

# Multidimensional analysis methods for flow cytometry

**Pushing the boundaries**

Selma van Staveren

The work described in this thesis was performed in the department of Respiratory Medicine and Laboratory of Translational Immunology of the University Medical Center Utrecht, Faculty of Medicine, Utrecht University, The Netherlands with financial support from the Netherlands Organisation for Scientific Research (NWO) in the context of a private public partnership Technology Area COAST, grant number 053.21.112.

ISBN                    978-94-6380-335-9

Cover Design:    Alfred van Rossum  
Lay out:            Proefschriftmaken.nl  
Printed by:        Proefschriftmaken.nl

© Selma van Staveren, 2019, Rotterdam, The Netherlands



# Multidimensional analysis methods for flow cytometry

**Pushing the boundaries**

## Multi-dimensionele analyse methoden voor flowcytometrie

Het verleggen van grenzen  
(met een samenvatting in het Nederlands)

### **Proefschrift**

ter verkrijging van de graad van doctor aan de  
Universiteit Utrecht  
op gezag van de  
rector magnificus, prof.dr. H.R.B.M. Kummeling,  
ingevolge het besluit van het college voor promoties  
in het openbaar te verdedigen op

dinsdag 21 mei 2019 des middags te 2.30 uur

door

**Selma van Staveren**

geboren op dinsdag 12 mei 1987  
te Utrecht

**Promotoren:**

Prof. dr. L. Koenderman

Prof. dr. W.P. de Voogt

**Copromotoren:**

Dr. N. Vrisekoop

Dr. J.J. Jansen

## Table of Contents

<b>Chapter 1</b>	General introduction	<b>7</b>
<b>Chapter 2</b>	A field-applicable method for flow cytometric analysis of granulocyte activation: Cryopreservation of fixed granulocytes	<b>27</b>
<b>Chapter 3</b>	Multi-dimensional flow cytometry analysis reveals increasing changes in the systemic neutrophil compartment during seven days of endurance exercise	<b>51</b>
<b>Chapter 4</b>	Transformation of flow cytometry data with OTflow. A necessity for multivariate analysis	<b>85</b>
<b>Chapter 5</b>	Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm	<b>147</b>
<b>Chapter 6</b>	Unsupervised identification of rare cell subsets by multicolor flow cytometry applying the multivariate algorithm ECLIPSE	<b>231</b>
<b>Chapter 7</b>	Phagocytosis of microplastics leads to accelerated cell death in human neutrophils	<b>291</b>
<b>Chapter 8</b>	Summary & General discussion	<b>325</b>
<b>Appendices</b>	Nederlandse samenvatting	<b>345</b>
	Dankwoord	<b>355</b>
	Curriculum Vitae	<b>363</b>
	List of publications	<b>365</b>

## CHAPTER 1



# General Introduction



To study health and disease, single-cell technologies are often used. Unravelling the heterogeneity between our cells is of key importance for our understanding of embryonic development, the differentiation of cells, homeostasis, and the pathophysiology of disease. Major advancements have resulted in a high-throughput single-cell technologies, measuring a dozen to thousands of characteristics on hundreds to millions of cells, resulting in enormous multi-dimensional datasets. Retrieving all the relevant information from large datasets like these, remains challenging.<sup>1-3</sup> Amongst the techniques creating large single cell datasets are single-cell RNA sequencing<sup>4,5</sup>, high-resolution microscopy<sup>6-9</sup>, and flow cytometry.<sup>10</sup>

Flow cytometry that is the main technique for obtaining single cell data applied in this thesis measures the phenotype of cells and particles in suspension. In flow cytometry less parameters per cell are measured when compared to single cell RNA sequencing but the technique allows the analysis of tens of thousands of cells per second. In addition, flow cytometry is a relatively cheap technique and is easy accessible. Many labs have one or multiple flow cytometers available and these can be operated by researchers themselves after a short training. Also flow cytometry is fast, measuring 30,000 - 100,000 cells per second. Furthermore, the results are visible during measurement, and the data are directly available after measurement. Fluorescently labelled monoclonal antibodies that bind to known receptors on the surface of cells or to proteins inside cells are added to (cells isolated from) the peripheral blood, or, for example the bone marrow or other tissues. Per single cell the fluorescence signal is measured, and from this data the phenotypes of the cells can be determined. Recent developments now allow the determination of an increasing number of variables per single cell. As a result, the datasets obtained from flow cytometry measurements increase their dimensionality. To deal with the increasingly 'deep' datasets, various multidimensional analysis methods have been developed for deep phenotyping of the cells measured.

### **Flow cytometry development**

In 1965 the first flow cytometer was built by Mack Fulwyler,<sup>11</sup> combining the properties of two instruments. The first was a Coulter Counter, which counts and measures the size of particles and cells suspended in a conducting medium.<sup>12</sup> Fulwyler also adapted Richard Sweet's principle behind the technique of the inkjet printer, which utilized electrostatic droplet deflection, mainly designed for oscillography. Droplets of ink were formed from a vibrating nozzle and these were electrostatically charged based on the input signal. The charged droplets passed through electrostatic plates and were thereby positioned on the paper report, which was located underneath the nozzle. The ink jet printer contained a waste gutter for surplus ink droplets.<sup>13</sup> Fulwyler's flow cytometer measured the suspended cells during passage through a Coulter orifice and subsequently the fluid stream would be broken up into droplets containing single cells. Based on cell characteristics like volume,

these droplets could be charged and deflected into the selected collection vessel, after passing between plates maintained at a high voltage of opposing polarities. This enabled the separation of two cell populations based on their characteristics. Notably, after sorting the cells were still viable.<sup>11</sup> The technology behind cell sorting advanced rapidly and in 1968 Wolfgang Göhde designed a microscope-based flow cytometer, which could measure cell-associated fluorescence. This instrument was built to measure and analyse fluorescence-based characteristics of cells and did not enable cell sorting. The cells were discarded after measurement.<sup>14</sup> One year later Herzenberg et al. developed the first fluorescence-activated cell sorter, and coined the popular term 'FACS'. FACS enabled sorting cell populations of interest based on fluorescence properties.<sup>15</sup> The first FACS, FACS-1 (BD Biosciences), could measure cellular light scatter properties as well as fluorescence of wavelengths above 530 nm. Over the past 50 years, together with the development of fluorescently labelled monoclonal antibodies, flow cytometers became more advanced in measuring multiple markers per cell. It is almost unbelievable that a few decades ago only 1 marker per cell was measured, while nowadays multicolour flow cytometry (MFC) panels commonly contain 10-30 markers.

A technical problem which arose when measuring multiple markers simultaneously per cell was spectral overlap. Fluorophores within one panel can be excited by multiple lasers and due to the relatively broad range of the excitation and emission wavelengths, the signal of one fluorophore might be picked up by several detectors. Thus, the signal measured in one fluorescent channel is not solely derived from one fluorophore.<sup>16</sup> When only 2 or 3 markers were measured per cell, this was not problematic, and most researchers compensated for the spectral overlap manually, a process called compensation. As common MFC panels contain 8-17 markers, manual compensation is not feasible anymore and compensation is now performed computationally. Single stain controls are measured to examine the spectral overlap of each fluorophore per channel measured. The percentage of spectral overlap per marker is calculated per channel, resulting in a matrix of  $N \times N$  dimensions,  $N$  being defined as the number of markers measured. In subsequent samples the spectral overlap is subtracted from each channel.<sup>16,17</sup> With increasing numbers of markers per laser in the MFC panels, the emission wavelengths of the different fluorophores overlap more and proper compensation becomes challenging.

Recent advancements in the flow cytometry field include the development of instruments in which spectral overlap is not an issue. A multispectral flow cytometer<sup>18</sup> collects the full spectral information per cell. Spectral unmixing software retrieves the separate emission spectra per fluorophore. Since more spectral information is used to retrieve the fluorescence signals, spectral flow cytometry seems a promising technique which might have less problems regarding compensation or unmixing and autofluorescence. Also,



the detector configuration is fixed and not explicitly tuned to a particular staining panel, which is contributing to more standardised measurements.<sup>18,19</sup>

Mass cytometry is a relatively new platform for deep phenotyping of single cells, which has been developed in 2009.<sup>20</sup> Instead of fluorescently labelled antibodies, for mass cytometry antibodies are tagged with heavy metals. Using time of flight mass spectrometry (TOF-MS) the presence of these metals can be measured at a single-cell level. Since many stable isotopes can be used, over 40 markers can be measured per cell. Also the signal to noise ratio is lower compared to flow cytometry.<sup>20,21</sup> However, the CyTOF measures cells at a slower rate of 500 cells per second, the instrument has a lower efficiency due to loss of cells during measurement, considerable instrument variability has been reported over time and consumables and parts are more expensive compared to flow cytometry.<sup>21,22</sup> For this thesis, I focussed on flow cytometry, but it is good to emphasize that the analysis algorithms are not restricted to flow cytometry datasets, as they are not restricted in dimensions. Thus, these multivariate analysis methods were used for the analysis of flow cytometry datasets as proof-of-principle, but they are also applicable to e.g. by CyTOF obtained datasets.

## Analysis of flow cytometric data

Flow cytometry can be used to answer a variety of research questions, as multi-variate comparison of cell phenotypes might be studied at different levels: cells within one individual, between individuals, between groups of individuals or within one or multiple individuals over time.

One of the main interests of our lab is to study the role neutrophils play in innate immune responses and in inflammatory diseases. Neutrophils are phagocytes able to engage, phagocytose and kill targets, such as bacteria and yeasts, that are smaller in size than the cell itself.<sup>23–25</sup> In contrast to the suggestion of older studies describing neutrophils as belonging to a short-lived (7–9 hrs)<sup>26</sup> and homogenous population of cells, the cells live longer (3–5 days)<sup>27</sup> and there is a clear heterogeneity in function between different subsets of cells.<sup>24,28–36</sup> For instance, using flow cytometry, our lab has previously identified at least three distinct neutrophil subsets, based on differential expression of FcγRIII (CD16) and L-selectin (CD62L). These three subsets were sorted using FACS, enabling subsequent functional assays and microscopic analysis. The subsets had morphological differently shaped nuclei and different functionalities.<sup>35</sup> In homeostasis one homogenous neutrophil population is present in the human blood, expressing both CD16 and CD62L to high levels, with a segmented nucleus. Interestingly, acute inflammation that can be evoked by administering lipopolysaccharide (LPS) intravenously to healthy subjects leads to an acute systemic neutrophil response that is associated with two additional neutrophil subtypes found in the human blood. Within 3 hours immature CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils, with a band-shaped nucleus, are recruited from the bone marrow.<sup>35</sup> These cells have a very good

bacterial killing capacity.<sup>33</sup> In the same timeframe CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophils appear in the peripheral blood. These cells have a hypersegmented nucleus and, in contrast to banded neutrophils, they possess T-cell suppressive properties<sup>35</sup> and are relatively poor bacterial killers.<sup>33</sup> These studies have broadened our view about neutrophils from short-lived phagocytosing cells to longer-lived cells with multiple functions.

Next to LPS-induced neutrophil heterogeneity, our lab also found CD16<sup>dim</sup>CD62L<sup>bright</sup> and CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophil phenotypes in patients submitted at the Intensive Care unit after severe trauma (Leliefeld et al, submitted). CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophils were also found in the peripheral blood in athletes during a short bout of anaerobic rowing exercise. (Hilvering et al, unpublished) Furthermore, the CD16<sup>dim</sup>CD62L<sup>bright</sup> and CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophil subsets were found in patients with active systemic-onset juvenile idiopathic arthritis<sup>37</sup> and upon acute sleep deprivation in healthy young men.<sup>38</sup>

### **Challenges in flow analysis: the issue of point-of-care vs the laboratory setting**

Aiming to extend our knowledge on neutrophil subsets after exercise and severe trauma we are faced with technical challenges. The analysis cannot be performed in a point-of-care context as the laboratory is far away when studying an exercise challenge outdoor and blood is drawn at different time points during the challenge. Similar problems are encountered when blood is drawn from trauma patients at the location of the accident.<sup>39</sup> Neutrophils are difficult to handle and extremely sensitive for ex vivo handling particularly in a non-laboratory setting. Peripheral blood neutrophils become easily activated outside the body, even when the blood remains untouched in a tube at room temperature (Spijkerman et al, unpublished). Therefore, measurements should be conducted quickly after drawing blood samples. This is problematic during fieldwork with relatively primitive equipment available. And unfortunately, unlike lymphocytes, neutrophils are not viable after freezing and thawing. Tackling these problems would allow many other studies of neutrophils in the field, even in developmental countries. Laboratories in developmental countries might not be fully equipped, e.g. lacking flow cytometers. Yet, specific diseases can be studied which are not prevalent in developed nations. So there is a clear unmet need to either transfer the technology to a point-of-care context and ship adequately fixed samples to the flow laboratory.

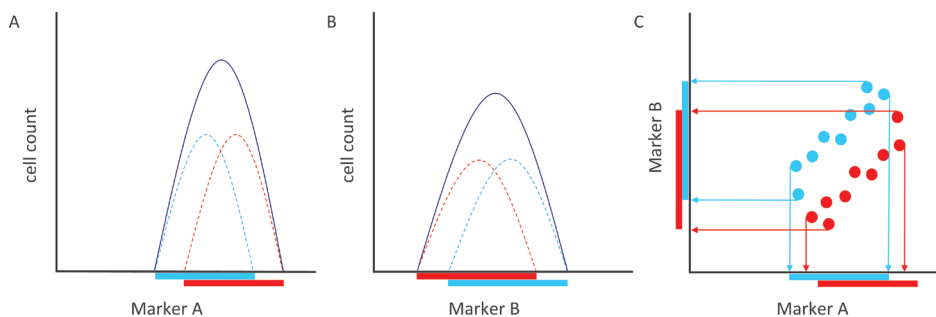
### **Challenges in flow analysis: intuitive diagnostics from a multidimensional data set.**

Apart from the importance of flow cytometry in research, it is daily used in clinical diagnostics departments to diagnose and monitor disease. A nice illustration is the importance of this technique in the identification of minimal residual disease (MRD) in patients treated for various types of lymphomas or leukemias. Upon treatment for

lymphoma or leukemia, malignant cells in the bone marrow may decrease. In some patients a small proportion (0.1-0.01%) of residual malignant cells (MRD cells) may be found. For various types of lymphomas and leukemias the number of MRD has been found to positively correlate to risk of relapse of the disease and therefore MRD detection has become increasingly important in monitoring and treatment protocols.<sup>40-44</sup> With increasing numbers of markers measured, flow cytometry contributes more often to the identification of relatively rare disease-specific cell types. The identification of rare cells is challenging, as they might become overshadowed by a vast majority of normal cells present in the sample. There is still much room for improvements in flow cytometry analysis and of the identification of rare cells. Current manual analysis of the data is time consuming and inter-observer variability has been reported.<sup>45</sup> Especially in a clinical diagnostics setting, this is undesirable. In the next section the drawbacks of conventional manual analysis will be explained in more detail.

