ID3 = ddMMMyyy of reception
- v. Sample ID4 = ddMMMyyy of acquisition

Example: Panel 1_32140001_UBO_NAVIOS_09JUL2014_09JUL2014

- d. In the panel 1 and the panel 2: **indicate** the absolute number of your Flow-Count fluorospheres to enable the cell count calculations (in the CAL box of the Acquisition Manager).
2. Acquire the samples.
 - a. Adjust PMT voltages of P1 (FSC-A) and P2 (SSC-A) parameters to position the LY region at the following coordinates: This will facilitate the comparisons between sites.





PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

- b. Be careful to the threshold that should enable the detection of the Flow-Count fluorospheres.
- c. Adjust compensations if needed
- d. The minimum of cells to be acquired are indicated in the following table:

	STOP PARAMETERS (2 options)	
	Cell COUNT	Time stop
Panel 1	1,000 B cells	10 min
Panel 2	2,000 Lin-DR+ cells	17 min
Panel 3	20,000 T cells	10 min
Panel 4	20,000 T cells	10 min
Panel 5	20,000 B cells	10 min
Panel 6	20,000 B cells	10 min
Panel 7	10,000 B cells	5 min
Panel 8	10,000 B cells	5 min
Panel 9	10,000 B cells	5 min

- e. Save all LMD files as mentioned above:

Panel x_Patient ID_INSTITUTION_INSTRUMENT_date of reception_date of acquisition

Example: Panel 1_32140001_UBO_NAVIOS_09072014_09072014

AT THE END: QUIT YOUR SPECIFIC PRECISESADS CONFIGURATION

- 3. Upload all your LMD files (QC8 Peaks Beads and Panels files) to the intranet of the PRECISESADS website
 - a. Complete the information's file: the address will be communicated by email.
 - b. Upload your LMD files: the website address will be communicated by email.



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

If you have any issue with this procedure you can contact:

UBO team: Christophe JAMIN christophe.jamin@univ-brest.fr

or Jacques-Olivier PERS pers@univ-brest.fr

and

Beckman Coulter reference team: Olivier JAEN ojaen@beckman.com

or Nicole WEIT nweit@beckman.com



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

ANNEX 1

FULL ADDRESSES of the participants to the flow cytometry acquisition

Note:

Participants 1 to 8: Task 5.1 and 5.2: panel 1 to panel 9
Participants 9 to 11: Task 5.1 only: panel 1 and panel 2

Task 5.1 and 5.2

1	UBO	JO Pers, C Jamin	Navios	Beckman Coulter
2	FPS	C Marañon	FACS Verse	BD
3	SAS CORDOBA	R Lopez, N Barbarroja	FACS Aria III	BD
4	IDIBELL	E Ballestar, L Ciudad, D Alvarez, N Ramirez	Gallios	Beckman Coulter
5	IRCCS	E Trombetta	FACS Canto II	BD
6	DRFZ	F Hiepe, R Biesen V Gerl	FACS Canto II	BD
7	UNIGE	C Chizzolini, A Zuber	Navios	Beckman Coulter
8	K.U.LEUVEN	R Lories, E De Langhe, T Cantaert	LSR Fortessa	BD
9	UCL	B Lauwerys, J Ducreux	FACS Canto II	BD

Task 5.1

10	CHPORTO	C Vasconcelos, E Neves	Navios	Beckman Coulter
11	MHH	T Witte	FACS Canto II	BD



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

1- UBO (Brest): Navios (BCoulter)

Mr Christophe JAMIN
Laboratoire d'Immunologie
CHRU Morvan
5 av Foch
BP824
29609 Brest
FRANCE

christophe.jamin@univ-brest.fr
pers@univ-brest.fr

2- FPS (Granada): FACS Verse (BD)

Miss Concepción MARAÑÓN
PTS Granada
Parque Tecnológico de Ciencias de la Salud
Avenida de la Ilustración 114
18016
Granada
SPAIN

concepcion.maranon@genyo.es
marta.alarcon@genyo.es

3- SAS (Cordoba): FACS Aria III (BD)

Rosario Lopez-Pedrerá
IMIBIC-Unidad de Investigación
Hospital Universitario Reina Sofía
Avenida Menéndez Pidal s/n
14004
Córdoba
SPAIN

Tels. +34957736542 / +34651319413

rosario.lopez.exts@juntadeandalucia.es
nuria.barbarroja.exts@juntadeandalucia.es



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

4- IDIBELL (Barcelona): Gallios (BCoulter)

Prof Esteban BALLESTAR and Dr Laura CIUDAD
Cancer Epigenetics and Biology Programme (PEBC)
Bellvitge Biomedical Research Institute
Avda. Gran Via 199-203
08908 L'Hospitalet de Llobregat
Barcelona
SPAIN

eballestar@idibell.cat

lciedad@idibell.cat

dalvarez@idibell.cat

nramirez@idibell.cat

5- IRCCS (Milan): FACS Canto II (BD)

Elena Trombetta
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico
via F. Sforza n° 35
Padiglione Invernizzi 5° piano - Servizio di Citometria
20122 - Milano (MI)
Italy

trombetta.elena@gmail.com

6- DRFZ (Berlin): FACS Canto II (BD)

Prof Falk HIEPE
Dr Robert BIESEN
Dr Velia GERL
Department of Rheumatology and Clinical Immunology, Charité
Universitätsmedizin Berlin
Charitéplatz 1
D-10117 Berlin
GERMANY

falk.hiepe@charite.de

Robert.Biesen@charite.de

gerl@drfz.de



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

7- UNIGE (Geneva): Navios (BCoulter)

Prof Carlo CHIZZOLINI
Immunology and Allergy
University Hospital and School of Medicine
Rue Gabrielle Perret-Gentil 4
Geneva 14
1211
SWITZERLAND

Carlo.Chizzolini@unige.ch
Aleksandra.Zuber@unige.ch

8- KUL (Leuven): Fortessa (BD)

Prof Rik Lories
Faculty of Medicine
Laboratory of Tissue Homeostasis and Disease
Skeletal Biology and Engineering Research Center (SBE)
Department of Development and Regeneration
O&N - Box 813
Herestraat 49
B3000 Leuven
Belgium

Rik.Lories@med.kuleuven.be
ellen.delanghe@uzleuven.be
tineke.cantaert@med.kuleuven.be

9- UCL (Louvain) : FACS Canto II (BD)

Julie Ducreux
Service de Rhumatologie
Cliniques Universitaires Saint-Luc & Institut de Recherche Expérimentale et Clinique
Université catholique de Louvain
Avenue Hippocrate 10
bte B2.5390
1200 Bruxelles
BELGIQUE

Bernard.Lauwerys@uclouvain.be
julie.ducreux@uclouvain.be



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

10- CHP (Porto): Navios (BCoulter)

Dr Esmeralda NEVES
Centro Hospitalar do Porto
Serviço de Imunologia EX-CICAP
Rua D. Manuel II
4050-345
Porto
PORTUGAL

cvcarlosvasconcelos@gmail.com
m.esmeraldaneves@gmail.com

11- MHH (Hannover): FACS Canto II (BD)

Prof Torsten WITTE
Klinik für Immunologie und Rheumatologie
Medizinische Hochschule Hannover
Carl-Neuberg-Str. 1
30625
Hannover
GERMANY

Witte.Torsten@mh-hannover.de

SOP transfert des données



Work Package 5

A large, stylized sunburst graphic composed of numerous blue, rounded rectangular bars radiating from a central point, creating a circular shape. The bars vary slightly in length and are arranged in a fan-like pattern.

Standard Operating Procedure #4
for the transfer of raw Flow-cytometry
files to QuartzBio



Summary

I- RAW FILE NAMING.....	3
II- FLOW CYTOMETRY REPORT.....	4
III- FILE TRANSFER.....	4
IV- AUTOMATED CHECKS.....	5
V- PROCESS OVERVIEW.....	5
Appendices.....	6
Appendix 1: The flow cytometry report form.....	6
Appendix 2: Table of institutions.....	9
Appendix 3: Instructions for using SFTP.....	10
Appendix 4: List of institutions and contact information.....	13



I- RAW FILE NAMING

It is of paramount importance that all the transferred files adhere strictly to the following convention. It will allow us to automate the tracking and validation of the data transfers.

A- Naming conventions

- All dates must be formatted like **ddMMMyyyy** where MMM is the three letter prefix of the english month (e.g. 22SEP2014 for the 22nd day of september, 2014).
- All spaces (e.g. "Panel x") and unwanted information must be suppressed (e.g. the "Tag Number" and the ".PR" for BC users). Therefore, file named should be renamed if necessary.

B- QC8PeaksBeads files

QC8PeaksBeads_\$(patient_id)_\$(institution)_\$(instrument)_\$(reception_date)_\$(acquisition_date).extension
where:

- **patient_id:** the patient ID as an integer
 - **institution:** the canonical name of the institution (cf Appendix 2)
 - **instrument:** the canonical name of the instrument (cf Appendix 2)
 - **reception_date:** the date of reception of the sample
 - **acquisition_date:** the date of acquisition of the raw data
 - **.extension:** the original extension of the file (e.g. .fcs or .LMD)
- example for BC users (LMD files):
QC8PeaksBeads 32140001 UBO_NAVIOS 22SEP2014 22SEP2014 001.LMD
must be renamed as:
QC8PeaksBeads_32140001_UBO_NAVIOS_22SEP2014_22SEP2014.LMD
 - example for BD users (FCS 3.0 files):
QC8PeaksBeads_32140001_FPS_VERSE_22SEP2014_22SEP2014.fcs

C- Panel files

Panel_\$(panel_nb)_\$(patient_id)_\$(institution)_\$(reception_date)_\$(acquisition_date).extension
where:

- **panel_nb:** the panel number
- example for BC users (LMD files):
Panel 1.PR 32140001 UBO_NAVIOS 09SEP2014 09SEP2014 001.LMD
must be renamed as:
Panel_1_32140001_UBO_NAVIOS_09SEP2014_09SEP2014.LMD
 - example for BD users (FCS 3.0 files):
Panel 1_32140001_FPS_VERSE_09SEP2014_09SEP2014.fcs
must be renamed as:
Panel_1_32140001_FPS_VERSE_09SEP2014_09SEP2014.fcs



II- FLOW CYTOMETRY REPORT

Before transferring the files to QuartzBio, you have to fill the Flow cytometry report by clicking on the following link:

https://docs.google.com/forms/d/1MoUXjObLuXKFGnwm_aPDEU7frGmCmMwz5-y77ATdY8Y/viewform?c=0&w=1&usp=mail_form_link

This report is important for the computational immunologist who will analyze the data.

Cf Appendix 1

III- FILE TRANSFER

The files will be transferred to the project server in a secure way using SFTP (cf appendix 3).

Each center/institution will connect with a dedicated account, a dedicated area and a pre-configured file system tree as follows:

- INSTITUTION
 - uploads
 - CROSS_SECTIONAL_PHASE2
 - CROSS_SECTIONAL_SUBSTUDY_PHASE1
 - INCEPTION
 - timepoint0
 - timepoint1
 - timepoint2

The first level is the **cohort** name, and (if applicable) the second level the **timepoint**.

The uploading process is:

1. connect using the institution credentials (cf Appendix 2)
2. locate the appropriate folder in which to upload (e.g. Institution/uploads/CROSS_SECTIONAL);
3. create (if not already done) in the folder a sub-folder named after the patient ID (e.g. uploads/CROSS_SECTIONAL/32140001)
4. upload in this directory the corresponding files (properly named)



5.



IV- AUTOMATED CHECKS

An access on the server will be created for UBO, and a software tool will be developed allowing, given a spreadsheet file containing the form entries, to produce a report (as a CSV file), listing among others:

- uploads that are correct
- uploads that are not correct (misnamed files)
- uploads that do not match anything
- missing uploads

This tool will allow UBO to have an overview of the state of the uploads.

V- PROCESS OVERVIEW

Before the data transfers / setup:

- SFTP Account creation and setup (with Quartz Bio)

Data transfer:

- Proper naming of the files
- Fill the flow cytometry report form
- Connect to the sftp server
- Locate the appropriate folder
- Create if needed the subfolder for the patient ID
- Transfer the files

if you have any issue with this procedure you can contact:

UBO team: Lucas LE LANN Lucas.Lelann@etudiant.univ-brest.fr

and

Quartzbio: Karl FORNER karl.forner@quartzbio.com



Appendices

Appendix 1: The flow cytometry report form

A screenshot of a web-based form titled 'PRECISESADS Flow Cytometry Report'. The form has a light blue background and contains the following fields and options:

- PRECISESADS** logo and text: 'Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases'
- FLOW CYTOMETRY REPORT** title
- *Obligatoire** (red text)
- Site *** with a sub-label 'Institution' and a text input field.
- Flow Cytometer *** with a list of checkboxes:
 - Navios
 - Gallios
 - Canto II
 - Verse
 - Aria
 - Fortessa
- Patient ID *** with a text input field.
- Day and Hour of sample drawing *** with dropdown menus for 'Jour', 'Mois', '2014', and 'h', and a 'min' dropdown.



innovative
medicines
initiative



PRECISESADS

Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

Day and Hour of sample arrival *

Jour Mois 2014 h min

Day and Hour of sample preparation *

Jour Mois 2014 h min

Day and Hour of flow cytometry acquisition *

Jour Mois 2014 h min

New batch of the Duracell Tubes *

No
 Yes

PMT Voltages in accordance with QC8 Peaks Beads *

Yes
 No

Temperature stability *

<22°C
 22-25°C
 >25°C

Minimum Number of cells reached : Tube 1 *

Yes
 No

Minimum Number of cells reached : Tube 2 *

Yes
 No

Minimum Number of cells reached : Tube 3 *

Yes
 No
 NA

Minimum Number of cells reached : Tube 4 *

Yes
 No
 NA

Minimum Number of cells reached : Tube 5 *

Yes
 No
 NA

Minimum Number of cells reached : Tube 6 *

Yes
 No
 NA

Minimum Number of cells reached : Tube 7 *

Yes
 No
 NA

Minimum Number of cells reached : Tube 8 *

Yes
 No
 NA

Minimum Number of cells reached : Tube 9 *

Yes
 No
 NA

Speed acquisition *

Normal
 Higher


Continuer »

Terminé à 50 %

Formé par Google Forms

Ce contenu n'est ni rédigé, ni cautionné par Google.
Signaler un cas d'utilisation abusive - Conditions d'utilisation - Clauses additionnelles




PRECISESADS
Molecular Reclassification to Find
Clinically Useful Biomarkers for
Systemic Autoimmune Diseases

FLOW CYTOMETRY REPORT

New batch of the Duraclone Tubes

If Yes: specify which panel

- Panel 1
- Panel 2
- Panel 3
- Panel 4
- Panel 5
- Panel 6
- Panel 7
- Panel 8
- Panel 9

[« Retour](#) [Envoyer](#)

N'envoyez jamais de mots de passe via Google Forms.

100 % - vous avez réussi.



Appendix 2: Table of institutions

Table of institutions abbreviations

“Your Institution”	Short name / User name	Instrument
UBO (Brest, France)	ubo	NAVIOS
FPS (Granada, Spain)	fps	VERSE
SAS (Cordoba, Spain)	sas	FORTESSA
IDIBELL (Barcelona, Spain)	idibell	GALLIOS
IRCCS (Milan, Italy)	irccs	CANTO2
DRFZ (Berlin, Germany)	drfz	CANTO2
UNIGE (Geneva, Switzerland)	unige	NAVIOS
KU Leuven (Leuven, Belgium)	kul	FORTESSA
UC Louvain (Louvain, Belgium)	ucl	CANTO2
CHP (Porto, Portugal)	chp	NAVIOS
MHH (Hannover, Germany)	mhh	CANTO2

Appendix 3: Instructions for using SFTP

On Windows operating systems

1. Install the Winscp sftp customer (freeware <http://winscp.net/eng/docs/introduction>) on your computer: Select installation package for the latest version <http://winscp.net/eng/download.php> (need help ? See http://winscp.net/eng/docs/guide_install)
2. Execute the Winscp application (Winscp.exe)
3. Fill the Log in gate: (http://winscp.net/eng/docs/guide_connect).

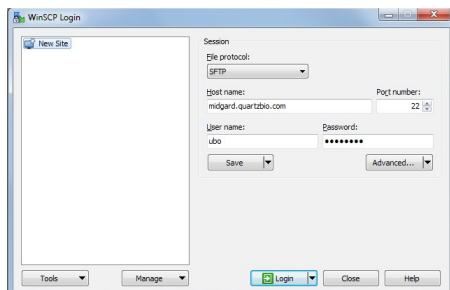
To log in you need the following information :

File protocol : SFTP

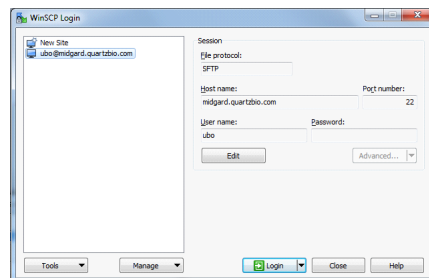
Host name : midgard.quartzbio.com

User name : your Short name/user name given in the Table of institutions (Appendix 2)

Password : given by Quartzbio



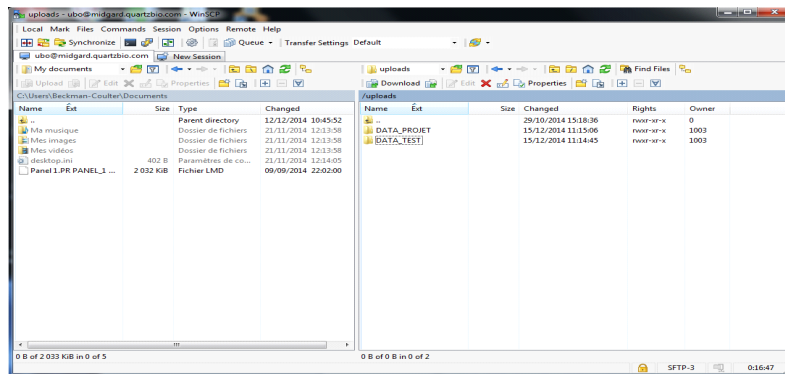
Click on save to facilitate the later access.