### **Conventional manual analysis of flow cytometry data**

With the advances in flow cytometry instruments and fluorophore based technology, MFC datasets increase in depth. Conventional flow cytometry analysis consists of 1D (plotting the fluorescent intensity of a certain marker on the x-axis against the cell count on the y-axis), 2D (plotting the intensities of two markers against each other on X- and Y-axes) and sometimes 3D analysis (plotting the intensities of three markers, on X-, Y- and Z-axes). If more than three parameters are measured, this type of analysis misses out on important information hidden in the unknown correlation between cells. This cryptic statement can be explained by the often found situation that expression levels partly overlap. In this example, separation of distinct cell populations by 1D analysis becomes difficult as these populations have broad overlapping fluorescent intensities for the single markers (Figure 1A, B). However, 2D analysis can help as the correlations between two markers are explored by a scatter plot showing the expression of one variable on the x-axis against a second variable on the y-axis per single cell, such as shown in Figure 1C. The expressions for both markers in 1D might not be significantly different, since they greatly overlap, but from a 2-dimensional scatter plot the two populations can be identified as separate. It will be clear that this problem increases when more markers are used and that multidimensional analysis methods (> 2D) will become increasingly important for the identification and characterization of cellular subsets.



**Figure 1. Correlations between multiple markers are essential for finding differences between groups**

Schematic representation of the expression of a cellular sample (purple line), consisting of two cell populations (red and blue dashed line), for marker A (A) and B (B) plotted in a histogram plot, or (C) plotted in a bi-dimensional scatter plot. Analysing the histograms (in A+B) it might seem as if there is only one homogeneous cell population present in the dataset, while in the scatter plot (C) two separate cell populations are identified.

Typically, in conventional sequential gating analysis cell populations of interest are manually selected, or “gated”, and further explored in histogram plots or bi-dimensional scatter plots. Subsequent gating steps are conducted to exclude cells that are not of interest and eventually the population of interest is reached. This has important disadvantages. Manual analysis is subjective and biased, as it is based on the knowledge and interests of the researcher analysing the data. Commonly, part of the cells in the sample are removed during subsequent gating steps, based on the presence or absence of specific markers chosen by the researcher beforehand. This leads to bias caused by the preconceived gating strategy. Importantly, the dataset might contain essential unknown information about putative correlations between cells that were removed on the basis of insufficient knowledge (see also the explanation in figure 1). The larger the multi-colour flow cytometry (MFC) panel, the more difficult it becomes to extract all the relevant information from the dataset. Also, individual researchers or research groups use different antibodies and a different order for the subsequent steps in the gating strategy. This makes it hard to reproduce findings between research groups. Last but not least, manual analysis is time-consuming, especially when the dataset consists of many samples, since the gating has to be checked for every sample individually.

Aiming to make flow cytometry a more standardised and reliable technique, the EuroFlow Consortium has developed multiple methods to measure and analyse patient samples using flow cytometry for diagnostics and prognostic (sub)classification of hematological

malignancies, as well as for evaluation of treatment effectiveness during follow-up. This includes the development of standard operating procedures, antibody panels and (automated) gating software “Infinicyt”.<sup>46–49</sup> Using Infinicyt manual analysis can be performed, as well as automated gating of patient samples suspected for Multiple Myeloma MRD, provided that the researcher used the specific EuroFlow antibody panels and employed the EuroFlow standard operating procedure for the measurement of cells.

### Multidimensional analysis of flow cytometry data

The increasing size and complexity of datasets and problems of manual analysis led to the need of automated unsupervised multi-dimensional analysis algorithms. By exploring the correlations between more than two parameters simultaneously, more cell populations may be distinguished from each other.

First **SPADE** (spanning tree progression of density normalized events)<sup>50</sup> was developed, which is a hierarchical cluster method that groups similar single cells together in a cluster which are visualised in a minimal spanning tree (MST). Generally, clusters with similar phenotypes are mapped closer to each other. As a third dimension a colour heatmap can be superimposed onto the SPADE clusters, depicting the relative expressions of a single marker. The disadvantage of SPADE is the loss of the single cell structure. Also the SPADE map does not show abundance of cells. The algorithm attempts to equalize the size of all clusters by down-sampling based on density. The advantage of this down-sampling is that also relatively rare cell populations should be visualised.<sup>50</sup> However, the abundance of different cell populations is very relevant information and now lost during analysis. Furthermore, the down-sampling step is very time-consuming. Another important disadvantage is caused by the minimal spanning tree visualisation. A circular branch of similar cells might be split in two branches, which implies that the clusters found in the two branches differ from each other, while they might contain similar cells.

Another recently developed cluster method is **flowSOM**.<sup>51</sup> Visually, the flowSOM result resembles SPADE, as the clusters are also visualised in a minimal spanning tree. However, there are important differences between the 2 cluster methods. FlowSOM employs self-organizing maps (SOM) for clustering, instead of hierarchical clustering. In contrast to SPADE, subsampling is not needed, which saves analysis time. In short, a predefined number of clusters (or nodes) is generated and from the dataset a random data point is assigned to the multidimensional node. One by one additional data points are assigned to the node nearest to the data point. Cell abundance is visualised in the model by the size of the nodes. Prior to flowSOM analysis the number of nodes has to be chosen. It is recommended to choose a high number of nodes (higher than the number of cell populations expected), in order for the algorithm to cluster all cell types in separate clusters. And when expecting rare cells, even a higher number has to be chosen of 200-

300 clusters. In flowSOM, the relative intensities of certain markers are indicated with the use of a piechart for each cluster.<sup>52</sup>

**viSNE**<sup>53</sup> is a dimensionality reduction technique that uses the t-stochastic neighbour embedding (t-SNE) algorithm. A two-dimensional plot is created showing single cells. Cells with similar high-dimensional properties are placed close to each other in the viSNE plot. Similar to SPADE, as a third dimension a colour heat map can be projected onto the viSNE plot, showing the relative expressions for a certain marker, or showing which sample the individual cells belong to. An important disadvantage of the viSNE algorithm is the necessity to down-sample the data. This is done to decrease the computational burden and analysis time and to prevent over-crowding of the viSNE plot.<sup>53</sup> When too many cells are projected in the map, no information can be retrieved from the position of the cells, as the map shows one amorphous population, without any structure. Contour-map visualisation is possible, based on density. However it is advised to down-sample the data to 30,000 cells.<sup>53</sup> This is problematic for rare cell discovery. Imagine a cell population which comprises of 0.01% of the total cells such as typically found for minimal residual disease. After down-sampling, only 3 cells of this population are left. For reliable MRD detection in AML at least 20 cells with the same aberrant phenotype should be found,<sup>40</sup> this means that the lower limit of MRD detection in viSNE is 0.67%, which is far above the 0.01% limit which can be reached by manual gating.<sup>44</sup> Down-sampling is also not favourable for the analysis of a dataset with multiple heterogenous samples, as less cells per sample and thus per cell population can be included in the analysis.

In tSNE, after creating a low-dimensional t-SNE map, an optimization step is performed. The difference in distance between 2 points in the high dimensional space vs. the low dimensional t-SNE map are measured. The position of the data-points is adjusted until the difference between the high-dimensional and low-dimensional distances is lowest. This optimization step is initiated randomly, so the difference between distances is analysed in a random order, and as a result different runs give different solutions.<sup>53</sup> It is advised to run viSNE multiple times per dataset and choose the visualisation with the lowest value for the objective function.<sup>54</sup> The fact that this stochastic method results in differences per run may be statistically correct, however it is time-consuming to perform multiple analyses per dataset and also very impractical that previously obtained results cannot be regenerated if the analysis was not saved. It is important to emphasize that cells which are clustered together are most similar to each other, however the relative distance between clusters in the tSNE map does not have a meaning.

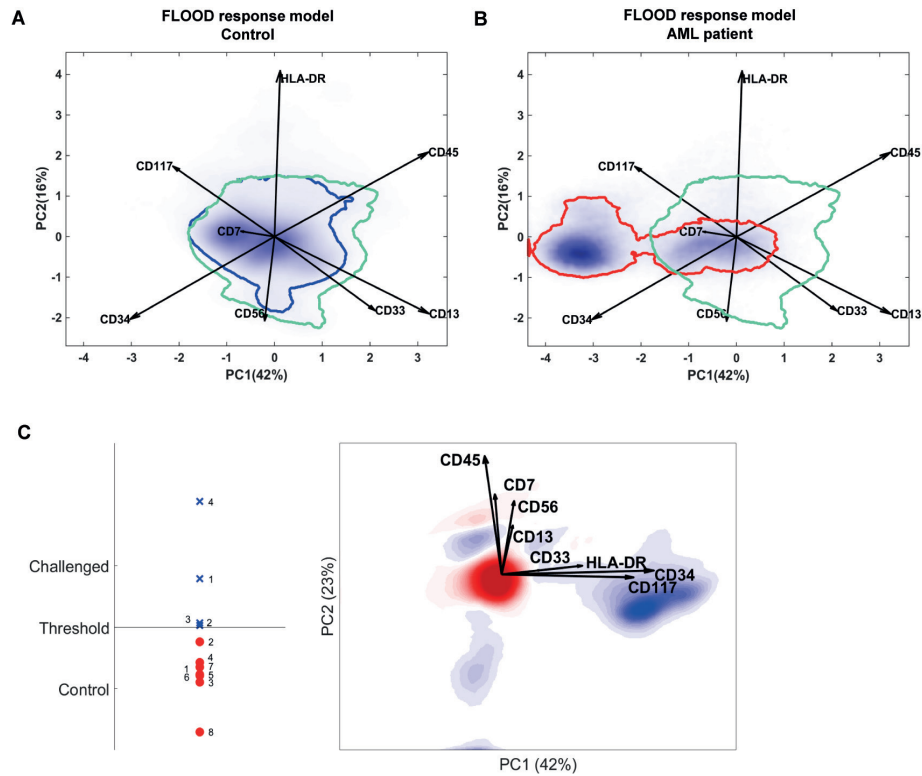
Next, it is impossible to project new samples in an existing viSNE model. When new samples are measured a totally new viSNE model has to be generated. Therefore, implementing viSNE in a clinical diagnostics setting for fast high-throughput analysis is

difficult, as with a new sample measured, a new viSNE model needs to be generated and thoroughly analysed. In the same line, it is also impossible to perform FACS sorting using tSNE as with every new incoming event, the tSNE space changes and the algorithm is not fast enough: before the cell can be assigned to a certain population in the tSNE space and the droplet can be charged accordingly, the droplet already ended up in the waste bucket.

To overcome the aforementioned SPADE and tSNE-related problems, our lab has developed **FLOOD**,<sup>55</sup> a dimensionality reduction method, using principal component analysis (PCA),<sup>56,57</sup> to subtract the high-dimensional disease specific single cell information and project this into two-dimensional plots. Samples are imported and a control group and a patient/ response group are defined. First a control model is generated to describe the variance within the control samples. The response or patient samples are then projected in the control model. The algorithm was constructed such that the control model does not explain the patient-specific or response-specific information. Therefore, the variance described within the control model is subtracted from the dataset, and subsequently a new patient-specific or response-specific PCA model is generated, the (e.g. disease specific) response model. The PCA loadings are visualised as vectors. The single cells are shown and the relative expressions of the various markers can be deduced from the position of the cells and the direction of the marker loadings, as shown in an example in Figure 2. FLOOD enables visualisation of the multi-dimensional expression levels of single-cells in a biplot and allows comparison between 2 groups of patients or time points. The distance between cell populations in the PCA model reflect the corresponding variance between the phenotypes of the cell subsets. An important advantage of FLOOD lies in the fact that PCA based analysis methods are fast, large datasets can be analysed without the need for downsampling and performing an analysis on a specific dataset multiple times, leads to identical results. Also, new data-points and even additional samples can be projected into an existing PCA model. Therefore, real-time FACS sorting is a real possibility using a multidimensional PCA analysis. An important disadvantage of FLOOD is that rare cell subsets might be missed due to the overwhelming abundance of other cell subsets in the dataset, and the variance of the latter will be more readily described within the first PC's. To circumvent this problem we have developed the ECLIPSE algorithm (see **Chapter 5** of this thesis).

Classification methods are another type of multi-dimensional method. DAMACY<sup>58</sup> is a PCA-based method, classifying individual samples as belonging to the pre-defined control or patient/responder group, based on the multi-dimensional similarities in marker expressions between samples. An example is shown in Figure 2C. The PCA space on the right shows which cellular phenotypes are more predominant in controls (red regions) or in patients (blue regions). Based on this difference in distribution of cells, performing cross-validation by taking one sample out sequentially, DAMACY predicts whether this

specific sample fits best in the control or patient group. As a result the accuracy, sensitivity and specificity of the cross-validated classification are calculated. Also a PCA-space shows which cellular phenotypes are more predominant in patients or controls.



**Figure 2. FLOOD and DAMACY analysis of bone marrow samples from healthy controls and an AML patients**

Multidimensional analysis of an AML dataset consisting of bone marrow samples of 8 healthy donors and 4 AML patients. PC1 and PC2 are plotted on the axes and the percentages (in brackets) show the % variance explained by each PC. In both plots, the cyan control benchmark encompasses 80% of the cells of all 8 healthy controls. The marker loadings are represented by arrows. In A) the FLOOD response space shows the cells (blue density plot) of a representative healthy bone marrow sample and the dark blue border encompasses 80% of the cells of this specific control. In B) the cells (blue density plot) of a bone marrow sample of an AML patient are projected. The red benchmark encompasses 80% of the cells of this specific AML patient. The marker loadings are represented by arrows. In B, the cells positioned outside of the control benchmark might be identified as leukemic cells and their expression profile can be derived from the direction of the loadings as  $CD34^{bright}CD117^{bright}CD7^{bright}CD13^{low}CD33^{low}CD45^{low}$ . C) DAMACY model of the same AML dataset. The right panel shows the DAMACY cell map of which relatively more cells of the control group are located in the red regions, while the cells of the AML group are relatively more represented by the blue regions. The longest loadings are CD34, CD117 and HLA-DR, pointing towards a blue cell population, mostly represented in AML patients. On the left the result of the cross-validation is depicted. The accuracy of the DAMACY model is 100%, as all samples have been correctly classified as control (red) or patient (blue).



### **Challenges in multidimensional analysis algorithms: rare cell detection**

Although automated multidimensional analysis methods are a great aid for unbiased and fast analysis of flow cytometry data, some challenges remain. All the above mentioned analysis methods are not specifically suited for the identification of rare cell subsets, as the variance between the predominant cell populations are dominating the analysis. Apart from the increased focus on major cell populations, the clustering in SPADE<sup>50</sup> and the down-sampling which is required for viSNE analysis,<sup>53</sup> are disadvantageous for the identification of rare cells. Rare cells might be clustered together with other cells in SPADE and down-sampling in viSNE may lead to decreased resolution or even to the unintentional disappearance of the potentially interesting rare cells.

### **Challenges in multidimensional analysis algorithms: acquisition and pre-processing of flow data**

Because multidimensional analysis describes the cells in such detail, standardized operating procedures are even more important than for manual analysis. It is very common that subtle changes in the definition of gates are used during manual analysis. These changes can occur by not following strict standardised operating procedures (SOPs) for cell isolation and antibody staining. These SOP's are essential for MVA and have been published.<sup>41,44,46,59</sup> However, when it comes to the multivariate analysis of flow cytometry datasets, reproducibility is an important recurring issue.

For instance, the choices in pre-processing options such as the type of transformation,<sup>17,60,61</sup> centring and scaling,<sup>62,63</sup> contribute to the above mentioned differences. Unfortunately, pre-processing of flow cytometry data is not a popular topic amongst the users of multidimensional methods. For reports showing multidimensional flow cytometry results, it is rather essential to know whether various pre-processing options have been explored and/or why a specific type of pre-processing was chosen. Generally this remains elusive as the whole issue is not discussed. At most, the used pre-processing steps are only shortly mentioned in the methods section. It is worth noting that multidimensional analysis results and the immunological conclusions may differ considerably when applying various pre-processing options prior to analysis, as Folcarelli et al experienced when applying different combinations of scaling and mean centring (Folcarelli et al, submitted).

## Thesis outline

The aim of this thesis was to push the boundaries of flow cytometry in terms of field applications and multidimensional analysis.

In **Chapter 2** we have developed a new method to isolate, fix and freeze neutrophils for enabling sampling in the field and later flow cytometry analysis in the laboratory.

The cryopreservation method was applied to study the exercise-induced neutrophil response during an intensive 8-day cycling tour, Tour for Life, from Italy to The Netherlands (**Chapter 3**). We studied the exercise-induced neutrophil phenotypes and examined the recovery of the neutrophil compartment in particular. We hypothesized that various neutrophil subsets might be recruited to the blood, as was found upon acute inflammation experimentally evoked by an LPS challenge. Using FLOOD analysis we multi-dimensionally compared the different degranulation and immune suppressing markers of exercise-induced neutrophils with the neutrophil compartment after an LPS challenge.

Preprocessing flow cytometry data has a great impact on subsequent results and conclusions. Flow cytometry data is transformed prior to multi-dimensional analysis, and the transformation parameter is commonly set by default. **Chapter 4** describes how arbitrary choices for the transformation parameters can have misleading effects on results obtained from conventional flow cytometry analysis as well as on multi-dimensional flow cytometry methods.

As described above, multi-dimensional analysis methods are promising techniques to subtract the information from large flow cytometry datasets. After the development of FLOOD<sup>55</sup> the collaboration of the Koenderman lab and Jansen group continued, with the aim of finding the needle (e.g. rare cell) in the haystack (e.g. total blood). In **Chapter 5** we present the new computational method ECLIPSE, which focusses on disease-specific cells by eliminating the normal-like cells from the dataset. Subsequently, **Chapter 6** shows how ECLIPSE is specifically suited for the identification of scarce cells, using a multiple myeloma dataset containing samples with Minimal Residual Disease (MRD) as a proof of principle.

As a member of the IMPACT consortium multiple private and academic partners with different fields of expertise were gathered to tackle the urgent societal question of the impact of environmental microplastics on human health. The consortium consisted of the department of Analytical Chemistry of the Radboud Univeriteit Nijmegen, the LaserLaB of the Vrije Universiteit Amsterdam, Rijkswaterstaat, CytoBuoy, KWR, Deltares and Beckman Coulter. The research project consisted of two main questions. First, aquatic particle detection. One of our partners is CytoBuoy, a company which manufactures

flow cytometry instruments for rapid analysis of particles, like algae, in the aquatic environment. In the meantime, our collaborators from the LaserLaB of the VU Amsterdam developed a technique to isolate environmental plastic particles from water samples and to measure these plastic particles using stimulated Raman scattering.<sup>64</sup> The Jansen group from the Radboud University developed different multi-dimensional analysis techniques which could be applied to the data generated from flow cytometry or stimulated Raman scattering.<sup>58,64</sup> Secondly we aimed to investigate the impact of microplastics on marine and human health. We hypothesized that if microplastics were to invade the human body, innate immune cells would phagocytose the microplastics much like foreign invaders and that the innate cells would have difficulty dealing with these particles. As a first proof of principle, we used standardized manufactured microplastics with a fluorescent core and performed in vitro experiments studying the effects of these particles on the phagocytic properties and lifespan of human peripheral blood neutrophils in **Chapter 7**.

## References

1. Patel D V., McGhee CN. Quantitative analysis of in vivo confocal microscopy images: A review. *Surv Ophthalmol.* 2013;58(5):466-475. doi:10.1016/j.survophthal.2012.12.003
2. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet.* 2015. doi:10.1038/nrg3833
3. Bacher R, Kendzierski C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 2016;17(1):1-14. doi:10.1186/s13059-016-0927-y
4. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.* 2011;12:87-98. doi:10.1038/nrg2934
5. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol.* 2018;18(1):35-45. doi:10.1038/nri.2017.76
6. Cornett DS, Reyzer ML, Chaurand P, Caprioli RM. MALDI imaging mass spectrometry: Molecular snapshots of biochemical systems. *Nat Methods.* 2007;4(10):828-833. doi:10.1038/nmeth1094
7. Vaysse PM, Heeren RMA, Porta T, Balluff B. Mass spectrometry imaging for clinical research-latest developments, applications, and current limitations. *Analyst.* 2017;142(15):2690-2712. doi:10.1039/c7an00565b
8. Lubeck E, Cai L. Single cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods.* 2013;9(7):743-748.
9. Beghin A, Kechkar A, Butler C, et al. Localization-based super-resolution imaging meets high-content screening. *Nat Methods.* 2017;14(12):1184-1190. doi:10.1038/nmeth.4486
10. Chattopadhyay P, Perfetto S, Gaylord B, et al. Toward 40+ Parameter Fluorescence Flow Cytometry. In: *XXIX Congress of the International Society for Advancement of Cytometry.* ; 2014.
11. Fulwyler MJ. Electronic separation of biological cells by volume. *Science (80-).* 1965. doi:10.1126/science.150.3698.910
12. DeBlois RW, Bean CP. Counting and sizing of submicron particles by the resistive pulse technique. *Rev Sci Instrum.* 1970. doi:10.1063/1.1684724
13. Sweet RG. High frequency recording with electrostatically deflected ink jets. *Rev Sci Instrum.* 1965;36(2):131-136. doi:10.1360/zd-2013-43-6-1064
14. Ditttrich W, Göhde W. Impulsfluorometrie bei Einzelzellen in Suspensionen. *Zeitschrift Naturforsch.* 1969;21(b):360-361.
15. Hulet HR, Bonner WA, Barrett J, Herzenberg LA. Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. *Science.* 1969;166(3906):747-749. doi:10.1126/science.166.3906.747
16. Roederer M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. *Cytometry.* 2001;45(3):194-205. <http://www.ncbi.nlm.nih.gov/pubmed/11746088>.
17. Tung JW, Parks DR, Moore WA, Herzenberg LA, Herzenberg LA. New approaches to fluorescence compensation and visualization of FACS data. *Clin Immunol.* 2004. doi:10.1016/j.clim.2003.11.016
18. Nolan JP, Condello D. Spectral flow cytometry. *Curr Protoc Cytom.* 2013;(SUPPL.63):1-13. doi:10.1002/0471142956.cy0127s63

19. Feher K, von Volkmann K, Kirsch J, Radbruch A, Popien J. Multispectral flow cytometry: The consequences of increased light collection. *Cytom Part A*. 2016;89(7):681-689. doi:10.1002/cyto.a.22888
20. Baranov VI, Lou X, Bandura DR, et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal Chem*. 2009;81(16):6813-6822. doi:10.1021/ac901049w
21. Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. *Cell*. 2016;165(4):780-791. doi:10.1016/j.cell.2016.04.019
22. Bendall SC, Simonds EF, Qiu P, et al. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science (80- )*. 2011;332(6030):687-696. doi:10.1126/science.1198704
23. Metchnikoff E. *Lectures on the Comparative Pathology of Inflammation*. Vol 1. Truber and Co., London, 1893; 1892. doi:10.1136/bmj.1.3197.581-c
24. Kolaczowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol*. 2013;13(3):159-175. doi:10.1038/nri3399
25. Gazendam RP, van de Geer A, Roos D, van den Berg TK, Kuijpers TW. How neutrophils kill fungi. *Immunol Rev*. 2016;273(1):299-311. doi:10.1111/imr.12454
26. Dancy JT, Deubelbeiss KA, Harker LA, Finch CA. Neutrophil kinetics in man. *J Clin Invest*. 1976;58(3):705-715. doi:10.1172/jci108517
27. Pillay J, Den Braber I, Vrisekoop N, et al. In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. *Blood*. 2010. doi:10.1182/blood-2010-01-259028
28. Kamp VM, Pillay J, Lammers J-WJ, Pickkers P, Ulfman LH, Koenderman L. Human suppressive neutrophils CD16bright/CD62Ldim exhibit decreased adhesion. *J Leukoc Biol*. 2012;92(5):1011-1020. doi:10.1189/jlb.0612273
29. Pillay J, Tak T, Kamp VM, Koenderman L. Immune suppression by neutrophils and granulocytic myeloid-derived suppressor cells: Similarities and differences. *Cell Mol Life Sci*. 2013;70(20):3813-3827. doi:10.1007/s00018-013-1286-4
30. Hellebrekers P, Hietbrink F, Vrisekoop N, Leenen LPH, Koenderman L. Neutrophil functional heterogeneity: Identification of competitive phagocytosis. *Front Immunol*. 2017;8(NOV). doi:10.3389/fimmu.2017.01498
31. Gallin JI. Human neutrophil heterogeneity exists, but is it meaningful? *Blood*. 1984;63(5):977-983. doi:dx.doi.org
32. Paíno T, Sarasquete ME, Paiva B, et al. Phenotypic, Genomic and Functional Characterization Reveals No Differences between CD138++ and CD138low Subpopulations in Multiple Myeloma Cell Lines. *PLoS One*. 2014;9(3):e92378. doi:10.1371/journal.pone.0092378
33. Leliefeld PHC, Pillay J, Vrisekoop N, et al. Differential antibacterial control by neutrophil subsets. *Blood Adv*. 2018;2(11):1344-1355. doi:10.1182/BLOODADVANCES.2017015578
34. Tak T, Wijten P, Heeres M, et al. Human CD62Ldimneutrophils identified as a separate subset by proteome profiling and in vivo pulse-chase labeling. *Blood*. 2017;129(26). doi:10.1182/blood-2016-07-727669

35. Pillay J, Kamp VM, van Hoffen E, et al. A subset of neutrophils in human systemic inflammation inhibits T cell responses through Mac-1. *J Clin Invest*. 2012;122(1):327-336. doi:10.1172/JCI57990
36. Marini O, Costa S, Bevilacqua D, et al. Mature CD10+ and immature CD10- neutrophils present in G-CSF-treated donors display opposite effects on T cells. 2017;129(10):1343-1357. doi:10.1182/blood-2016-04-713206.The
37. ter Haar NM, Tak T, Mokry M, et al. Reversal of Sepsis-Like Features of Neutrophils by Interleukin-1 Blockade in Patients With Systemic-Onset Juvenile Idiopathic Arthritis. *Arthritis Rheumatol*. 2018;70(6):943-956. doi:10.1002/art.40442
38. Christoffersson G, Vågesjö E, Pettersson US, et al. Acute sleep deprivation in healthy young men: Impact on population diversity and function of circulating neutrophils. *Brain Behav Immun*. 2014;41(1):162-172. doi:10.1016/j.bbi.2014.05.010
39. Hazeldine J, Naumann DN, Toman E, et al. Prehospital immune responses and development of multiple organ dysfunction syndrome following traumatic injury: A prospective cohort study. *PLoS Med*. 2017;14(7):1-29. doi:10.1371/journal.pmed.1002338
40. Grimwade D, Freeman SD. Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for "prime time"? *Blood*. 2014;124(23):222-233. doi:10.1182/blood-2014-05-577593.This
41. Ikoma MRV, Beltrame MP, Ferreira SIACP, et al. Proposal for the standardization of flow cytometry protocols to detect minimal residual disease in acute lymphoblastic leukemia. *Rev Bras Hematol Hemoter*. 2015;37(6):406-413. doi:10.1016/j.bjhh.2015.07.012
42. Landgren O, Devlin S, Boulad M, Mailankody S. Role of MRD status in relation to clinical outcomes in newly diagnosed multiple myeloma patients: a meta-analysis. *Bone Marrow Transplant*. 2016;51(12):1565-1568. doi:10.1038/bmt.2016.222
43. Paiva B, Cedena M, Puig N, et al. Minimal residual disease monitoring and immune profiling in multiple myeloma in elderly patients. 2016;127(25):3165-3175. doi:10.1182/blood-2016-03-705319.The
44. Schuurhuis GJ, Heuser M, Freeman S, et al. Minimal/measurable residual disease in AML: consensus document from ELN MRD Working Party. *Blood*. 2018;131(12):blood-2017-09-801498. doi:10.1182/blood-2017-09-801498
45. Wood BL. Principles of minimal residual disease detection for hematopoietic neoplasms by flow cytometry. *Cytom Part B - Clin Cytom*. 2016;90(1):47-53. doi:10.1002/cyto.b.21239
46. Van Dongen JJM, Lhermitte L, Böttcher S, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia*. 2012;26(9):1908-1975. doi:10.1038/leu.2012.120
47. Flores-Montero J, Sanoja-Flores L, Paiva B, et al. Next Generation Flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. *Leukemia*. 2017;31(10):2094-2103. doi:10.1038/leu.2017.29
48. Pedreira CE, Costa ES, Barrena S, et al. Generation of flow cytometry data files with a potentially infinite number of dimensions. *Cytom Part A*. 2008;73(9):834-846. doi:10.1002/cyto.a.20608

49. Costa ES, Pedreira CE, Barrena S, et al. Automated pattern-guided principal component analysis vs expert-based immunophenotypic classification of B-cell chronic lymphoproliferative disorders: A step forward in the standardization of clinical immunophenotyping. *Leukemia*. 2010;24(11):1927-1933. doi:10.1038/leu.2010.160
50. Qiu P, Simonds EF, Bendall SC, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol*. 2011;29(10):886-893. doi:10.1038/nbt.1991
51. Van Gassen S, Callebaut B, Van Helden MJ, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytom Part A*. 2015;87(7):636-645. doi:10.1002/cyto.a.22625
52. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: Helping to make sense of high-dimensional immunology data. *Nat Rev Immunol*. 2016;16(7):449-462. doi:10.1038/nri.2016.56
53. Amir ED, Davis KL, Tadmor MD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol*. 2013;31(6):545-552. doi:10.1038/nbt.2594
54. Van Der Maaten L. tSNE FAQ. <https://lvdmaaten.github.io/tsne/#faq>. Published 2019.
55. Jansen JJ, Hilvering B, van den Doel A, et al. FLOOD: FLOW cytometric Orthogonal Orientation for Diagnosis. *Chemom Intell Lab Syst*. 2016;151:126-135. doi:10.1016/J.CHEMOLAB.2015.12.001
56. Jolliffe IT. *Principal Component Analysis. Second Edition.*; 2002. doi:10.2307/1270093
57. Timmerman ME, Kiers HAL. Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*. 2003. doi:10.1007/BF02296656
58. Tinnevelt GH, Kokla M, Hilvering B, et al. Novel data analysis method for multicolour flow cytometry links variability of multiple markers on single cells to a clinical phenotype. *Sci Rep*. 2017;7(1). doi:10.1038/s41598-017-05714-1
59. Orozco AF, Lewis DE. Flow cytometric analysis of circulating microparticles in plasma. *Cytom Part A*. 2010;77(6):502-514. doi:10.1002/cyto.a.20886
60. Parks DR, Roederer M, Moore WA. A new "logicle" display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. *Cytom Part A*. 2006. doi:10.1002/cyto.a.20258
61. Finak G, Perez J-M, Weng A, Gottardo R. Optimizing transformations for automated, high throughput analysis of flow cytometry data. *BMC Bioinformatics*. 2010;11(1):546. doi:10.1186/1471-2105-11-546
62. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*. 2006. doi:10.1186/1471-2164-7-142
63. Novo D, Wood J. Flow cytometry histograms: Transformations, resolution, and display. *Cytom Part A*. 2008;73(8):685-692. doi:10.1002/cyto.a.20592
64. Zada L. Fast microplastics identification with stimulated Raman scattering microscopy. 2018;(February):1136-1144. doi:10.1002/jrs.5367