You can select the saved connection parameters and click on log in to access the server.

4. upload

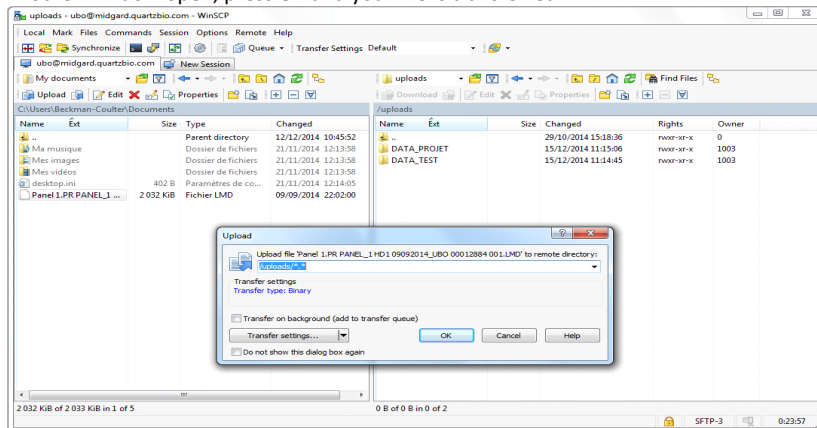
The left tab is your computer and the right tab is the server. On the right side you should find your institution. You can navigate, like in windows, to get to the right folder.



To upload a file:

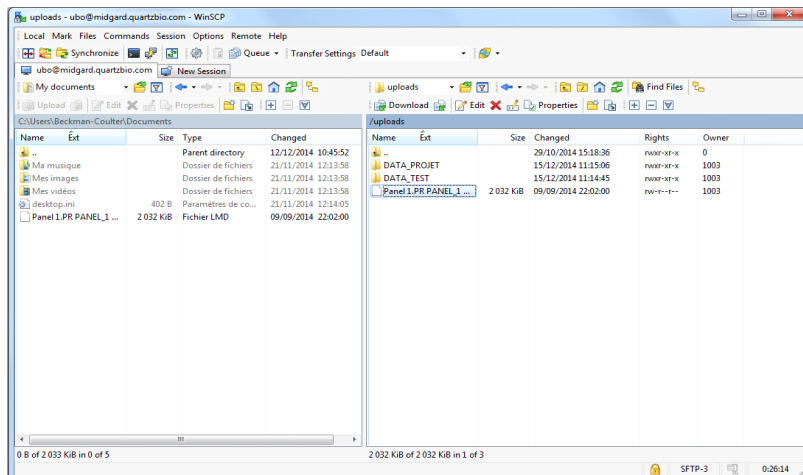
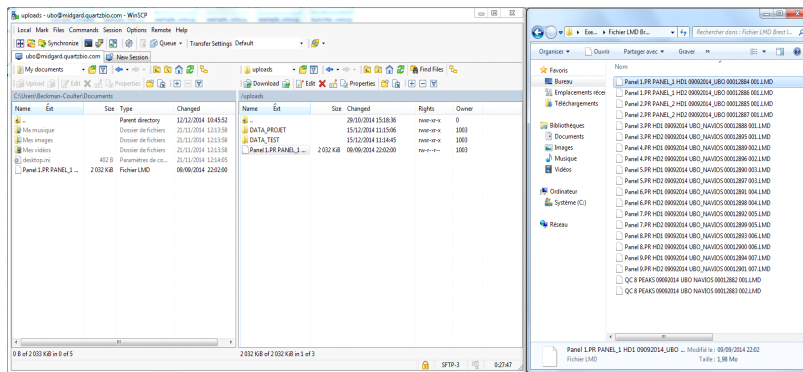
Select the files you want to send on the left side and then click on the upload button (left side, just above the files).

Another window open, press OK and your file is transferred.





Another solution is to open the folder in another window, select your files, drag and drop it on the right tab.



Example of the successful transfer of Panel 1.PR_PANEL_1 HD1_09092014_UBO_00012884_001.LMD



Appendix 4: List of institutions and contact information

1- UBO (Brest):

Jacques-Olivier Pers
Christophe JAMIN
Laboratoire d'Immunologie
CHRU Morvan
5 av Foch
BP824
29609 Brest
FRANCE

christophe.jamin@univ-brest.fr
pers@univ-brest.fr

2- FPS (Granada):

Concepción MARAÑÓN
PTS Granada
Parque Tecnológico de Ciencias de la Salud
Avenida de la Ilustración 114
18016
Granada
SPAIN
concepcion.maranon@genyo.es
marta.alarcon@genyo.es

3- SAS (Cordoba):

Rosario Lopez-Pedraza
IMIBIC-Unidad de Investigación
Hospital Universitario Reina Sofia
Avenida Menéndez Pidal s/n
14004
Córdoba
SPAIN
Tels. +34957736542 / +34651319413
rosario.lopez.exts@juntadeandalucia.es



4- IDIBELL (Barcelona):

Esteban BALLESTAR
Laura CIUDAD
Cancer Epigenetics and Biology Programme (PEBC)
Bellvitge Biomedical Research Institute
Avda. Gran Via 199-203
08908 L'Hospitalet de Llobregat
Barcelona
SPAIN

eballestar@idibell.cat
lciedad@idibell.cat
dalvarez@idibell.cat
namirez@idibell.cat

5- IRCCS (Milan):

Elena Trombetta
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico
via F. Sforza n° 35
Padiglione Invernizzi 5° piano - Servizio di Citometria
20122 - Milano (MI)
Italy

trombetta.elena@gmail.com

6- DRFZ (Berlin):

Falk HIEPE
Robert BIESEN
Velia GERL
Department of Rheumatology and Clinical Immunology, Charité
Universitätsmedizin Berlin
Charitéplatz 1
D-10117 Berlin
GERMANY

falk.hiepe@charite.de
Robert.Biesen@charite.de
gerl@drfz.de



7- UNIGE (Geneva):

Carlo CHIZZOLINI
Immunology and Allergy
University Hospital and School of Medicine
Rue Gabrielle Perret-Gentil 4
Geneva 14
1211
SWITZERLAND

Carlo.Chizzolini@unige.ch
Aleksandra.Zuber@unige.ch

8- KUL (Leuven):

Rik Lories
Faculty of Medicine
Laboratory of Tissue Homeostasis and Disease
Skeletal Biology and Engineering Research Center (SBE)
Department of Development and Regeneration
O&N - Box 813
Herestraat 49
B3000 Leuven
Belgium

Rik.Lories@med.kuleuven.be
ellen.delanghe@uzleuven.be
tineke.cantaert@med.kuleuven.be

9- UCL (Louvain) :

Julie Ducreux
Service de Rhumatologie
Cliniques Universitaires Saint-Luc & Institut de Recherche Expérimentale et Clinique
Université catholique de Louvain
Avenue Hippocrate 10
bte B2.5390
1200 Bruxelles
BELGIQUE

Bernard.Lauwerys@uclouvain.be
julie.ducreux@uclouvain.be



10-CHP (Porto):

Esmeralda NEVES
Centro Hospitalar do Porto
Serviço de Imunologia EX-CICAP
Rua D. Manuel II
4050-345
Porto
PORTUGAL

cvcarlosvasconcelos@gmail.com
m.esmeraldaneves@gmail.com

11-MHH (Hannover):

Torsten WITTE
Klinik für Immunologie und Rheumatologie
Medizinische Hochschule Hannover
Carl-Neuberg-Str. 1
30625
Hannover
GERMANY