## CHAPTER 2

2



# A field-applicable method for flow cytometric analysis of granulocyte activation: Cryopreservation of fixed granulocytes

Karin de Ruiter,<sup>1</sup> Selma van Staveren,<sup>2,3</sup> Bart Hilvering,<sup>2</sup> Edward Knol,<sup>4</sup> Nienke Vrisekoop,<sup>2</sup> Leo Koenderman,<sup>2</sup> Maria Yazdanbakhsh<sup>1</sup>

1 Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands

2 Department of Respiratory Medicine, Laboratory of Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

3 TI-COAST, Science Park, Amsterdam, The Netherlands

4 Department of Dermatology and Allergology, Laboratory of Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

Upon activation granulocytes upregulate several adhesion molecules (CD11b) and granule proteins (CD35, CD66b) and shed surface L-selectin (CD62L). These changes in expression, as assessed by flow cytometry, can be used as markers for activation. Whereas these markers are usually studied in fresh blood samples, a new method is required when samples are collected at a field site with no direct access to a flow cytometer. Therefore, we developed and tested a field-applicable method in which fixed leukocytes were cryopreserved. Using this method, the intensity of granulocyte activation markers was compared to samples that were either stained fresh, or fixed prior to staining but not cryopreserved. In addition, the response to an in vitro stimulation with fMLF was determined. While we observed differences in marker intensities when comparing fresh and fixed granulocytes, similar intensities were found between fixed cells that had been cryopreserved and fixed cells that did not undergo cryopreservation. Although fixation using FACS lysing solution might lead to membrane permeabilization, activation markers, and the responsiveness to fMLF or eotaxin could still be clearly measured. This method will, therefore, enable future studies of granulocyte activation in settings with limited resources and will allow simultaneous analysis of samples collected at different time points.

## Introduction

Granulocytes form a class of innate effector cells and consist of eosinophils, neutrophils, and basophils, all characterized by the presence of lobulated nuclei and secretory granules in their cytoplasm. Upon inflammation, granulocytes are recruited from the circulation to the site of tissue injury or infection, where they become activated and exert their effector functions. To adhere to the endothelium, granulocytes express L-selectin (CD62L), and several selectin ligands which mediate “rolling” along the vessel endothelium.<sup>1,2</sup> This is followed by firm adhesion to and transmigration through endothelial cells, which is mainly mediated by the integrins Mac1 (CD11b/CD18) and VLA4 (CD49d/CD29, only expressed by eosinophils and basophils).<sup>1,2</sup>

In addition to adhesion molecules, surface expression of proteins normally residing inside granule membranes can be used as markers for degranulation and, therefore, activation.<sup>3</sup> Relevant are complement receptor 1 (CD35), mediating binding and phagocytosis of C3b coated particles and immune complexes, present in secretory vesicles, and CEACAM-8 (CD66b) which is present in specific granules.<sup>3</sup>

Whereas neutrophils demonstrate a rapid upregulation of CD11b, CD35 and CD66b after in vitro activation with inflammatory agonists such as *N*-Formyl-Met-Leu-Phe (fMLF), the level of CD62L reduces as a result of stimulation-induced shedding.<sup>4,5</sup> Similarly, eosinophils respond to the chemokine eotaxin (amongst others) by upregulating CD11b and shedding their CD62L.<sup>6,7</sup> Therefore, quantification of these markers by flow cytometry can be used as read-out for the activation state and responsiveness of the cells for inflammatory mediators.<sup>4,5,8–14</sup>

To study the expression of activation markers on granulocytes by flow cytometry, the use of lysed whole blood is preferred over cell isolation techniques as the latter could lead to nonspecific activation.<sup>15</sup> In addition, purification procedures often require more time and a larger volume of blood compared to using whole blood. Therefore, total leukocytes are usually analysed by multicolour flow cytometry after staining whole blood with fluorescently-labelled antibodies followed by multiple gating.<sup>4,5,16</sup> Whereas this method is suitable for laboratories with direct access to flow cytometers, circumstances are different when blood is collected at a field site with only basic laboratory infrastructure. Hence, a novel method is required in which samples can be collected in a setting with limited resources and stored for long time periods until analysis in a fully equipped laboratory.

Fixation methods are often used to preserve cells and studies have successfully demonstrated the applicability of granulocyte fixation prior to staining.<sup>8,10,17</sup> However, fixation can modify cell morphology as measured by forward scatter (FSC) and side

scatter (SSC) of leukocytes.<sup>8,17–19</sup> Fixation may also affect granulocyte cell membrane permeability<sup>18–20</sup> and can affect antigen epitopes and, thereby, their interaction with the monoclonal antibodies.<sup>9</sup> This requires careful evaluation of each antigen-antibody combination.

As the cryopreservation of freshly isolated granulocytes has been proven difficult because of high susceptibility to damage during the cryopreservation method,<sup>21</sup> cryopreservation of fixed lysed whole blood samples might be a suitable method for long-term storage. This would allow sample collection at the field site requiring only pipettes, a centrifuge and a -80°C freezer, followed by transport to a central laboratory for flow cytometric analysis. Moreover, long-term storage allows researchers to analyse samples collected at different time points simultaneously.

Nemes et al. previously provided proof-of-principle that cryopreservation leads to accurate quantification of cell subsets (granulocytes, lymphocytes, monocytes, T cells and B cells) by flow cytometry in fixed whole blood samples.<sup>17</sup> Our study tested the hypothesis that also granulocyte activation can be determined in such samples by comparing the expression of granulocyte activation markers on samples that were either fresh (no fixation), fixed prior to staining, or fixed and cryopreserved prior to staining. In addition, we set out experiments to measure the response of granulocytes to an in vitro stimulation with fMLF (neutrophils or basophils) or eotaxin (CCL11; eosinophils) in cells fixed and cryopreserved after activation.

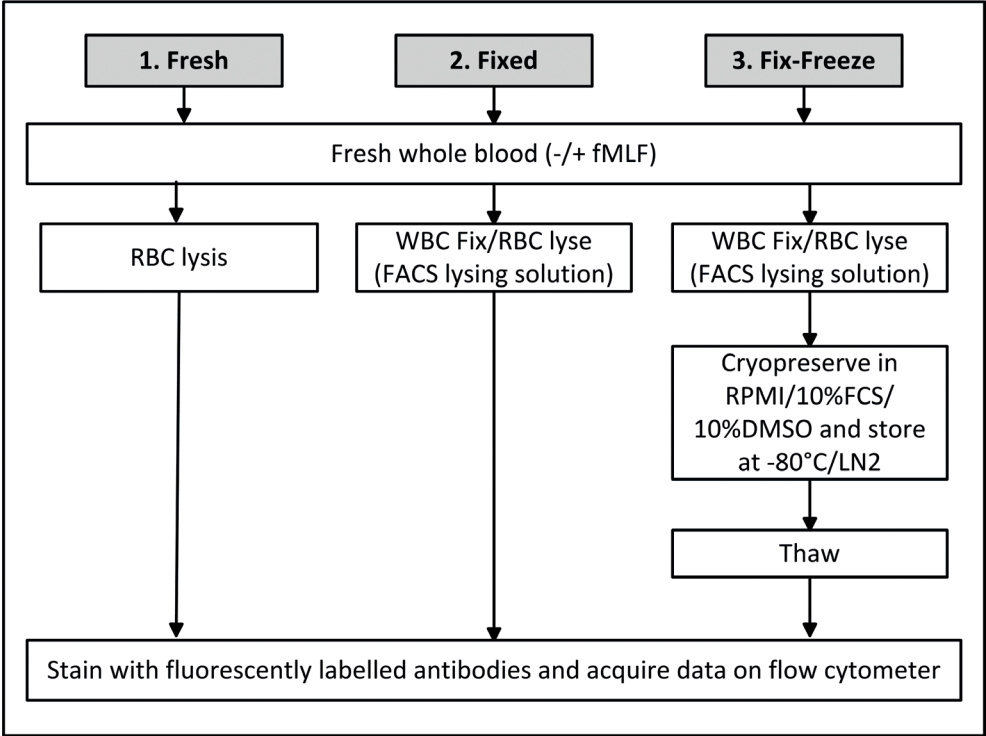
The applicability of cryopreservation of fixed cells was subsequently tested in a controlled laboratory model of acute inflammation<sup>22</sup> and in a field study in Flores, Indonesia.<sup>23</sup> We now show that cryopreservation of fixed total leukocyte populations allows the determination of activated granulocytes, either activated by chemokines in vitro or by inflammation in vivo.

## Materials & Methods

### Sample preparation

Venous blood of three healthy volunteers was collected in sodium heparin vacutainers (BD Biosciences, Franklin Lakes, NJ, USA) and handled within 30 minutes. Different methods were compared, depicted in Figure 1, and for each method 2 polystyrene round-bottom tubes containing 200  $\mu$ L of blood per donor were pre-incubated for 5 min in a 37°C waterbath, followed by a 5 min-stimulation at 37°C with fMLF ( $10^{-5}$  M; Sigma, Saint Louis, MO, USA) or left unstimulated. Subsequently, samples were subjected to the following treatments: (1) Erythrocyte lysis using lysing buffer consisting of 155 mM  $\text{NH}_4\text{Cl}$  and 10 mM  $\text{Na}_2\text{EDTA}$ , with a pH of 7.0 and osmolality of 298-305. Cells were transferred to ice and to each sample 900  $\mu$ L lysing buffer was added. Cells were centrifuged at 1600 rpm for 5 min and washed with first 500  $\mu$ L lysing buffer, followed by washing and resuspension in FACS buffer (PBS complemented with 0.5% BSA and 2 mM EDTA) and directly stained. (2) Erythrocyte lysis and leukocyte fixation using FACS lysing solution (BD Biosciences, #349202) for 15 min at room temperature. Following the recommended procedures, the FACS lysing solution (stock solution 10x) was diluted, and 4 mL of the diluted reagent was used for each whole blood sample. After lysis the cells were centrifuged, washed with RPMI 1640 containing 10% heat-inactivated foetal calf serum (FCS), centrifuged again and finally resuspended in FACS buffer and stored (<8 hrs) at 4°C until staining. (3) Erythrocyte lysis and leukocyte fixation using FACS lysing solution as described above (Treatment 2). Following lysis, cells were washed with 10% FCS/RPMI and resuspended in 10% FCS/10% dimethyl sulfoxide (DMSO)/RPMI. Cryovials containing the cell suspension were placed at -80°C. The cryopreserved fixed cells were thawed on the bench and subsequently transferred to 5 ml polystyrene round-bottom tubes containing 1 mL 10% FCS/RPMI. After centrifugation at 1600 RPM for 5 min, cells were resuspended in FACS buffer and stored (<2 hrs) at 4°C until staining. In the results section, Treatments 1, 2 and 3 will be referred to as "Fresh", "Fixed" and "Fix & Freeze", respectively.

Cryopreservation of fixed leukocytes (Treatment 3), was applied in a clinical trial conducted in a rural area at Flores island in Indonesia to study the effect of soil-transmitted helminth infections on granulocyte activation. This SUGARSPIN trial was approved by the ethics committee of Faculty of Medicine, Universitas Indonesia (FKUI), filed by the ethics committee of Leiden University Medical Center (LUMC), and registered as a clinical trial (<http://www.isrctn.com/ISRCTN75636394>). The study protocol was published previously.<sup>23</sup> Blood samples were collected after obtaining written consent and processed according to Treatment 3, with the only difference that cells were stored in liquid nitrogen after placement at -80°C for a minimum of 4 hours. In addition to the stimulation with fMLF described above, whole blood was stimulated for 5 min at 37°C with eotaxin (CCL11;  $10^{-7}$  M, R&D systems, Abingdon, UK) to study in vitro eosinophil activation.



**Figure 1. Experiment overview.**

### Neutrophil subset identification after LPS challenge

To assess whether neutrophil activation can also be detected after using the method of fixation and cryopreservation following in vivo stimulation, cells from a large endotoxin trial (NCT02629874 at [www.clinicaltrials.gov](http://www.clinicaltrials.gov)) were used. In short, healthy male volunteers were enrolled after screening and were prehydrated. U.S. Reference *E. coli* endotoxin (Lot Ec-5; Centre for Biologic Evaluation and Research, U.S. Food and Drug Administration, Bethesda, MD, USA) was used in this study. Endotoxin was reconstituted in 5 mL saline and injected as a single intravenous bolus (2 ng/kg) during 1 min at  $t = 0$ . The study protocol was approved by the Ethics Committee of the Radboud University Nijmegen Medical Centre and complies with the Declaration of Helsinki and Good Clinical Practice guidelines. Subjects gave written informed consent.

At  $t = 180$  min, blood samples, anticoagulated with sodium heparin, were taken from the arterial catheter. The blood samples were processed according to Treatments 1 and 3 described above. For Treatment 3, cells were first stored at  $-30^{\circ}\text{C}$  for 8 days for operational reasons (no  $-80^{\circ}\text{C}$  freezer at sampling site), before they were transferred to  $-80^{\circ}\text{C}$ .

### Flow cytometric analysis

Irrespective of the method used, cells were counted by microscopy after erythrocyte lysis and 500,000 leukocytes were transferred to a 96-well V-bottom microplate and incubated for 30 min at 4°C with anti-CD35-FITC (E11, Biolegend, San Diego, USA), anti-CD66b-PerCP/Cy5.5 (G10F5, Biolegend), anti-CD193-PE (5e, Biolegend), anti-CD16-PE/CF594 (3G8, BD Biosciences), anti-CD69-PE/Cy5 (FN50, Biolegend), anti-CD3/CD19/CD20/CD56-APC (UCHT1, HIB19, 2H7, 5.1H11, Biolegend), anti-CD11b-APC/eF780 (ICRF44, eBioscience Inc., San Diego, USA), anti-CD203c-BV421 (NP4D6, Biolegend), anti-CD14-BV510 (M5E2, Biolegend) and anti-CD62L-BV605 (DREG-56, BD Biosciences). Cells collected after LPS challenge were stained with anti-CD16-Pacific Orange (3G8, Beckman Coulter) and CD62L-BV650 (DREG-56, Biolegend). Compensation beads (BD Biosciences) were stained with each antibody separately and run before acquisition to calculate the compensation matrix using BD FACS Diva flow cytometry software.

As we aimed to compare three methods, all samples were measured on the same day and stained with the same antibody mix, with exception of the samples that were collected in the endotoxin clinical trial or as part of the SUGARSPIN study. Cells were measured on a LSR Fortessa flow cytometer (BD Biosciences), equipped with 405, 488, 561 and 640 lasers. Data were analysed with FlowJo software (Treestar Inc., Ashland, OR, USA) and median fluorescence intensity (MFI) data are displayed. Representative gating schemes are shown in Supplementary Figure S1. Although neutrophils and eosinophils clearly differ in their FSC/SSC properties, we used the expression of CD193 (in fresh samples) or the autofluorescent signal (in fixed samples) to gate eosinophils. Neutrophils were gated based on their FSC/SSC properties and subsequently separated from monocytes based on their CD14<sup>+</sup>/CD11b<sup>+</sup> expression. Basophils were gated based on their expression of CD193 and CD203c and lack of the lineage markers CD3, CD19, CD20 and CD56.

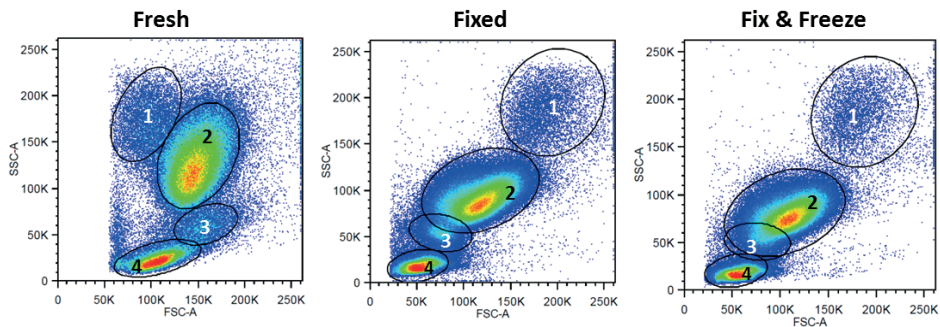
### Statistical analysis

Data analysis was performed using GraphPad Prism version 7 for Windows (GraphPad Software, San Diego, CA, USA) and this software was also used to make graphs. Paired *t* test was used to assess the effect of fMLP stimulation on the expression of activation markers in neutrophils and basophils. The effect of stimulation on markers expressed by neutrophils, eosinophils and basophils from samples collected in a field study was tested using Wilcoxon matched-pairs signed rank test. *P* values ≤ 0.05 were considered statistically significant.

## Results

### Fixation leads to changes in FSC/SSC profiles

After fixation, granulocytes displayed different forward scatter/ side scatter (FSC/SSC) profiles compared to fresh cells suggesting a change in morphology (Figure 2). Eosinophils showed a shift in FSC and became more autofluorescent after fixation (Supplementary Figure S2). No increase in autofluorescence was observed for neutrophils (data not shown). After fixation, neutrophils exhibited a lower SSC signal, resulting in an overlap between neutrophils and monocytes. The amount of overlap increased after freeze-thawing, and it was, therefore, important to add a monocyte marker (e.g. CD14) and/or a specific neutrophil marker (e.g. CD66b) to the panel to be able to discriminate between these cell types. Except for neutrophils, fixed leukocytes that had been cryopreserved displayed similar FSC-SSC properties compared to fixed cells that did not undergo cryopreservation.



**Figure 2. FSC/SSC properties of granulocytes.**

Numbers indicate different cell populations: (1) eosinophils, (2) neutrophils, (3) monocytes and (4) lymphocytes. After fixation, eosinophils change in FSC/SSC and monocytes and neutrophils show overlap in FSC/SSC properties. This overlap is further amplified after freeze-thawing the cells. Basophils overlap with lymphocytes in FSC/SSC properties.

### Neutrophils

#### ***Cellular activation and responsiveness to fMLF can be measured by activation markers after fixation and cryopreservation***

We observed changes in marker intensities when comparing fresh and fixed unstimulated neutrophils (Figure 3, Table 1). The measured intensities of CD11b, CD35 and CD66b were elevated after fixation when compared to fresh cells, whereas the detected intensity of CD62L on neutrophils was lower after fixation compared to fresh cells. These changes were consistent in all three donors. No major differences in the marker intensities were observed when fixed neutrophils that had been cryopreserved were compared to fixed cells that did not undergo cryopreservation (Figure 3), indicating that the observed



differences with fresh cells are a consequence of fixation, and are not caused by the process of freeze-thawing.

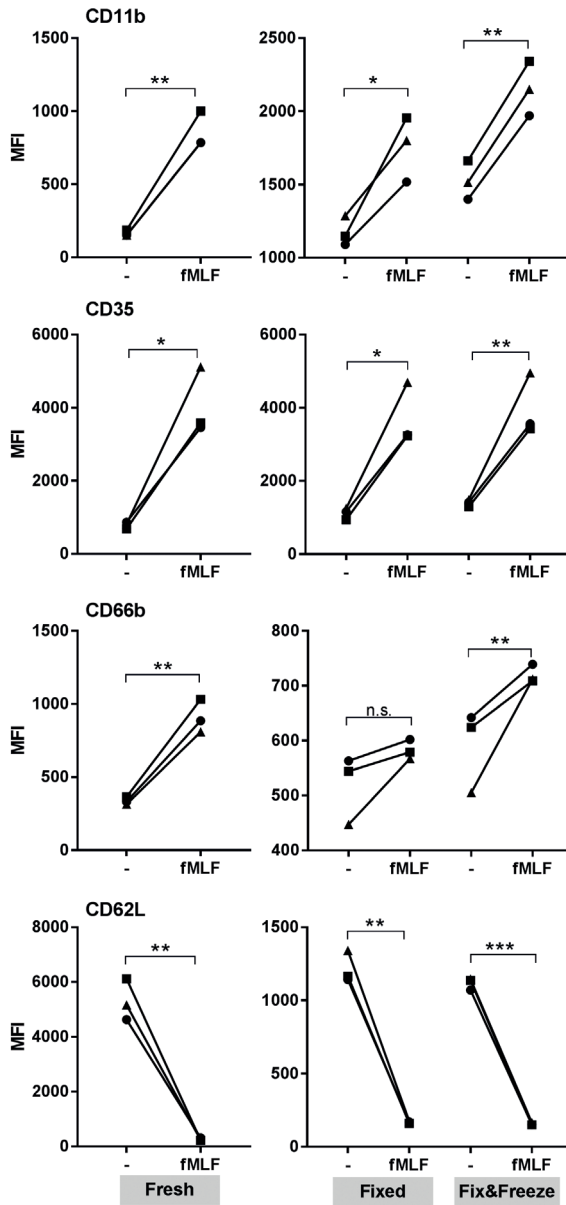
To assess whether it is possible to detect activation of neutrophils after fixation and cryopreservation, we measured the levels of activation markers in samples that were incubated with and without fMLF. Despite the changes in marker intensities that were described above, fMLF activation could still further increase the expression of measured activation markers in fixed and cryopreserved neutrophils (Figure 3, Table 1). The detected responsiveness of CD35 and CD11b to fMLF was similar in fresh and fixed neutrophils, irrespective of freezing. Regarding CD66b, we observed an increased expression after stimulation with fMLF in fresh neutrophils as well as fixed cells. However, in fixed cells, the increase in CD66b expression was rather weak compared to fresh cells, indicating that CD66b might not be suitable for studying functional capacity of neutrophils in terms of responding to a stimulus when this field applicable method is used. CD62L is shed from the membrane upon activation, and despite the overall lower intensity of CD62L in fixed neutrophils, we could still observe this response after activation with fMLF.

**Table 1. Expression of CD11b, CD35, CD66b, and CD62L on neutrophils and the response to fMLF**

Neutrophils	Fresh Mean MFI (SD)		Fixed Mean MFI (SD)		Fix & Freeze Mean MFI (SD)	
	Medium	fMLF	Medium	fMLF	Medium	fMLF
CD11b	164 (20)	857 (124)	1175 (101)	1758 (222)	1525 (132)	2154 (186)
CD35	788 (92)	4054 (921)	1113 (160)	3729 (832)	1399 (99)	3981 (845)
CD66b	337 (25)	909 (113)	518 (62)	583 (18)	590 (74)	720 (17)
CD62L	5304 (753)	261 (49)	1216 (108)	169 (8)	1118 (41)	156 (8)

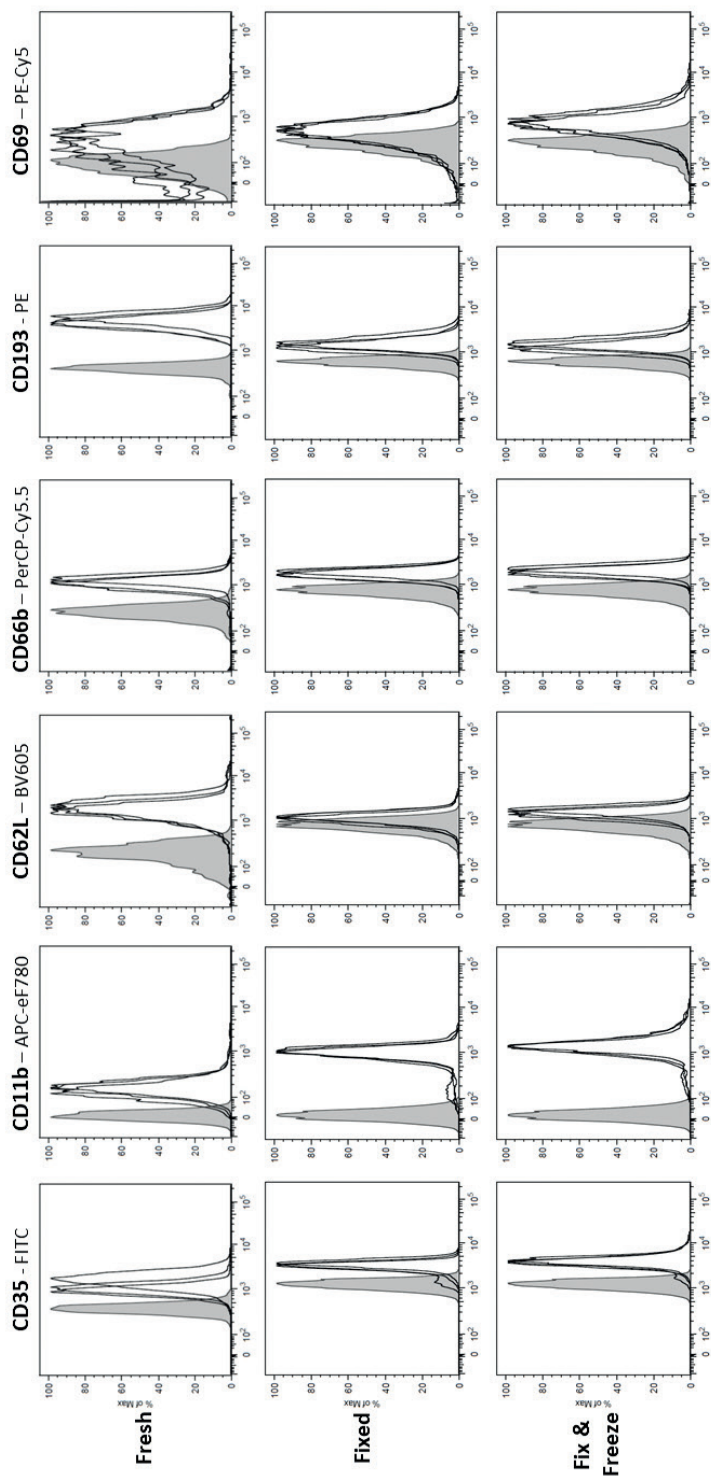
**Three neutrophil subsets can be identified based on CD16 and CD62L expression after fixation and cryopreservation**

The experiments described above relied on the in vitro activation of leukocytes by fMLF. We next studied leukocyte subsets that have been described in fresh cells following in vivo stimulation by a systemic pro-inflammatory stimulus, LPS. Thus, since CD62L intensities were found to be decreased after fixation, we assessed whether we would still be able to identify the three different neutrophil subsets that were previously described on fresh cells.<sup>22</sup> Although both CD16 and CD62L intensities were lower after fixation and cryopreservation, we were able to identify the three different neutrophil subsets (1- CD16<sup>bright</sup>CD62L<sup>bright</sup>; 2- CD16<sup>dim</sup>CD62L<sup>bright</sup>; 3- CD16<sup>bright</sup>CD62L<sup>dim</sup>) based on expression of these two markers (Supplementary Figure S3).



**Figure 3. Expression of CD11b, CD35, CD66b and CD62L on neutrophils and the response to fMLF.**

Median fluorescent intensity (MFI) of markers on neutrophils left unstimulated and stimulated with fMLF. Neutrophils were stained either fresh (Fresh), after fixation (Fixed) or after being fixed, cryopreserved and thawed (Fix & Freeze). Symbols (circle, triangle, square) represent three different donors. \*indicates analysis with paired t test. \*,  $P \leq 0.05$ . \*\*,  $P \leq 0.01$ . \*\*\*,  $P \leq 0.001$ . n.s., not significant.



**Figure 4. Expression of CD35, CD11b, CD62L, CD66b, CD193 and CD69 on eosinophils.** Grey histograms represent unstained eosinophils while the black lines represent three different donors. Cells were stained either fresh (Fresh), after fixation (Fixed) or after being fixed, cryopreserved and thawed (Fix & Freeze).