Witte.Torsten@mh-hannover.de

A.7 SCRIPTS

A.7.1 Script R de normalisation intra-centre

Script 1: Script R de calcul et d'application de la normalisation sur les fichiers de cytométrie

```
#### Loading packages ####
suppressPackageStartupMessages(
{
  library(flowCore)
  library(flowStats)
  library(flowViz)
  library(ggcyto)
})
require(Ckmeans.1d.dp)

#### functions ####

# Change the scale to be between 0 and 1024 #

KalTransform <- function(transformId, a){
  t = new("transform", .Data = function(x) (x/a))
  t@transformationId = transformId
  t
}

# Definition of the 8 peaks beads MFI #
# Input are an 8 peaks beads fcs file and the path to this file #

border <- function(fcs, eight_peaks){
  name_file <- basename(eight_peaks)
  somePDFPath = paste("C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/REPORTS/8
    PEAKBEADS/PLOT/8peakbeads_norm_clustering_", name_file, ".pdf", sep="")
  pdf(file=somePDFPath)

  median_peak <- vector("list", 8)
  channels.estLoel <- colnames(fcs)[c(12)]

  test_lmd <- grepl(".LMD", eight_peaks, fixed=TRUE)
  if (test_lmd){
    KalT <- KalTransform(transformId="KalT", a=1024)
    mult=1024
  } else{
    KalT <- KalTransform(transformId="KalT", a=256)
    mult=256
  }

  translist <- transformList(fluc.channels, KalT)
  after_tmp <- transform(fcs, translist)
  arcTrans <- arcsinhTransform(transformationId="defaultArcsinhTransform", a=0, b=1, c=0)
  translist <- transformList(fluc.channels, arcTrans)
  after <- transform(after_tmp, translist)

  test_IDIBELL <- grepl("IDIBELL", eight_peaks, fixed=TRUE)
  if (test_lmd){
    if (test_IDIBELL){
      # Add a step to get cleaner FS/SS plot for the center IDIBELL #

      rgate <- rectangleGate(filterId="RG", "FS-A"=c(100, 1000000), "SS-A"=c
        (100000, 1000000))
      xyplot('SS-A'~'FS-A', after, filter=rgate)
      fres <- filter(after, rgate)
      fcs_split <- split(after, fres)
      beads <- fcs_split[1]$'RG+'

      lg <- flowStats::lymphGate(beads, channels = c("SS-A", "FS-A"), filterID=
        rgate1, scale= 3.0, plot = F)
      xyplot('SS-A'~'FS-A', beads, filter=lg$2gate)
    } else{
      lg <- flowStats::lymphGate(after, channels = c("SS-A", "FS-A"), scale=
        3.0, plot = F)
      xyplot('SS-A'~'FS-A', after, filter=lg$2gate)
    }
  } else{
    lg <- flowStats::lymphGate(after, channels = c("SSC-A", "FSC-A"), scale= 5.0, plot = F)
    xyplot('SSC-A'~'FSC-A', after, filter=lg$2gate)
  }

  #number of peaks
  i=1
  for (channel in fluc.channels){
    nb_clust <- function(type) {
      switch(type, "FITC.A"=8,
        "PE.A"=8,
        "PC5.5.A"=6,
        "PC7.A"=6,
        "APC.A"=8,
        "APC.AF750.A"=8,
```

```

                                "PB.A"=8,
                                "KO.A"=8)
    }
    k <- nb_clust(channel)
    if (grepl("UBO",name_file) & (channel == "APC.AF750.A")){
        k <- 9
    }
    result <- Ckmedian.l.d.dp(lg$x@exprs[,channel], k)
    m<-result$center

    a=1
    if (grepl("UCL",name_file) & (channel == "KO.A")){
        median_peak[[i]]<-(sinh(m)*mult)[-8]
    } else if (grepl("KUL",name_file) & (channel == "APC.A")){
        median_peak[[i]]<-(sinh(m)*mult)[-8]
    } else {
        median_peak[[i]]<-sinh(m)*mult
        a=0
    }
    names(median_peak)<-fluo.channels
    i=i+1

    plot(density(lg$x@exprs[,channel],bw = 0.001),main = channel)
    points(x=result$center,y = rep(3:4,length.out=k),col=1:k,pch=0,cex=1)

}
dev.off()
return(median_peak)
}

#### End of functions ####

args <- commandArgs(trailingOnly = TRUE)
PEAKSDataPath<-args[1]
REFDataPath<-args[2]
flowDataPath<-args[3]

fluo.channels<- c("FITC.A","PE.A","PC5.5.A","PC7.A","APC.A","APC.AF750.A","PB.A","KO.A")
PB_REF <- flowCore::read.FCS(file.path(REFDataPath),dataset=2)

test_nb_colname <- grepl("MHH",REFDataPath, fixed=TRUE)
test_parameters1<-grepl("DRFZ_CANTO2_12SEP2017",REFDataPath, fixed=TRUE)
test_parameters2<-grepl("DRFZ_CANTO2_21OCT2015",REFDataPath, fixed=TRUE)

if (test_nb_colname & length(colnames(PB_REF@exprs))<12){
    colnames(PB_REF@exprs)[3:10]<- fluo.channels
    colnames(PB_REF)[3:10]<- fluo.channels
} else if (test_parameters1 | test_parameters2){
    colnames(PB_REF@exprs)[8:15]<- fluo.channels
    colnames(PB_REF)[8:15]<- fluo.channels
} else{
    colnames(PB_REF@exprs)[7:14]<- fluo.channels
    colnames(PB_REF)[7:14]<- fluo.channels
}

PB_center <- flowCore::read.FCS(file.path(PEAKSDataPath),dataset=2)
colnames(PB_center@exprs)[7:14]<- fluo.channels
colnames(PB_center)[7:14]<- fluo.channels

border_lim_REF<-border(PB_REF,REFDataPath)
border_lim_center<-border(PB_center,PEAKSDataPath)

test_lmd <- grepl(".LMD",PEAKSDataPath, fixed=TRUE)
if (test_lmd){
    ext=".LMD"
    mult=1024
} else{
    ext=".fcs"
    mult=256
}

y<-do.call(cbind,border_lim_REF)/mult
x<-do.call(cbind,border_lim_center)/mult

refnfile=basename(REFDataPath)
refCSVPath = paste("C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/REPORTS/SPEAKBEADS/
MATRICES/",refnfile,"_REF.csv",sep="")
write.csv(y,refCSVPath)

nfile=basename(PEAKSDataPath)
CSVPath = paste("C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/REPORTS/SPEAKBEADS/MATRICES/
",nfile,".csv",sep="")
write.csv(x,CSVPath)

trans <-do.call('rbind', lapply( fluo.channels, function(chan) coefficients(lm(y[,chan] ~ x[,chan]
))) ) )
rownames(trans) <- fluo.channels

```

```

colnames(trans) <- c('alpha', 'beta')

nfile=basename(PEAKSDataPath)
transCSVPath = paste("C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/REPORTS/NORMALISATION/
trans_", nfile, ".csv", sep="")
write.csv(trans, transCSVPath)

ErrorTransPath<-"C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/REPORTS/NORMALISATION/high_
alpha_norm.txt"

for (chan_t in fluo.channels){
  intra_alpha <-trans[chan_t, 'alpha']
  if (abs(intra_alpha) > 1){
    write(paste(nfile, chan_t, intra_alpha, sep=";"), ErrorTransPath, append = TRUE)
  }
  trans[chan_t, 'alpha']<-0
}

normalisation <- lapply(fluo.channels, function(n) return(function(x) trans[n, 'alpha'] + x*
trans[n, 'beta']))
names(normalisation) <- fluo.channels

P_fcs <- list.files(pattern = ext, flowDataPath, recursive=F, full = TRUE)

for (fcs_file in P_fcs){
  fcs <- flowCore::read.FCS(file.path(fcs_file), dataset=2)
  colnames(fcs)[7:14]<- fluo.channels
  if (is.null(fcs@description$SPILL)){
    colnames(fcs@description$'SPILLOVER')<- fluo.channels
  } else {
    colnames(fcs@description$SPILL)[1:8]<- fluo.channels
  }
  X <- fcs@exprs/mult

  for (chan in fluo.channels){
    x.norm <- normalisation[[chan]](X[,chan])
    fcs@exprs[,chan] <- x.norm*mult
  }

  filename = paste(fcs_file, "intra.fcs", sep="")
  write.FCS(fcs, filename, delimiter = "\\")

  test_QC <- grepl("QC8", fcs_file, fixed=TRUE)
  if (test_QC){
    ErrorPath<-"C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/REPORTS/8
PEAKBEADS/CHECK_NORM/8peakbeads_norm_mistake.txt"
    border_norm_center<-border(fcs, filename)
    xnorm<-do.call(cbind, border_norm_center)/mult

    dif_norm <- ((y-xnorm)/y)*100
    for (chan in fluo.channels){
      nb_peak <- function(type) {
        switch(type, "FITC.A"=8,
                  "PE.A"=8,
                  "PC5.5.A"=6,
                  "PC7.A"=6,
                  "APC.A"=8,
                  "APC.AF750.A"=8,
                  "PB.A"=8,
                  "KO.A"=8)
      }
      starting_peak <- function(type) {
        switch(type, "FITC.A"=2,
                  "PE.A"=2,
                  "PC5.5.A"=2,
                  "PC7.A"=3,
                  "APC.A"=2,
                  "APC.AF750.A"=4,
                  "PB.A"=2,
                  "KO.A"=2)
      }
      peak_nb<-nb_peak(chan)
      peak<- starting_peak(chan)
      while(peak <= peak_nb){
        peak_value_dif <- dif_norm[peak,chan]
        if (peak_value_dif > 5){
          write(paste(basename(fcs_file), chan, peak, peak_value_dif,
                      sep=";"), ErrorPath, append = TRUE)
        }
        peak<-peak+1
      }
    }
  }
}

print("End_of_the_intra-center_normalisation_for_this_inclusion")