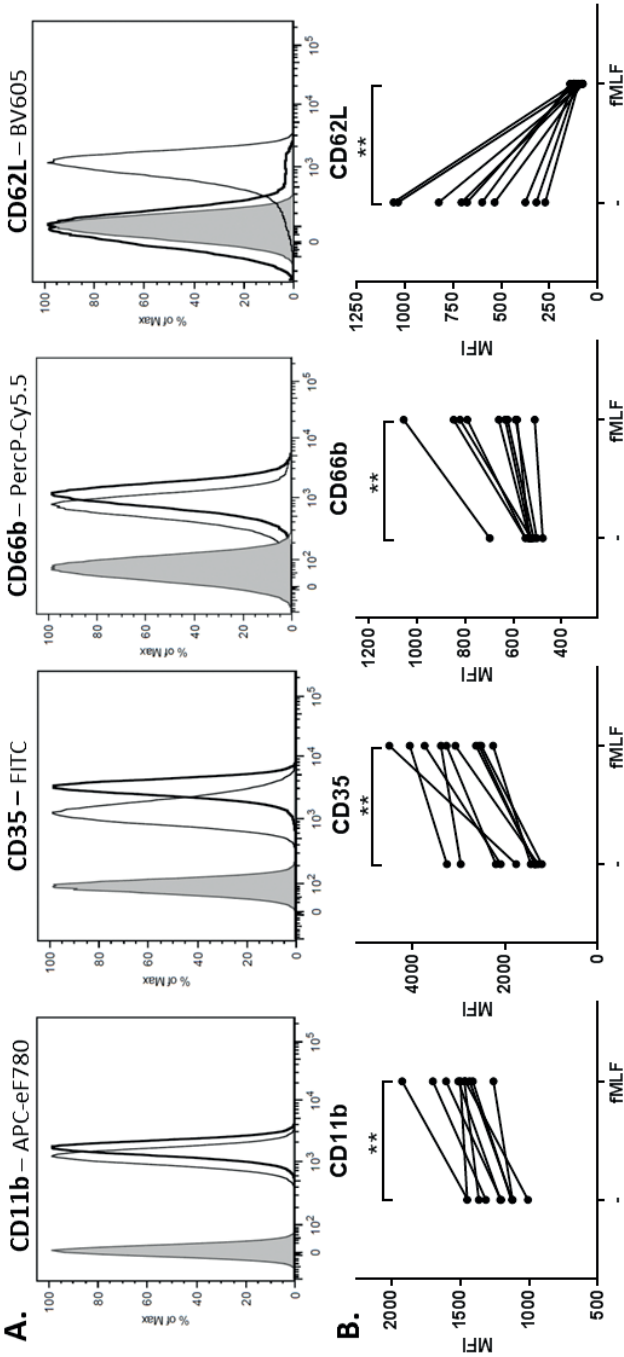
### Eosinophils and Basophils

The intensities of all eosinophil markers studied (CD193, CD66b, CD35, CD11b, CD62L and CD69) were similar between fixed eosinophils that had been cryopreserved and fixed cells that did not undergo cryopreservation (Figure 4). However, similar to the situation with neutrophils, differences were observed in the intensities of activation markers for eosinophils when comparing fresh and fixed cells. The measured intensities of CD35 and CD66b were slightly higher after fixation, while the intensity of CD11b was strongly increased and CD62L detection was lower in fixed compared to fresh cells.

Basophils also respond to fMLF<sup>24</sup> and the intensity of CD11b, CD35 and CD203c was increased after stimulation (Supplementary Figure S4, Table S1). Like the expression of these markers in unstimulated cells, the responsiveness to fMLF was similar between fresh, fixed and cryopreserved cells.

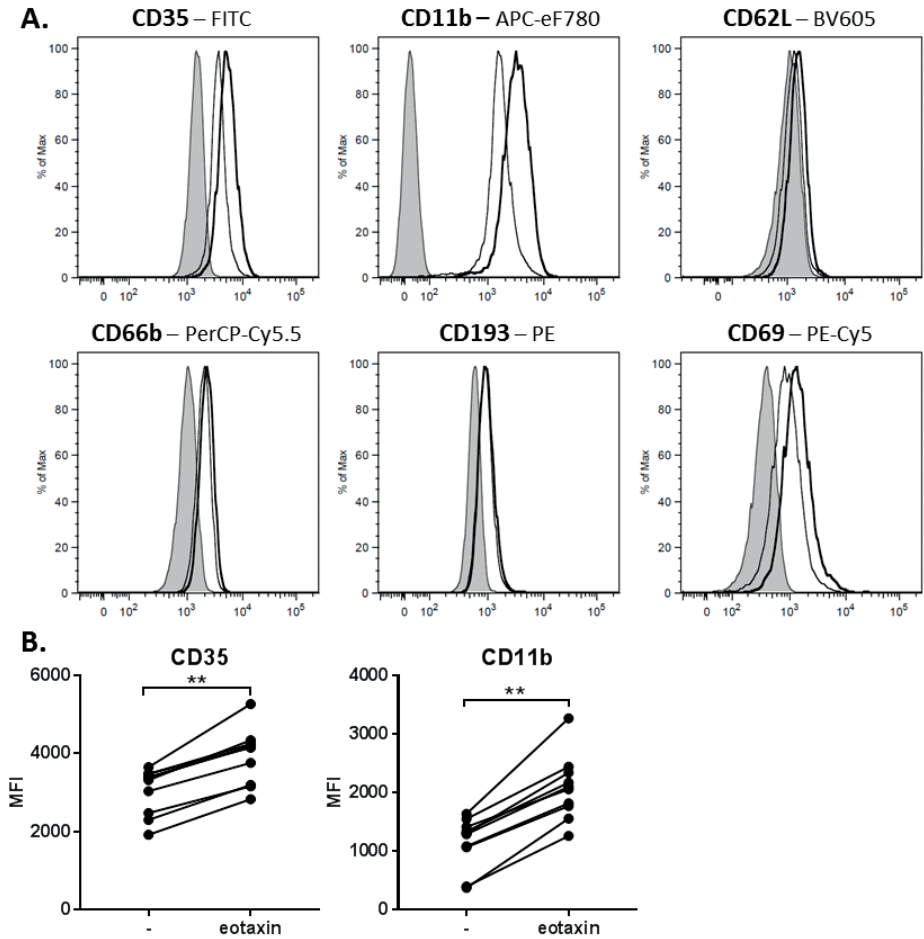
### Field study application

The aim of this study was to set up a method to assess granulocyte activation and responsiveness in a field study with limited laboratory equipment. After testing the method to fix and cryopreserve leukocytes for flow cytometric analysis, this method was applied in a clinical trial in a rural area in Indonesia.<sup>23</sup> Blood samples from 300 subjects were processed according to the method described and stored in liquid nitrogen for up to two years until analysis. Figures 5 and 6 and Supplementary Figure S5 show representative samples collected during this study. Even after long-term storage we could measure the expression of activation markers on neutrophils (Figure 5A), eosinophils (Figure 6A) and basophils (Supplementary Figure S5A). Moreover, the response of neutrophils and basophils to fMLF evoked in samples in the field was clearly present (Figure 5B, Supplementary Figure S5B, Table S2). To study the responsiveness of eosinophils, we stimulated whole blood with eotaxin and observed an increased intensity of CD35 and CD11b when compared to unstimulated cells (Figure 6B, Table S2). Although not compared to fresh cells, this shows that the activation effect in eosinophils can be measured after fixation and cryopreservation.



**Figure 5. Expression of CD11b, CD35, CD66b and CD62L on neutrophils collected in a field study.**

Samples were cryopreserved for 2 years in liquid nitrogen. (A) Histograms of a representative sample, filled lines display unstained neutrophils, the thin black lines unstimulated cells and the thick black lines fMLF-stimulated cells. (B) Median fluorescent intensity (MFI) of markers in neutrophils from 10 representative donors left unstimulated and stimulated with fMLF. Means and standard deviations are shown in Supplementary Table S2. \*indicates analysis with Wilcoxon matched-pairs signed rank test. \*\*,  $P \leq 0.01$ .



**Figure 6. Expression of CD35, CD11b, CD62L, CD66b, CD193 and CD69 on eosinophils collected in a field study.**

Samples were cryopreserved for 2 years in liquid nitrogen. A) Histograms of a representative sample, filled lines represent unstained eosinophils, the thin black lines unstimulated cells and the thick black lines eotaxin-stimulated cells. B) Median fluorescent intensity (MFI) of CD35 and CD11b in eosinophils from 10 donors left unstimulated and stimulated with eotaxin. Means and standard deviations are shown in Supplementary Table S2. \*indicates analysis with Wilcoxon matched-pairs signed rank test. \*\*,  $P \leq 0.01$ .

## Discussion

Here, we show that the activation status of neutrophils, eosinophils, and basophils and their responsiveness to stimuli such as fMLF and eotaxin, can be studied in whole-blood samples that were lysed, fixed and cryopreserved. Previously, Nemes et al. showed that cryopreservation of fixed leukocytes provides for an accurate quantification of cell subsets based on the expression of lineage-specific markers (CD66b, CD14, CD3, CD19, CD16), as analysed by flow cytometry.<sup>17</sup> However, they had not tested whether the method was also suitable for assessing the activation status of granulocytes. We show that cryopreservation of granulocytes allows detection of activation of granulocytes *in vitro* in response to fMLF and eotaxin, as well as activation *in vivo* during acute inflammation evoked by experimental endotoxemia in healthy volunteers. It will now be possible to measure differences in innate immune responses at field sites where no flow cytometers are available for direct analysis.

Cryopreservation of fixed granulocytes induces some differences in cellular characteristics that preclude direct comparison with analysis of fresh cells. For instance, the differences detected in FSC/SSC profiles of granulocytes and marker intensities between cryopreserved and fresh cells were mainly caused by fixation. The increased intensity of CD11b, CD35 and CD66b in fixed neutrophils might be a consequence of fixation-induced membrane permeabilization, potentially leading to intracellular staining in the granules.<sup>19,20</sup> This is particularly relevant for markers normally residing in the granular compartment, which become expressed upon degranulation. Only 5% of the total cellular content of the CD11b-CD18 (Mac-1) complexes in unstimulated neutrophils is expressed on the plasma membrane, whereas 95% colocalized with specific granules and secretory vesicles.<sup>25</sup> Therefore, the increased intensity of CD11b in fixed versus fresh neutrophils might reflect some staining of the intracellular pool of CD11b molecules. Importantly, despite the putative membrane permeability caused by fixation, we could still observe a significant effect of fMLF in neutrophils (in particular for CD35, CD11b and CD62L) and basophils (in particular for CD203c, CD35 and CD11b), and of eotaxin in eosinophils (in particular for CD35 and CD11b). This activation effect was also clearly present in samples collected in the field which had been cryopreserved for up to two years.

Hamblin et al. observed no intracellular leukocyte staining of p8,14 (expressed in all circulating monocytes and neutrophils) nor DNA staining after fixation with 0.4% formaldehyde.<sup>8</sup> This suggests that different fixation methods might have different effects in terms of cell permeabilization. However, the use of FACS lysing solution is a quick and easy method and, therefore, suitable to be used in the field.

We also observed that eosinophils became more autofluorescent after fixation and it is, therefore, difficult to determine in eosinophils whether the differences in marker intensity between fresh and fixed cells are a result of cell membrane permeabilization, affected antigen epitopes, or increased autofluorescence. The effect of autofluorescence on the intensity of a certain marker depended on the fluorescent channel being used, as we observed no shift in the channel used to detect APC/eF780. CD193 (chemokine receptor CCR3 / eotaxin receptor) is a well-known marker for eosinophils and we observed a lower detection of CD193 in fixed compared to fresh eosinophils. Combined with an increased autofluorescent signal in the PE-channel, we recommend not to use CD193 as single eosinophil marker when analysing fixed, cryopreserved granulocytes. CD193 functioned well as a basophil marker, irrespective of fixation.<sup>26,27</sup> For similar reasons, we advise to use a CD62L antibody with a different fluorophore than BV605, or to increase the amount of antibody when staining fixed eosinophils to be able to detect CD62L.

In summary, these results demonstrate that fixation and cryopreservation of lysed whole blood samples can be applied at field sites with limited laboratory infrastructure. We showed that the activation status of granulocytes and responsiveness to fMLF or eotaxin could be measured after cryopreservation, allowing to measure changes in the innate immune response in patients. This will enable future studies of granulocyte activation in settings with limited resources and allows the parallel analysis of samples collected at different time points and at different sites.



## References

1. Rosenberg HF, Dyer KD, Foster PS. Eosinophils: Changing perspectives in health and disease. *Nat Rev Immunol*. 2013. doi:10.1038/nri3341
2. Kolaczowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol*. 2013;13(3):159-175. doi:10.1038/nri3399
3. Cowland JB, Borregaard N. Granulopoiesis and granules of human neutrophils. *Immunol Rev*. 2016. doi:10.1111/imr.12440
4. Mann BS, Chung KF. Blood neutrophil activation markers in severe asthma: Lack of inhibition by prednisolone therapy. *Respir Res*. 2006. doi:10.1186/1465-9921-7-59
5. Berends C, Hoekstra MO, Dijkhuizen B, De Monchy JGR, Gerritsen J, Kauffman HF. Expression of CD35 (CR1) and CD11b (CR3) on circulating neutrophils and eosinophils from allergic asthmatic children. *Clin Exp Allergy*. 1993. doi:10.1111/j.1365-2222.1993.tb00277.x
6. Santamaria LF, Palacios JM, Beleta J. Inhibition of eotaxin-mediated human eosinophil activation and migration by the selective cyclic nucleotide phosphodiesterase type 4 inhibitor rolipram. *Br J Pharmacol*. 1997. doi:10.1038/sj.bjp.0701243
7. Tenscher K, Metzner B, Schöpf E, Norgauer J, Czech W. Recombinant human eotaxin induces oxygen radical production, Ca(2+)-mobilization, actin reorganization, and CD11b upregulation in human eosinophils via a pertussis toxin-sensitive heterotrimeric guanine nucleotide-binding protein. *Blood*. 1996.
8. Hamblin A, Taylor M, Bernhagen J, et al. A method of preparing blood leucocytes for flow cytometry which prevents upregulation of leucocyte integrins. *J Immunol Methods*. 1992. doi:10.1016/0022-1759(92)90231-H
9. Finn A, Rebuck N. Measurement of adhesion molecule expression on neutrophils and fixation. *J Immunol Methods*. 1994. doi:10.1016/0022-1759(94)90048-5
10. Torsteinsdóttir I, Arvidson NG, Hällgren R, Håkansson L. Enhanced expression of integrins and CD66b on peripheral blood neutrophils and eosinophils in patients with rheumatoid arthritis, and the effect of glucocorticoids. *Scand J Immunol*. 1999. doi:10.1046/j.1365-3083.1999.00602.x
11. Mawhorter SD, Stephany DA, Ottesen EA, Nutman TB. Identification of surface molecules associated with physiologic activation of eosinophils. Application of whole-blood flow cytometry to eosinophils. *J Immunol*. 1996.
12. Chiba T, Kamada Y, Saito N, et al. RANTES and eotaxin enhance CD11b and CD18 expression on eosinophils from allergic patients with eosinophilia in the application of whole blood flow cytometry analysis. In: *International Archives of Allergy and Immunology*. ; 2005. doi:10.1159/000085426
13. Yoon J, Terada A, Kita H. CD66b Regulates Adhesion and Activation of Human Eosinophils. *J Immunol*. 2007;179(12):8454-8462. doi:10.4049/jimmunol.179.12.8454
14. Zhao L, Xu S, Fjaertoft G, Pauksen K, Håkansson L, Venge P. An enzyme-linked immunosorbent assay for human carcinoembryonic antigen-related cell adhesion molecule 8, a biological marker of granulocyte activities in vivo. *J Immunol Methods*. 2004. doi:10.1016/j.jim.2004.08.009

15. Kuijpers TW, Tool AT, van der Schoot CE, et al. Membrane surface antigen expression on neutrophils: a reappraisal of the use of surface markers for neutrophil activation. *Blood*. 1991.
16. Fujiwara RT, Cançado GGL, Freitas PA, et al. Necator americanus infection: A possible cause of altered dendritic cell differentiation and eosinophil profile in chronically infected individuals. *PLoS Negl Trop Dis*. 2009. doi:10.1371/journal.pntd.0000399
17. Nemes E, Kagina BMN, Smit E, et al. Differential leukocyte counting and immunophenotyping in cryopreserved ex vivo whole blood. *Cytom Part A*. 2015. doi:10.1002/cyto.a.22610
18. Carulli G, Sbrana S, Azzarà A, et al. Detection of eosinophils in whole blood samples by flow cytometry. *Commun Clin Cytom*. 1998. doi:10.1002/(SICI)1097-0320(19981215)34:6<272::AID-CYTO5>3.0.CO;2-L
19. Tiirikainen MI. Evaluation of red blood cell lysing solutions for the detection of intracellular antigens by flow cytometry. *Cytometry*. 1995. doi:10.1002/cyto.990200410
20. Vuorte J, Jansson SE, Repo H. Evaluation of Red Blood Cell Lysing Solutions in the Study of Neutrophil Oxidative Burst by the DCFH Assay. *Cytometry*. 2001;43(4):290-296. doi:10.1002/1097-0320(20010401)43:4<290::AID-CYTO1061>3.0.CO;2-X
21. Vian AM, Higgins AZ. Membrane permeability of the human granulocyte to water, dimethyl sulfoxide, glycerol, propylene glycol and ethylene glycol. *Cryobiology*. 2014. doi:10.1016/j.cryobiol.2013.11.004
22. Pillay J, Kamp VM, van Hoffen E, et al. A subset of neutrophils in human systemic inflammation inhibits T cell responses through Mac-1. *J Clin Invest*. 2012;122(1):327-336. doi:10.1172/JCI57990
23. Tahapary DL, de Ruiter K, Martin I, et al. Helminth infections and type 2 diabetes: A cluster-randomized placebo controlled SUGARSPIN trial in Nangapanda, Flores, Indonesia. *BMC Infect Dis*. 2015. doi:10.1186/s12879-015-0873-4
24. Chirumbolo S, Vella A, Ortolani R, et al. Differential response of human basophil activation markers: A multi-parameter flow cytometry approach. *Clin Mol Allergy*. 2008;6. doi:10.1186/1476-7961-6-12
25. Sengeløv H, Kjeldsen L, Diamond MS, Springer TA, Borregaard N. Subcellular localization and dynamics of Mac-1 (αmβ2) in human neutrophils. *J Clin Invest*. 1993. doi:10.1172/JCI116724
26. Hausmann O V., Gentinetta T, Fux M, Ducrest S, Pichler WJ, Dahinden CA. Robust expression of CCR3 as a single basophil selection marker in flow cytometry. *Allergy Eur J Allergy Clin Immunol*. 2011. doi:10.1111/j.1398-9995.2010.02431.x
27. Uguccioni M, Mackay CR, Ochensberger B, et al. High expression of the chemokine receptor CCR3 in human blood basophils. Role in activation by eotaxin, MCP-4, and other chemokines. *J Clin Invest*. 1997. doi:10.1172/JCI119624

Supplementary Material

Supplementary Figures

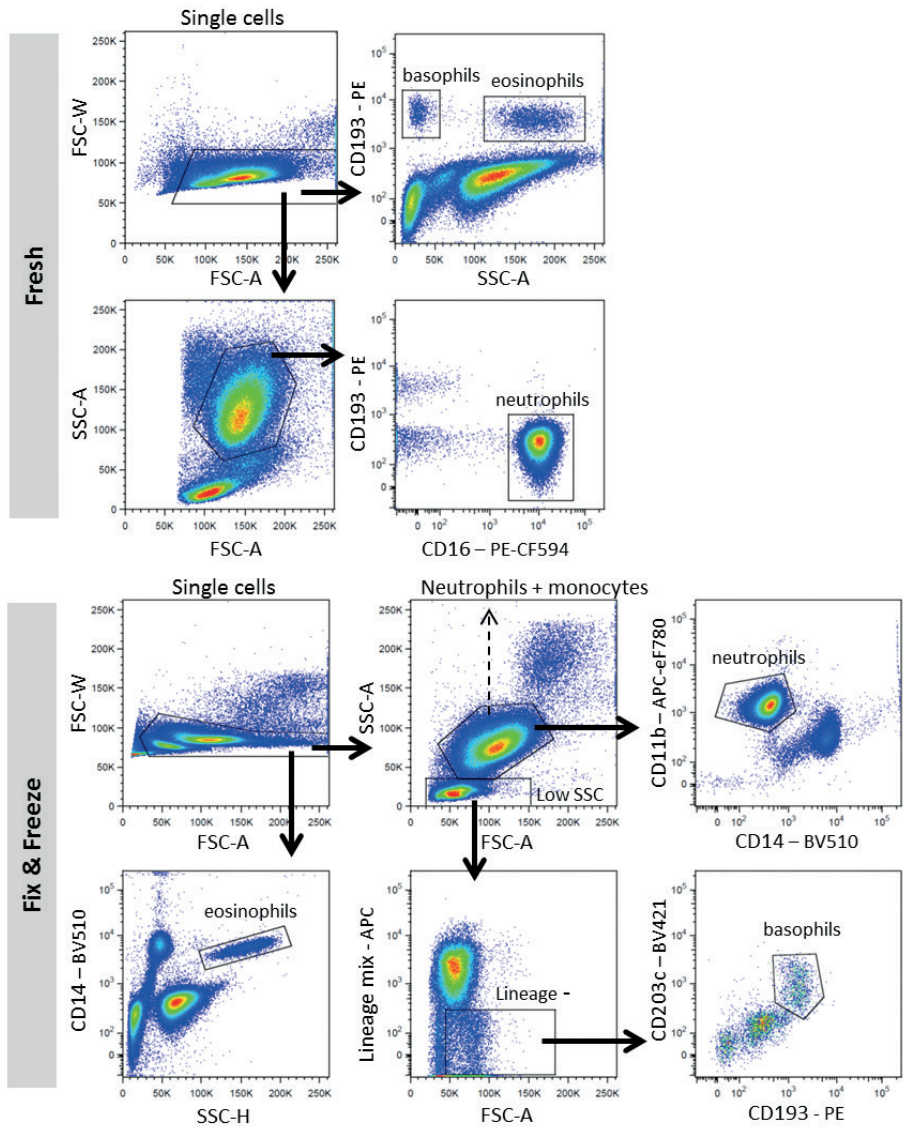
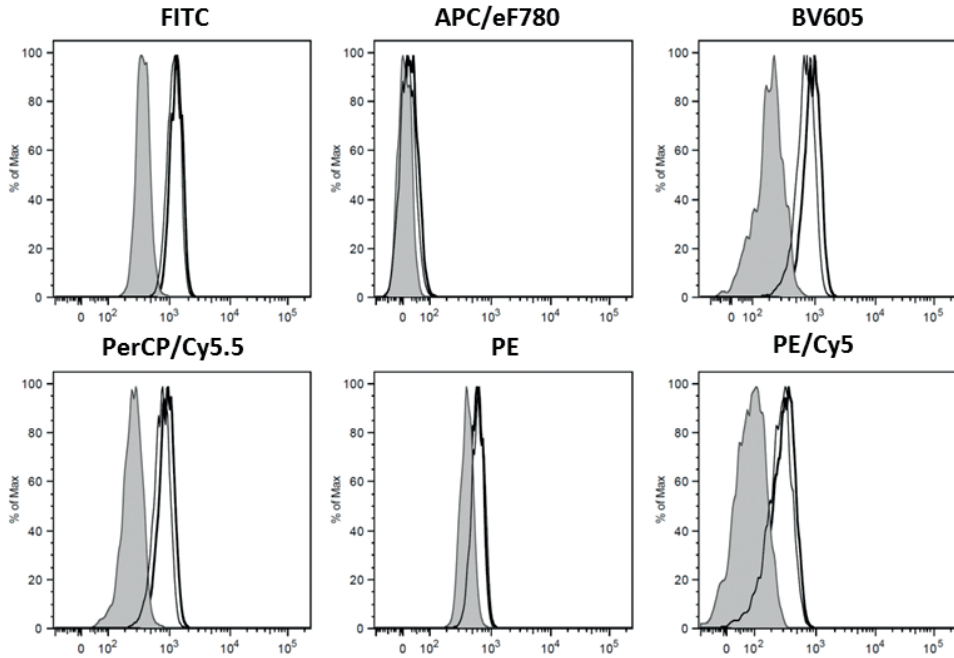


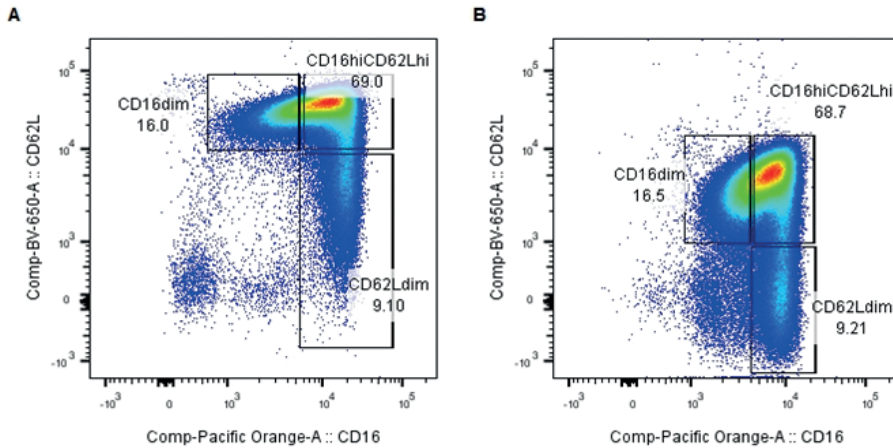
Figure S1. Gating strategies.

The gating strategy for the analysis of eosinophils, neutrophils and basophils in a representative fresh sample (upper panel) and fixed-cryopreserved sample (lower panel) is shown.



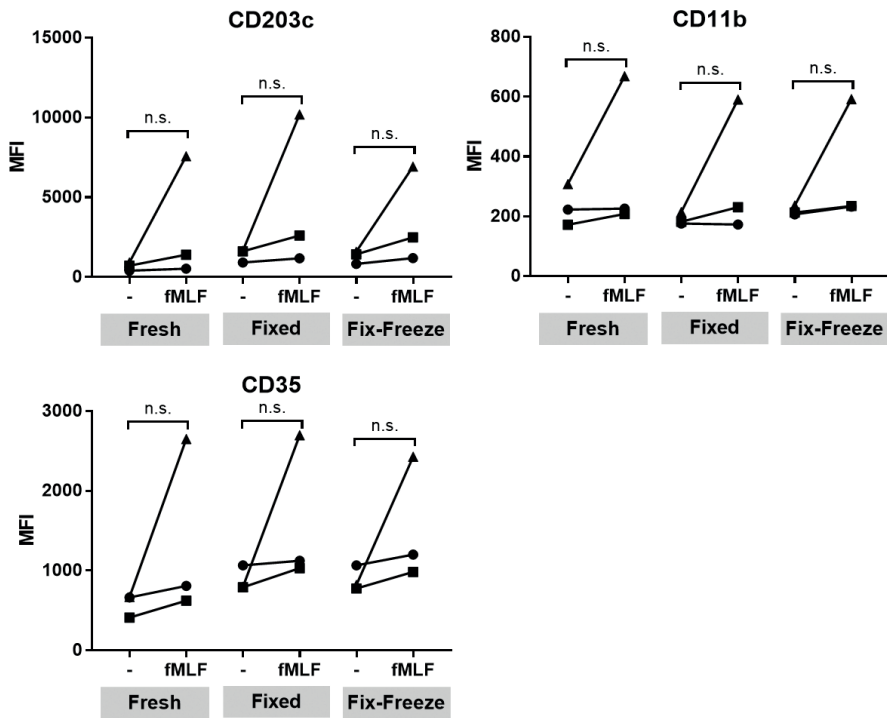
**Figure S2. Increased autofluorescence of eosinophils after fixation.**

The filled lines represent fresh, non-fixed, unstained eosinophils, while the thin black lines display fixed unstained cells that did not undergo cryopreservation and the thick black lines fixed, unstained cells after cryopreservation.



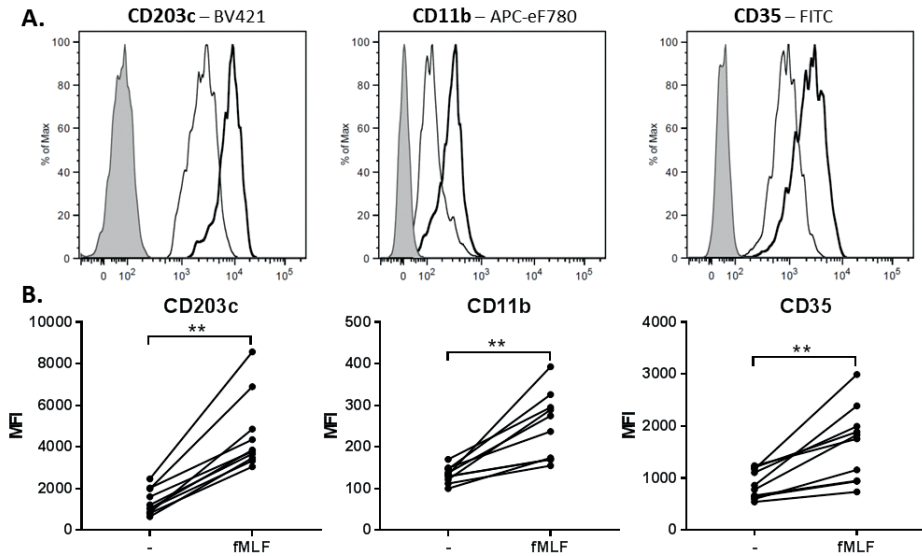
**Figure S3. Expression of CD203c, CD11b and CD35 on basophils and the response to fMLF.**

A) Median fluorescent intensity (MFI) of markers on basophils left unstimulated and stimulated with fMLF. B) The response of basophils to fMLF expressed as the difference in MFI between stimulated and unstimulated cells. The symbols (circle, triangle, square) represent three different donors. Means and standard deviations are shown in Supplementary Table S1. Data analysed with paired *t* test. n.s., not significant.



**Figure S4. CD16/CD62L expression of neutrophils that were activated in vivo in response to intravenous LPS infusion.**

Two extra neutrophil subsets ( $CD16^{dim}CD62L^{hi}$  and  $CD16^{hi}CD62L^{dim}$ ) can be identified in both A) fresh and B) fixed and cryopreserved cells.



**Figure S5. Expression of CD203c, CD11b and CD35 on basophils collected in a field study.**

Samples were cryopreserved for 2 years in liquid nitrogen. A) Histograms of a representative sample, filled lines display a fluorescence-minus-one (FMO) sample, the thin black line unstimulated cells and the thick black line fMLF-stimulated cells. B) Median fluorescent intensity (MFI) of markers in basophils from 10 donors left unstimulated and stimulated with fMLF. Means and standard deviations are shown in Supplementary Table S2. \*indicates analysis with Wilcoxon matched-pairs signed rank test. \*\*,  $P \leq 0.01$ .

Supplementary Tables

Supplementary table S1.