```

A.7.2 Script Python de correction de l'effet lot

Script 2: Script Python de calcul et d'application de la correction de l'effet batch sur les fichiers de cytométrie

```
#!/usr/bin/python
# Lucas Le Lann
#12/10/2018

### Import
from collections import defaultdict
from function_cytometry_results import *
from numpy import median
from datetime import datetime

import os.path
import glob, os
import matplotlib.pyplot as plt
import matplotlib.cbook as cbook
import numpy as np
import pandas as pd
import csv
import re

### FONCTION

def get_DATE(OMICID):
    #path="C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/PATIENTS/"
    path="E:/PRECISESADS_DATA/PATIENTS/"
    for PHASE in glob.glob(path+"*"):
        for center in glob.glob(PHASE+"*"):
            for patient in glob.glob(center+"*"):
                OMIC=patient.split("\\")[1]
                if OMIC==OMICID:
                    list_files=glob.glob(patient+"*")
                    DATE=list_files[0].split("_")[-1].split(".")[0]
                    return DATE

    return 0

def formating_date(date):
    if date == "DATE" or "-" in date:
        return date
    elif date == "":
        date="DATE"
    else:
        print date
        date=date.replace(". fcs", "")
        date=date.replace(". LMD", "")
        date=date.replace(". csv", "")

        date=date.replace("JAN", "/01/")
        date=date.replace("FEB", "/02/")
        date=date.replace("FEV", "/02/")
        date=date.replace("MAR", "/03/")
        date=date.replace("APR", "/04/")
        date=date.replace("AVR", "/04/")
        date=date.replace("MAY", "/05/")
        date=date.replace("MAI", "/05/")
        date=date.replace("JUN", "/06/")
        date=date.replace("Jun", "/06/")
        date=date.replace("JUL", "/07/")
        date=date.replace("AUG", "/08/")
        date=date.replace("AGU", "/08/")
        date=date.replace("SEP", "/09/")
        date=date.replace("OCT", "/10/")
        date=date.replace("Oct", "/10/")
        date=date.replace("OKT", "/10/")
        date=date.replace("NOV", "/11/")
        date=date.replace("DEC", "/12/")
        date=date.replace("DIC", "/12/")

        date=date.replace("T", "")
        date=date.replace("IL", "")
        date=date.replace("_", "")
        date=date.replace(".", "")
        tmp_date=date.split("/")
        print tmp_date
        if len(list(tmp_date[2]))== 2:
            tmp_date[2]="20"+tmp_date[2]
        newdate=tmp_date[2]+"-"+tmp_date[1]+"-"+tmp_date[0]

    return newdate
```

```

def search_date(OMICID):

    PHASEI={}
    PHASEII={}
    INCEPTION={}

    with open("C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/TRANSMART/transmart_15_06_
    2018_PHASE I_II.tsv","r") as transmart_phase:
        header=0
        transmart_phase.seek(0)
        for line in transmart_phase:
            if header==0:
                header=1
                line=line.split("\t")
                col_nb=0
                for head in line:
                    if "number" in head:
                        OMICID_col=col_nb
                    elif "Sampling\Date_Of_Sampling" in head:
                        DATE_col=col_nb
                    elif "CS_Phase" in head:
                        PHASE_col=col_nb

                col_nb+=1
            else:
                line=line.split("\t")
                OMICID_tm=line[OMICID_col]
                #DATE_tm=trad_date(line[DATE_col])
                DATE_tm=line[DATE_col]
                if "I1" in line[PHASE_col]:
                    PHASEII[OMICID_tm]=DATE_tm
                elif "I" in line[PHASE_col]:
                    PHASEI[OMICID_tm]=DATE_tm
                #else:
                #print line[PHASE_col]

    with open("C:/Users/Beckman-Coulter/Documents/PRECISESADS_DATA/TRANSMART/transmart_15_06_
    2018_inception.tsv","r") as transmart_phase:
        header=0
        for line in transmart_phase:
            if header==0:
                header=1
                line=line.split("\t")
                col_nb=0
                for head in line:
                    if "Sampling\OMIC_number\M000" in head:
                        OMICID_M00_col=col_nb
                    elif "Sampling\OMIC_number\M006" in head:
                        OMICID_M06_col=col_nb
                    elif "Sampling\OMIC_number\M014" in head:
                        OMICID_M14_col=col_nb
                    elif "Date_Of_Sampling\M000" in head:
                        DATE_M00_col=col_nb
                    elif "Date_Of_Sampling\M006" in head:
                        DATE_M06_col=col_nb
                    elif "Date_Of_Sampling\M014" in head:
                        DATE_M14_col=col_nb

                col_nb+=1
            else:
                line=line.split("\t")
                OMICID_M00_tm=line[OMICID_M00_col]
                DATE_M00_tm=line[DATE_M00_col]
                INCEPTION[OMICID_M00_tm]=DATE_M00_tm+"_M0"

                OMICID_M06_tm=line[OMICID_M06_col]
                DATE_M06_tm=line[DATE_M06_col]
                INCEPTION[OMICID_M06_tm]=DATE_M06_tm+"_M6"

                OMICID_M14_tm=line[OMICID_M14_col]
                DATE_M14_tm=line[DATE_M14_col]
                INCEPTION[OMICID_M14_tm]=DATE_M14_tm+"_M14"

    OMICID="\N"+OMICID+"\N"
    if OMICID in PHASEII.keys():
        if get_DATE(OMICID) == 0:
            DATE=PHASEII[OMICID]
        else:
            DATE=get_DATE(OMICID)

    elif OMICID in PHASEI.keys():
        if get_DATE(OMICID) == 0:
            DATE=PHASEI[OMICID]
        else:
            DATE=get_DATE(OMICID)

    elif OMICID in INCEPTION.keys():
        if get_DATE(OMICID) == 0:
            DATE=INCEPTION[OMICID].split("_")[0]
            DATE=DATE.replace("\N","")
        else:
            DATE=get_DATE(OMICID)

```

```

else:
    DATE="DATE"
    print "MISSING_"+ OMICID

DATE=formating_date(DATE)

return DATE

def get_coef(PANEL,CENTER,CHANNEL,DATE):
    coefFile = open("./MFI/MATRICES/COEF_BATCH/matrice_panel_"+PANEL+"_coef_batch.csv", 'r')

    if DATE == "2015-10-19":
        col=2
    elif DATE == "2017-04-14":
        col=3
    else:
        print "meh"

    for line in coefFile:
        line=line.split(";")
        if line[0] == CENTER and line[1]==CHANNEL:
            return line[col]

def formatting_month(month):

    month=month.replace("JAN","01")
    month=month.replace("FEB","02")
    month=month.replace("FEV","02")
    month=month.replace("MAR","03")
    month=month.replace("APR","04")
    month=month.replace("AVR","04")
    month=month.replace("MAY","05")
    month=month.replace("MAI","05")
    month=month.replace("JUN","06")
    month=month.replace("Jun","06")
    month=month.replace("JUL","07")
    month=month.replace("AUG","08")
    month=month.replace("AGU","08")
    month=month.replace("SEP","09")
    month=month.replace("OCT","10")
    month=month.replace("Oct","10")
    month=month.replace("OKT","10")
    month=month.replace("NOV","11")
    month=month.replace("DEC","12")
    month=month.replace("DIC","12")

    month=month.replace("T","")
    month=month.replace("IL","")
    month=month.replace("_","")
    month=month.replace(".","")

    return month

def testdate(date):
    #print "entre fonction"
    match=re.search("^\d{2}\D{3}\d{2}",date)
    if match:

        sdate=list(date)
        #print sdate
        month=sdate[2]+sdate[3]+sdate[4]

        newdate="20"+sdate[5]+sdate[6]+"-"+formatting_month(month)+"-"+sdate[0]+sdate[1]
        return newdate
    else:
        return date

### Fin des fonctions

print "start"

print "_clear_folders"
clear_panel=False
if clear_panel:
    #CLEAR panels
    for fichiers in glob.glob("./MFI/CORRECTED_MFI/*.csv"):
        os.remove(fichiers)
    for fichiers in glob.glob("./MFI/BATCH_MFI/*.csv"):
        os.remove(fichiers)
    for fichiers in glob.glob("./MFI/CENTER_MFI/*.csv"):
        os.remove(fichiers)
    print "MFI_BATCH_folder_clear"

print "define_list_of_blacklisted_OMICs"

#setting the dictionary
list_panel=["P1","P2","P3","P4","P5","P6","P7","P8","P9"]
BL={}

```

```

WL={}
for element in list_panel:
    BL[element]=[]
    WL[element]=[]

blacklist=open("detailedblacklist.txt","r")
whitelist=open("detailedwhitelist.txt","r")
##create list of whitelisted OMIC ID and the PANEL
for line_w in whitelist:
    OMICID_w=line_w.split(";")[0]
    PANELID_w=line_w.split(";")[2].replace("\n","")
    if PANELID_w=="ALL":
        for ID_w in WL.keys():
            WL[ID_w].append(OMICID_w)
    else:
        WL[PANELID_w].append(OMICID_w)

##create list of Blacklisted OMIC ID and the PANEL that are not in the whitelisted list
for line in blacklist:
    OMICID=line.split(";")[0]
    PANELID=line.split(";")[2].replace("\n","")

    if PANELID=="ALL":
        for ID in BL.keys():
            BL[ID].append(OMICID)
    else:
        if OMICID not in WL[PANELID]:
            BL[PANELID].append(OMICID)
print "OMICs_blacklist_{}_done".format(PANELID)

print "###_{}_standardized_channel_names_and_add_DATE_to_csv_file_###".format(PANELID)

# setting some parameters
count_panel=1
panel_nb_wanted=9

if True:
    while count_panel <= panel_nb_wanted:
        list_REF=[]
        for file in glob.glob("./MFI/RAW_MFI/*"+str(count_panel)+".csv"):
            with open(file,"r") as MFI_file:

                #print file
                IC=file.split("\\")[1]
                OMIC=IC.split("_")[0]
                date=search_date(OMIC)
                newMFI_file=file.replace("RAW_MFI","CORRECTED_MFI")
                newMFI_file=newMFI_file.replace("DATE",date.replace("\n",""))
                #print newMFI_file
                if not os.path.isfile(newMFI_file) and OMIC not in BL["P"+str(count_panel)]:
                    with open(newMFI_file,"a") as newMFI_file:
                        header=0
                        for line in MFI_file:
                            #print line
                            sline=line.split("\t")
                            REF=sline[3]
                            REF=REF.replace("-",".")
                            list_REF.append(REF)
                            if count_panel==1:
                                if "CD3" in REF or "APC.AF750" in REF:
                                    REF="CD3_APC.AF750"
                                elif "CD19" in REF or ("APC" in REF and not "APC.AF750" in REF):
                                    REF="CD19_APC"
                                elif "CD16" in REF or "FITC" in REF:
                                    REF="CD16_FITC"
                                elif "CD15" in REF or "PE" in REF:
                                    REF="CD15_PE"
                                elif "CD56" in REF or "PC5.5" in REF:
                                    REF="CD56_PC5.5"
                                elif "CD14" in REF or "PC7" in REF:
                                    REF="CD14_PC7"
                                elif "CD4" in REF or "PB" in REF:
                                    REF="CD4_PB"
                                elif "CD8" in REF or "KO" in REF:
                                    REF="CD8_KO"
                                else:
                                    print REF
                            elif count_panel==2:
                                if "CD123" in REF or ("APC" in REF and not "APC.AF750" in REF):
                                    REF="CD123_APC"

```

```

elif "CD1c" in REF or "FITC" in
REF:
REF= "CD1c_FITC"
elif "Lin" in REF or "PE" in REF:
REF= "Lin_PE"
elif "CD141" in REF or "PC5" in
REF:
REF= "CD56_PC5.5"
elif "CD11c" in REF or "PC7" in
REF:
REF= "CD11c_PC7"
elif "HLA" in REF or "PB" in REF:
REF= "HLADR_PB"
else:
print REF

elif count_panel == 3:
if "CD57" in REF or "FITC" in REF
:
REF= "CD57_FITC"
elif "CD45RA" in REF or "PE" in
REF:
REF= "CD45RA_PE"
elif "CD27" in REF or "PC5" in
REF:
REF= "CD27_PC5.5"
elif "CD62L" in REF or "PC7" in
REF:
REF= "CD62L_PC7"
elif "CD38" in REF or ("APC" in
REF and not "APC.AF750" in
REF):
REF= "CD38_APC"
elif "CD3" in REF or "APC.AF750"
in REF:
REF= "CD3_APC.AF750"
elif "CD4" in REF or "PB" in REF:
REF= "CD4_PB"
elif "CD8" in REF or "KO" in REF:
REF= "CD8_KO"
else:
print REF

elif count_panel == 4:
if "gd_TCR" in REF or "FITC" in
REF:
REF= "gd_TCR_FITC"
elif "ab_TCR" in REF or "PE" in
REF:
REF= "ab_TCR_PE"
elif "CD25" in REF or "PC5" in
REF:
REF= "CD25_PC5.5"
elif "CD127" in REF or "PC7" in
REF:
REF= "CD127_PC7"
elif "CD69" in REF or ("APC" in
REF and not "APC.AF750" in
REF):
REF= "CD69_APC"
elif "CD3" in REF or "APC.AF750"
in REF:
REF= "CD3_APC.AF750"
elif "CD4" in REF or "PB" in REF:
REF= "CD4_PB"
elif "CD8" in REF or "KO" in REF:
REF= "CD8_KO"
else:
print REF

elif count_panel == 5:
if "IgD" in REF or "FITC" in REF:
REF= "IgD_FITC"
elif "CD267" in REF or "PE" in
REF:
REF= "CD267_PE"
elif "CD5" in REF or "PC5" in REF
:
REF= "CD5_PC5.5"
elif "CD27" in REF or "PC7" in
REF:
REF= "CD27_PC7"
elif "CD19" in REF or ("APC" in
REF and not "APC.AF750" in
REF):
REF= "CD19_APC"
elif "CD24" in REF or "APC.AF750"
in REF:
REF= "CD24_APC.AF750"
elif "CD38" in REF or "PB" in REF
:
REF= "CD38_PB"
else:
print REF

```



```

elif count_panel == 6:
    if "CD43" in REF or "FITC" in REF
    :
        REF= "CD43_FITC"
    elif "CD69" in REF or "PE" in REF
    :
        REF= "CD69_PE"
    elif "CD5" in REF or "PC5" in REF
    :
        REF= "CD5_PC5.5"
    elif "CD27" in REF or "PC7" in
    REF:
        REF= "CD27_PC7"
    elif "CD11b" in REF or ("APC" in
    REF and not "APC.AF750" in
    REF):
        REF= "CD11b_APC"
    elif "CD20" in REF or "PB" in REF
    :
        REF= "CD20_PB"
    elif "CD3" in REF or "KO" in REF:
        REF= "CD3_KO"
    else:
        print REF

elif count_panel == 7:
    if "CD64" in REF or "FITC" in REF
    :
        REF= "CD64_FITC"
    elif "CD32" in REF or "PE" in REF
    :
        REF= "CD32_PE"
    elif "CD18" in REF or "PC5" in
    REF:
        REF= "CD18_PC5.5"
    elif "CD14" in REF or "PC7" in
    REF:
        REF= "CD14_PC7"
    elif "CD19" in REF or ("APC" in
    REF and not "APC.AF750" in
    REF):
        REF= "CD19_APC"
    elif "CD11b" in REF or "APC.AF750
    " in REF:
        REF= "CD11b_APC.AF750"
    elif "CD16" in REF or "KO" in REF
    :
        REF= "CD16_KO"
    else:
        print REF

elif count_panel == 8:
    if "CD35" in REF or "FITC" in REF
    :
        REF= "CD35_FITC"
    elif "CD59" in REF or "PE" in REF
    :
        REF= "CD59_PE"
    elif "CD14" in REF or "PC7" in
    REF:
        REF= "CD14_PC7"
    elif "CD19" in REF or ("APC" in
    REF and not "APC.AF750" in
    REF):
        REF= "CD19_APC"
    elif "CD3" in REF or "APC.AF750"
    in REF:
        REF= "CD3_APC.AF750"
    elif "CD21" in REF or "PB" in REF
    :
        REF= "CD21_PB"
    elif "CD16" in REF or "KO" in REF
    :
        REF= "CD16_KO"
    else:
        print REF

elif count_panel == 9:
    if "CD46" in REF or "FITC" in REF
    :
        REF= "CD46_FITC"
    elif "CD55" in REF or "PE" in REF
    :
        REF= "CD55_PE"
    elif "CD14" in REF or "PC7" in
    REF:
        REF= "CD14_PC7"
    elif "CD19" in REF or ("APC" in
    REF and not "APC.AF750" in
    REF):
        REF= "CD19_APC"
    elif "CD3" in REF or "APC.AF750"
    in REF:
        REF= "CD3_APC.AF750"

```

```

elif "CD16" in REF or "KO" in REF
:
REF= "CD16_KO"
else:
print REF

else:
print count_panel

if header==0:
header=1
newline=sline[0]+\t"+sline[1]+\
t"+sline[2]+\t"+REF+"\t"+
sline[4].replace("\n","")+
tDATE\n"
else:
newline=sline[0]+\t"+sline[1]+\
t"+sline[2]+\t"+REF+"\t"+
sline[4].replace("\n","")+
t"+date+"\n"

newMFI_file.write(newline)

count_panel+=1

print "###CORRECTED_DATE_done###"

### COMPUTE COEF BATCH ###
print "###BEGINING_CALCULE_BATCH_COEF###"

list_centre=("DRFZ", "FPS", "IRCCS", "IDIBELL", "KUL", "SAS", "UBO", "UCL", "UNIGE")
list_chan=('channel', 'FITC', 'PE', 'PC5.5', 'PC7', 'APC', 'AF750', 'PB', 'KO')

if True:
panel=1
while panel<=panel_nb_wanted:
value_dict={}
if panel == 1:
coef_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_1_coef_batch.csv"
med_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_1_med_batch.csv"
list_pop_selected=("CD15lowCD16high_Neutrophils_CD16_FITC", "PMN_CD15_PE",
"CD3negCD56pos_NKcells_CD56_PC5.5", "CD14highCD16neg
classicalMonocytes_CD14_PC7", "CD19pos_Bcells_CD19_APC", "CD3pos
Tcells_CD3_APC.AF750", "CD4pos_Tcells_CD4_PB", "CD8pos_Tcells_CD8_KO")
list_centre=("CHP", "DRFZ", "FPS", "IRCCS", "IDIBELL", "MHH", "KUL", "SAS", "UBO",
"UCL", "UNIGE")

elif panel == 2:
coef_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_2_coef_batch.csv"
med_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_2_med_batch.csv"
list_pop_selected=("mDC1_CD1c_FITC", "mDC2_CD56_PC5.5", "mDC_CD11c_PC7", "
BASO_CD123_APC", "DRpos_linneg_HLADR_PB")
list_centre=("CHP", "DRFZ", "FPS", "IRCCS", "IDIBELL", "MHH", "KUL", "SAS", "UBO",
"UCL", "UNIGE")

elif panel == 3:
coef_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_3_coef_batch.csv"
med_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_3_med_batch.csv"
list_pop_selected=("CD4 TEMRA_57_pos27neg_CD57_FITC", "CD4 TN_CD45RA_PC5.5",
"CD4 TCM_57neg27_pos_CD27_PC5.5", "CD8 TN_CD62L_PC7", "CD8 TN_57
neg27_pos_38_pos_CD38_APC", "Lymphocytes_T_CD3_APC.AF750", "CD4pos_CD4_
PB", "CD8pos_CD8_KO")

elif panel == 4:
coef_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_4_coef_batch.csv"
med_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_4_med_batch.csv"
list_pop_selected=("CD4pos_gdpos_abneg_gdTCR_FITC", "CD8pos_gdneg_abpos_
abTCR_PE", "CD4pos_CD25highCD127low_CD25_PC5.5", "CD4pos
CD25highCD127low_CD127_PC7", "CD4pos_CD69pos_CD25neg_CD69_APC", "T
CELLS_CD3_APC.AF750", "CD4pos_CD8neg_CD4_PB", "CD4neg_CD8pos_CD8_KO")

elif panel == 5:
coef_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_5_coef_batch.csv"
med_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_5_med_batch.csv"
list_pop_selected=("LB_mature_naive_IgD_FITC", "CD267pos_CD5neg_CD267_PE",
"CD5pos_CD27pos_CD5_PC5.5", "Plasmablasts_CD27_PC7", "B_cells_CD19_APC",
"IgDM_CD38pos_CD5neg_CD267neg_CD24pos_CD24_APC.AF750", "
Plasmablasts_CD38_PB")

elif panel == 6:
coef_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_6_coef_batch.csv"
med_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_6_med_batch.csv"
list_pop_selected=("CD20pos_CD27neg_CD43pos_CD43_FITC", "CD20pos_CD69pos_
CD69_PE", "B1_CD5pos_CD11bpos_CD5_PC5.5", "B1_CD27_PC7", "B1_CD5neg_
CD11bpos_CD11b_APC", "CD20pos_CD69pos_CD20_PB")

elif panel == 7:
coef_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_7_coef_batch.csv"
med_file= "/MFI/MATRICES/COEF_BATCH/matrice_panel_7_med_batch.csv"
list_pop_selected=("CD16pos_CD32pos_CD64pos_CD64_FITC", "CD16pos_CD32pos_
CD64neg_CD32_PE", "CD19pos_in_Bcells_CD19_APC", "CD14pos_CD11bpos_
CD18pos_CD11b_APC.AF750", "CD16pos_in_PMN_CD16_KO")

elif panel == 8:

```

```

coef_file= "./MFI/MATRICES/COEF_BATCH/matrice_panel_8_coef_batch.csv"
med_file= "./MFI/MATRICES/COEF_BATCH/matrice_panel_8_med_batch.csv"
list_pop_selected=("CD16pos_CD32pos_CD64pos_CD35_FITC","Tcells_CD59pos_
CD59_PE","CD14pos_in_Mono_CD14_PC7","Bcells_CD19_APC","Tcells_CD3_
APC.AF750","CD16pos_CD21pos_CD35pos_CD21_PB","CD16pos_in_PMN_CD16_KO
")

elif panel == 9:
coef_file= "./MFI/MATRICES/COEF_BATCH/matrice_panel_9_coef_batch.csv"
med_file= "./MFI/MATRICES/COEF_BATCH/matrice_panel_9_med_batch.csv"
list_pop_selected=("CD16pos_CD35pos_CD46pos_CD46_FITC","CD16pos_CD55pos_
CD46pos_CD55_PE","CD14pos_in_Mono_CD14_PC7","Tcells_CD3_APC.AF750","
CD16pos_in_PMN_CD16_KO")

else:
    print "Panel_number_not_implemented"

os.remove(coef_file)
f=open(coef_file,"a")
f.write("CENTER;CHANNEL;X2015.10.19;X2017.04.14\n")
f.close()

os.remove(med_file)
f=open(med_file,"a")
f.close()

for panel_csv in glob.glob("./MFI/CORRECTED_MFI/*PANEL_"+str(panel)+".csv"):
    first_line=True

    site=panel_csv.split("_")[2]
    if "CHAR" in site:
        site="DRFZ"
    if site == "IDIBEL":
        site="IDIBELL"
    for line in open(panel_csv,"r"):
        if first_line:
            OMIC=line.split("\t")[0]
            first_line=False
        else:
            line=line.split("\t")
            pop=line[1]
            ref=line[3]
            pop_ref=pop+" "+ref
            value = line[4]
            date=line[5].replace("\n","")
            date=testdate(date)

            ## determination which "batch" the patient is in
            if "DATE" not in date and date != '':
                date=date.replace("\n","")
                date=datetime.strptime(date, '%Y-%m-%d')
                first_batch=datetime.strptime("2015-06-01", '%Y-%
m-%d')
                second_batch=datetime.strptime("2015-10-19", '%Y
-%m-%d')
                third_batch=datetime.strptime("2017-04-14", '%Y-%
m-%d')
            #print line[3]
            channel=line[3].split("_")[1]
            if date < second_batch: #and date > first_batch:
                batch="1"
            #elif date < first_batch:
            #batch= "zero"
            else:
                if date < third_batch:
                    batch="2"
                else:
                    batch="3"

            if pop_ref in list_pop_selected:
                batch = str(batch)
                if batch not in value_dict.keys():
                    value_dict[batch]={}
                    for centre in list_centre:
                        value_dict[batch][centre]={}
                        for refpop in list_pop_selected:
                            value_dict[batch][centre][refpop]=[]
                if value != "NA" and value != "" and value != "N/
A":
                    value_dict[batch][site][pop_ref].append(
                        float(value))

batch_med=value_dict
for lot in value_dict.keys():
    for center in value_dict[lot].keys():
        for pop in value_dict[lot][center].keys():
            if value_dict[lot][center][pop] != []:

                number_used_for_med=len(value_dict[lot][center][
pop])
                #if center=="UBO" and "CD19" in pop:
                #print sum(value_dict[lot][center][pop])

```

```

# print len(value_dict[lot][center][pop])
med=median(value_dict[lot][center][pop])
batch_med[lot][center][pop]=med

mf=open(med_file,"a")
mf.write(lot+" "+center+" "+pop+" "+str(med)+
        "+str(number_used_for_med)+"\n")
mf.close()

if "1" not in value_dict.keys():
    batch_ref="2"
else:
    batch_ref="1"

print "#####"
print panel
print value_dict.keys()

for center in value_dict[batch_ref].keys():
    for pop in value_dict[batch_ref][center].keys():

        med_ref=value_dict[batch_ref][center][pop]
        med_center_2=value_dict["2"][center][pop]
        med_center_3=value_dict["3"][center][pop]

        # print "#####"
        # print center
        # print pop
        # print med_ref
        # print med_center_2
        if isinstance(med_ref, float) and isinstance(med_center_2, float)
            and isinstance(med_center_3, float):
            coef_batch_2=float(med_ref)/float(med_center_2)
            coef_batch_3=float(med_ref)/float(med_center_3)

            cf=open(coef_file,"a")
            chan=pop.split("_")[1]
            cf.write(center+" "+chan+" "+str(coef_batch_2)+" "+str(
                coef_batch_3)+"\n")
            cf.close()

        panel+=1
    print "END_COMPUTING_COEF_BATCH"

### APPLICATION OF THE COEF BATCH
print "###BEGINING_APPLY_BATCH_COEF###"

for MFI_files in glob.glob("./MFI/CORRECTED_MFI/*.csv"):
    # print MFI_files
    new_MFI_files=MFI_files.replace("CORRECTED_MFI","BATCH_MFI")
    # print MFI_files.split("/")
    CENTER=MFI_files.split("_")[2]
    if "CHAR" in CENTER:
        CENTER="DRFZ"
    elif CENTER=="IDIBEL":
        CENTER="IDIBEL"
    elif CENTER=="IDIBEL":
        CENTER="IDIBEL"
    panel=MFI_files.split("_")[-1].replace(".csv","")
    MFI_file=open(MFI_files,"r")
    new_MFI_file=open(new_MFI_files,"w")

    for ligne in MFI_file:
        ligne=ligne.split("\t")
        date=ligne[5].replace("\n","")
        date=testdate(date)
        if "DATE" not in date and date != '':
            date=date.replace("\",",",",")
            date=datetime.strptime(date, '%Y-%m-%d')
            second_batch=datetime.strptime("2015-10-19", '%Y-%m-%d')
            third_batch=datetime.strptime("2017-04-14", '%Y-%m-%d')
            channel=ligne[3].split("_")[1]
            if "HLA" in ligne[3]:
                channel="linneg_HLADR"
            if date < second_batch:
                coef=1
            else:
                if date < third_batch:
                    coef=get_coef(panel,CENTER,channel,"2015-10-19")
                else:
                    coef=get_coef(panel,CENTER,channel,"2017-04-14")

            if "Lin" in ligne[3] or "lin" in channel:
                coef=1
            if ligne[4]=="NA":
                newMFI="NA"
            else:
                print ligne[4] + "_" + str(coef)
                if coef == None:
                    print panel
                    print channel
                    print MFI_files
                    print date
                # if isinstance(med_ref, float(ligne[4])):
                try:

```

```

                                newMFI=float (ligne [4]) *float (coef)
        except :
            newMFI=ligne [4]
        # else :
        #     newMFI=ligne [4]
        # newMFI=round (newMFI, 2)
    else :
        newMFI=ligne [4]
        text=ligne [0]+ "\t"+ligne [1]+ "\t"+ligne [2]+ "\t"+ligne [3]+ "\t"+str (newMFI)+ "\t"+str
            (date)+ "\n"
        new_MFI_file . write (text)
MFI_file . close ()
new_MFI_file . close ()
print "_END_APPLYING_COEF_BATCH"

```

A.7.3 Script Python de correction de l'effet centre

Script 3: Script Python de calcul et d'application de la correction de l'effet centre sur les fichiers de cytométrie

```
#!/usr/bin/python
# Lucas Le Lann
#12/10/2018

from collections import defaultdict
from numpy import median
from datetime import datetime

import os.path
import glob, os
import numpy as np
import re
import subprocess

### FUNCTION
def get_coef_center(PANEL,CENTER,CHANNEL):
    coefFile = open("./MATRICES/COEF/matrice_panel_"+str(PANEL)+"_coef_"+CENTER+".csv", 'r')
    for line in coefFile.readlines():
        line=line.split(";")
        if CHANNEL == line[0].replace("\n",""):
            slo=line[1].replace("\n","")
            return slo

### end FUNCTION

print "#####BEGINING_COMPUTING_CENTER_COEF#####"
### COMPUTE COEF CENTER
panel=1
panel_nb_wanted=2
while panel<=panel_nb_wanted:
    value_dict={}
    if panel == 1:
        coef_file = "./MATRICES/MEDIANS/matrice_panel_1_med "
        list_cen=("CHP", "DRFZ", "FPS", "IDIBELL", "IRCCS", "KUL", "MHH", "SAS", "UBO", "UCL", "UNIGE")
    elif panel == 2:
        coef_file = "./MATRICES/MEDIANS/matrice_panel_2_med "
        list_cen=("CHP", "DRFZ", "FPS", "IDIBELL", "IRCCS", "KUL", "MHH", "SAS", "UBO", "UCL", "UNIGE")
    else:
        print "Panel_number_not_implemented"

    for cen in list_cen:
        #os.remove(coef_file+cen+".csv")
        f=open(coef_file+cen+".csv","w")
        f.write("CENTER;CHANNEL;POPULATION;MEDIANS;NB_USED\n")
        f.close()

    for panel_csv in glob.glob("./BATCH_MFI/*"+str(panel)+".csv"):
        first_line=True
        site=panel_csv.split("_")[2]

        if "CHAR" in site:
            site="DRFZ"
        if site == "IDIBEL":
            site="IDIBELL"

        for line in open(panel_csv, "r"):
            if first_line:
                OMIC=line.split("\t")[0]
                first_line=False
            else:
                line=line.split("\t")
                pop=line[1]
                ref=line[3]
                value = line[4]
                chan=ref.split("_")[0]

                if site not in value_dict.keys():
                    value_dict[site]={}
                if ref not in value_dict[site].keys():
                    value_dict[site][ref]={}
                if pop not in value_dict[site][ref].keys():
                    value_dict[site][ref][pop]=[]

                if value != "NA" or value != "":
                    try:
                        if float(value) > 1000.0:
                            print "extract_values_greater_than_1000"
                            log_save_file=open("./MATRICES/VALUE_USED
                                /log.csv", "a")
```

```

log_save_file.write(str(panel)+" "+site +
                    "+OMIC+" "+ref+" "+pop+" "+str(
                    value)+"\n")
else:
    value_dict[site][ref][pop].append(float(
        value))
except:
    print str(OMIC)+"_error_in_" + str(value)+ "pop_"
        + pop

center_med=value_dict
for center in value_dict.keys():
    for reference in value_dict[center].keys():
        for population in value_dict[center][reference].keys():
            med=median(value_dict[center][reference][population])
            if "nan" in str(med) or len(value_dict[center][reference][
                population]) <= 5:
                print "missing_median"
            else:
                med_save_file=open("./MATRICES/MEDIANS/matrice_panel_"+
                    str(panel)+"_med_"+center+".csv","a")
                med_save_file.write(center + " "+reference+" "+population
                    + " "+str(med)+" "+str(len(value_dict[center][
                    reference][population]))+"\n")
                med_save_file.close()
                val_save_file=open("./MATRICES/VALUE_USED/matrice_panel_"
                    +str(panel)+"_values_"+center+".csv","a")
                val_save_file.write(center + " "+reference+" "+population
                    + " "+str(med))
                for val in value_dict[center][reference][population]:
                    val_save_file.write(" "+str(val))
                val_save_file.write("\n")
                val_save_file.close()

        panel+=1

coef_center()
print "_END_COMPUTING_COEF_CENTER"
print "_APPLY_COEF_CENTER_"
panel=1
while panel<=panel_nb_wanted:
    for MFI_files in glob.glob("./BATCH_MFI/*"+str(panel)+".csv"):
        new_MFI_files=MFI_files.replace("BATCH_MFI","CENTER_MFI")

        CENTER=MFI_files.split("_")[2]
        if "CHAR" in CENTER:
            CENTER="DRFZ"
        if CENTER == "IDIBEL":
            CENTER="IDIBELL"

        omic_test=MFI_files.split("_")[1]
        omic_test=omic_test.replace("MFI\\", "")
        MFI_file=open(MFI_files,"r")
        new_MFI_file=open(new_MFI_files,"w")
        first_ligne=True

        for ligne in MFI_file:
            if first_ligne:
                text=ligne
                new_MFI_file.write(text)
                first_ligne=False
            else:
                ligne=ligne.split("\t")
                channel=ligne[3]
                print str(panel)+" "+CENTER+" "+channel
                coef_value=get_coef_center(panel,CENTER,str(channel))

                if ligne[4] == "NA" or ligne[4] == "" or ligne[4] == "N/A" :
                    newMFI="NA"
                else:
                    coef_value=coef_value.replace(",",".")
                    newMFI=(float(ligne[4])*float(coef_value))
                    newMFI=round(newMFI,2)

                text=ligne[0]+" \t"+ligne[1]+" \t"+ligne[2]+" \t"+ligne[3]+" \t"+str(
                    newMFI)+" \t"+ligne[5]
                new_MFI_file.write(text)

        MFI_file.close()
        new_MFI_file.close()

    panel+=1

```

Titre : Elaboration d'une procédure standardisée d'harmonisation des données de cytométrie en flux dans le cadre d'études multicentriques.

Mots clés : Cytométrie en flux, bio-informatique, immunologie, standardisation.

Résumé : L'objectif de la thèse est d'assurer la comparabilité des données de cytométrie générées au cours d'une étude multi-centrique. Ce travail de thèse s'inscrit dans le cadre du projet PRECISESADS. Ce projet européen cherche à reclassifier les maladies auto-immunes systémiques en utilisant une approche dite "omique" pour obtenir des signatures biologiques. Cette approche regroupe des outils comme la génomique, la protéomique ... et la cytométrie en flux.

Le nombre important d'individus inclus dans le projet rend obligatoire l'utilisation d'outils informatiques pour automatiser l'analyse des milliers de fichiers de cytométrie obtenus au cours des cinq années du projet.

Les fréquences des populations cellulaires extraites des fichiers sont comparables entre les centres mais l'intensité médiane de fluorescence (IMF) des molécules étudiées ne l'est pas malgré une phase de normalisation des données. La cause de cette non-comparabilité est une combinaison d'un effet lot et d'un effet centre. Ces deux effets peuvent être corrigés à l'aide de coefficients spécifiques. La normalisation et la correction des effets lot et centre par l'élaboration de script R et sous python, permettent d'obtenir des IMF comparables entre les centres. Au final, ce travail de thèse a permis d'établir une nouvelle procédure standardisée utilisable dans tous les projets multi-centriques prospectifs d'analyse de données de cytométrie en flux.

Title : Set up of a standardized procedure to harmonize flow cytometry data as part of multi-center studies.

Keywords : Flow cytometry, bio-informatic, immunology, standardization.

Abstract : The aim of this thesis is to ensure the comparability of flow cytometry data within the context of multi-center studies. This thesis work is part of the PRECISESADS project. This European project seeks to reclassify the systemic autoimmune diseases using "omic" data to find useful biological signatures. This encompasses tools like genomic, proteomic ... and flow cytometry.

The inclusion of numerous individuals in the project make the use of informatics tools a must for the analysis automation of the thousands flow cytometry files obtained during the 5-years period of the project.

Cell populations frequencies extracted from the files are comparable between centers but that is not the case for the median of fluorescence intensities (MFI) of the studied molecules, despite a normalization step. The origin of this incomparability is due to a combination of a batch effect and a center effect. Those two effects can be corrected with specific coefficients. The normalization and correction of both batch and center effect thanks to the elaboration of new R script and python script allow the production of comparable MFI between centers. Overall, this thesis work established a new standardized procedure, efficient for any multi-center projects of flow cytometry data analysis.