Basophils	Fresh Mean MFI (SD)		Fixed Mean MFI (SD)		Fix & Freeze Mean MFI (SD)	
	Medium	fMLF	Medium	fMLF	Medium	fMLF
CD203c	682 (260)	3172 (3844)	1397 (417)	4661 (4840)	1288 (401)	3542 (3007)
CD11b	234 (69)	368 (261)	191 (21)	332 (226)	219 (15)	353 (207)
CD35	582 (148)	1361 (1122)	886 (157)	1618 (938)	888 (157)	1538 (779)

Expression of CD203c, CD11b and CD35 on basophils and the response to fMLF.

Supplementary table S2.

Neutrophils	Medium	fMLF
CD11b	1237 (151)	1527 (181)
CD35	1895 (728)	3203 (735)
CD66b	543 (59)	713 (164)
CD62L	641 (278)	105 (23)
Eosinophils	Medium	Eotaxin
CD35	3044 (596)	3933 (714)
CD11b	1147 (438)	2079 (550)
Basophils	Medium	fMLF
CD203c	1374 (615)	4587 (1779)
CD11b	133 (20)	249 (80)
CD35	880 (275)	1663 (718)

Expression of activation markers on neutrophils, eosinophils and basophils from 10 representative donors collected in a field study presented as mean MFI and standard deviation.

# 3

## CHAPTER 3



# Multi-dimensional flow cytometry analysis reveals increasing changes in the systemic neutrophil compartment during seven consecutive days of endurance exercise

Selma van Staveren<sup>1,2</sup>, Twan ten Haaf<sup>3</sup>, Margot Klöpping<sup>1</sup>, Bart Hilvering<sup>1</sup>, Gerjen H. Tinnevelt<sup>4,2</sup>, Karin de Ruiter<sup>5</sup>, Maria F. Piacentini<sup>6</sup>, Bart Roelands<sup>7</sup>, Romain Meeusen<sup>7,8</sup>, Jos J. de Koning<sup>3,9</sup>, Jeroen J. Jansen<sup>4</sup>, Nienke Vrisekoop<sup>1</sup> and Leo Koenderman<sup>1</sup>

1 Department of Respiratory Medicine, Laboratory of Translational Immunology, UMC Utrecht, The Netherlands

2 TI-COAST, Amsterdam, The Netherlands

3 Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, MOVE Research Institute Amsterdam, The Netherlands

4 Analytical Chemistry, Institute for Molecules and Materials, Radboud University Nijmegen, The Netherlands

5 Department of Parasitology, LUMC Leiden, The Netherlands

6 Department of Human Movement and Sport Sciences, University of Rome "Foro Italico", Italy

7 Human Physiology Research Group, Vrije Universiteit Brussel, Brussels, Belgium.

8 School of Public Health, Tropical Medicine and Rehabilitation Sciences, James Cook University, Queensland, Australia.

9 Department of Exercise and Sports Science, University of Wisconsin-La Crosse, WI.

Endurance exercise is associated with a transient increase in neutrophil counts in the peripheral blood. Here we investigate the impact of intensified endurance exercise on the neutrophil compartment. We hypothesized that intensified endurance exercise leads to mobilization of neutrophil subsets, which are normally absent in the blood. Furthermore, we followed the potential build-up of neutrophil activation and the impact on overnight recovery of the neutrophil compartment during a seven-day cycling tour.

The neutrophil compartment was studied in 28 healthy amateur cyclists participating in an eight-day strenuous cycling tour. Blood samples were taken at baseline, after 4 days and after 7 days of cycling. The neutrophil compartment was analyzed in terms of numbers and its phenotype by deep phenotyping of flow cytometry data with the multi-dimensional analysis method FLOOD.

Repeated endurance exercise led to a gradual increase in total neutrophil counts over the days leading to a 1.26 fold-increase (95%CI 1.01-1.51  $p=0.0431$ ) in the morning of day 8. Flow cytometric measurements revealed the appearance of 2 additional neutrophil subsets: CD16brightCD62Ldim and CD16dimCD62Lbright. A complex change in neutrophil phenotypes was present characterized by decreased expression of both CD11b and CD62L and marked increased expression of LAIR-1, VLA-4 and CBRM1/5. The changes in expression were found on all neutrophils present in the blood. Strikingly, in strong contrast to our findings during acute inflammation evoked by LPS challenge, these neutrophils did not upregulate classical degranulation markers. In fact, our FLOOD analysis revealed that the exercise induced neutrophil phenotype did not overlap with the neutrophil subsets arising upon acute inflammation.

In conclusion, during multiple days of endurance exercise the neutrophil compartment does not regain homeostasis overnight. Thereby our study supports the concept of a build-up of inflammatory cues during repeated endurance exercise training, causing a prolonged change of the systemic neutrophil compartment.

## Introduction

Peripheral blood neutrophil counts increase upon aerobic and anaerobic single exercise bouts and recover within 6-24 hours.<sup>1</sup> Multiple non-mutually exclusive mechanisms underlie blood neutrophilia, but little is known regarding the mechanisms of neutrophil recruitment and recovery upon prolonged exercise. A major part of the recruitment of neutrophils after at least short-term exercise is originating from the so-called marginated pool, which consists of neutrophils that are not free flowing in the blood and are (at least in part) associated with the vasculature. This pool is thought to be in complete equilibrium with the free flowing pool.<sup>2-7</sup> Steroids and/or epinephrine are well-known mediators that can liberate these cells from the vasculature. Steroids have been implicated in relation to neutrophilia evoked by exercise, because rises in plasma adrenocorticotrophic hormone (ACTH) and cortisol levels coincide with the demargination of neutrophils.<sup>4,8,9</sup> In addition, treatment with high concentrations of glucocorticoids leads to neutrophilia.<sup>10</sup> On the other hand, stable or even decreasing levels of plasma ACTH and cortisol have also been described upon exercise. These latter studies question the hypothesis that systemic rises of these hormones are the main mechanism for the observed neutrophilia.<sup>11-14</sup> Furthermore, neutrophil counts can increase up to 3-4 fold upon exercise,<sup>1</sup> whereas only 50% of total blood neutrophils is thought to be present in the marginated pool.<sup>3,4</sup> Therefore, the observed increase in neutrophils does not seem to be solely explained by demargination.

Neutrophilia can also be caused by the mobilization of extra neutrophil subsets from other tissue sites (as reviewed by Pillay et al).<sup>15</sup> Under homeostasis peripheral blood neutrophils are homogeneous and express FcγRIII (CD16) and L-selectin (CD62L) within a narrow range, typically characterized as CD16<sup>bright</sup>CD62L<sup>bright</sup> neutrophils. We have previously described the mobilization of two additional neutrophil subsets in response to acute inflammation such as evoked by a systemic LPS challenge in healthy volunteers. These mobilized subsets can be identified by differential expression of CD16 and of L-selectin.<sup>16</sup> One subset consists of CD16<sup>dim</sup>CD62L<sup>bright</sup> cells, with predominantly band-shaped nuclei. These cells are more immature than the CD16<sup>bright</sup>CD62L<sup>bright</sup> cells and are likely recruited from the bone marrow.<sup>17</sup> Interestingly, previous publications reporting neutrophilia after endurance exercise also reported a specific increase in neutrophils with a band-shaped nucleus.<sup>18-20</sup> Thus, one might hypothesize that this same CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophil subset is recruited during neutrophilia evoked by exercise. The second extra subset found in the peripheral blood upon LPS challenge consists of CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophils with mainly hypersegmented nuclei that exhibit a more activated and immunosuppressive phenotype.<sup>16</sup> Interestingly, this latter subset is remarkably inefficient in killing bacteria such as *Staphylococcus aureus*.<sup>21</sup> The relative increases in these subsets can have a considerable impact on the expression of the classical activation/degranulation markers CD35 and CD11b of the total neutrophil pool.<sup>16</sup> This possibly explains why studies staining

for a single marker or a very limited number of markers after single exercise bouts have provided conflicting results.<sup>22–24</sup>

Currently, multi-color analysis of immune cells is the standard in medical immunology but this needs new multi-dimensional methods to adequately analyze these data. This has led to the development of data analysis tools that provide a multi-dimensional view on all markers expressed on individual cells. We have developed such a tool, FLOOD<sup>25</sup>, that enables the recognition of multi-dimensional differences in neutrophil phenotypes by staining with many markers in addition to CD62L and CD16. Using this analysis method, we have previously reported that only normal neutrophils (CD16<sup>bright</sup>CD62L<sup>bright</sup>) are mobilized from the marginated pool upon a single anaerobic exercise bout.<sup>25</sup>

To get more insight into the potential role of neutrophils in the exercise-induced immune response during relatively long periods of increased training we investigated neutrophils during seven days of increased intensive endurance training. We applied multi-dimensional FLOOD analysis of peripheral blood neutrophils allowing deep phenotyping of these cells. We tested the hypothesis that the neutrophil compartment in peripheral blood would fail to recover overnight upon repetitive days of strenuous exercise.

## Materials and Methods

### Subjects, study design and blood sampling

#### Tour for Life

The subjects were all healthy recreational cyclists, 11 male and 19 female aged 25–57 years (mean±SD age 40.8±10.8 y, BMI 23.5±2.1 kg/m<sup>2</sup>). For demographics see Table 1. At baseline the VO<sub>2</sub>max was 51.8±6.3 ml/kg/min and a peak power output of 4.12±0.57 W/kg was measured (performance level 2<sup>26,27</sup>). Subjects reported an estimated training volume of 144±46 km per week during preparation. Written informed consent was obtained from all participants. The study was conducted in accordance with the Declaration of Helsinki and approved by the institutional ethical committee.

Table 1 - Demographics

Baseline characteristics for the subjects	mean ± SD
Gender (M/F)	19/11
Age (y)	40.8±10.8
BMI (kg/m <sup>2</sup> )	23.5±2.1
VO <sub>2</sub> max (ml/min/kg)	51.8±6.3
PPO (W/kg)	4.12±0.57
training volume pre TFL (km/week)	144±46

BMI: Body Mass Index; PPO: Peak Power Output

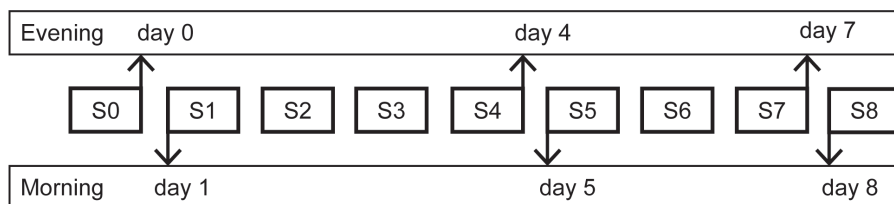
This study was conducted during the Tour for Life (TFL). The subjects cycled from Italy to The Netherlands in 8 consecutive days (1 stage per day) together with 400 other contestants. The length, altimeters and duration per stage is shown in Table 2. The start of each stage was between 07:00h and 08:00h and the finish between 16:00h and 19:30h, depending on the fitness level of the subject and the length of the stage. At 3 evening time point and at 3 mornings during the TFL blood was sampled (Figure 1). The researchers followed the cyclists in a mobile laboratory van that enabled blood sampling, processing and freezing.

At the mornings of day 1, day 5 and day 8 and the evenings of day 0, day 4 and day 7, blood of the contestants was collected. For cellular analyses a venous blood sample (4 ml) was collected at each occasion by venipuncture from an antecubital vein in a vacutainer tube (BD Vacutainer) containing sodium heparin as anticoagulant. For hormonal analysis at each occasion venous blood samples were collected in an EDTA vacutainer tube (3 ml, VenoSafe) and in a serum tube containing gel and clot activator (8 ml, Venosafe).

Table 2 - Training load per stage

Stage (day)	Distance (km)	Climbing (m)	Mean duration ± SD (min)
1	110	2800	387 ± 65
2	160	4200	592 ± 112
3	175	2900	564 ± 127
4	190	1900	531 ± 121
5	129	2450	420 ± 108
6	175	900	414 ± 60
7	215	2200	558 ± 83
8	110	1200	N.D.
Total	1264	18550	3466 ± 462

N.D.: no data available



**Figure 1. Study design.**

Blood was collected at 6 different time points during the TFL: before the start of the TFL in the evening of day 0 and in the morning of day 1; halfway during the TFL in the evening directly after stage 4 and in the morning before stage 5; at the end of the TFL in the evening directly after stage 7 and in the morning before stage 8. SX = Stage X.

### Experimental human endotoxemia model

Experiments were part of a larger endotoxin trial (NCT02629874 at [www.clinicaltrials.gov](http://www.clinicaltrials.gov)). Healthy male volunteers 19-24 years old were enrolled after screening and were prehydrated from  $t = -1$  h to  $t = 0$  (2.5% glucose/ 0.45% saline at a continuous rate of 1.5 L/h) through a cannula placed in an antecubital vein. U.S. Reference E. coli endotoxin (O:113, NIH Pharmaceutical Development Section, Bethesda, MD, USA) was used as systemic challenge in this study. Endotoxin was reconstituted in 5 ml saline and injected as a single intravenous dose of 2 ng/kg during 1 min at  $t = 0$ . After LPS injection, subjects were infused with hydration fluid at a continuous rate of 150 mL/h. Heart rate and blood pressure of the volunteers were monitored from  $t = -1$  h until discharge at  $t = 8$  h, as well as the course of LPS-induced symptoms.<sup>28</sup> The study protocol was approved by the Ethics Committee of the Radboud University Nijmegen Medical Centre, The Netherlands and complies with the Declaration of Helsinki and ICH Good Clinical Practice guidelines. Written informed consent was obtained from all participants.

At  $t=180$  min blood samples, anticoagulated with sodium heparin, were taken from the arterial catheter.

### Cell counts and flow cytometry

For the TFL samples as well as the LPS samples, 1 ml of sodium heparin blood was analyzed by a field applicable method that has been previously described in detail.<sup>29</sup> In Short, 1 ml of blood was diluted in 20 ml 1:10 FACS lysing solution (Becton Dickinson, Mountain View, CA) and incubated at room temperature for 15 min for red blood cell lysis and simultaneous leukocyte fixation. The samples were centrifuged at  $430 \times g$  for 5 min at room temperature. The cells were washed twice with RPMI + 10%FCS. After the last washing step, the cells were transferred to a  $-30^{\circ}\text{C}$  freezer in RPMI + 10%FCS + 10%DMSO.<sup>29</sup> Ten days after collection and freezing, the samples were transferred to a  $-80^{\circ}\text{C}$  freezer.

At the time of analysis the cells were thawed by adding 1 ml RPMI + 10%FCS at room temperature. The samples were centrifuged 430 x g for 5 min at 4°C and the cells were re-suspended in PBS + 0.32% sodium citrate + albumin (4 g/l). Cell counts were determined by an automated cell counter (CellDyn Emerald, Abbott Diagnostics). Per sample 500.000 cells were stained with a combination of 10 monoclonal antibodies and measured on a BD-LSR Fortessa flow cytometer (Becton Dickinson, Mountain View, CA). The antibodies used were as follows: CD35-FITC (clone E11), CD64-APC (clone 10,1), CBRM1/5-Alexa Fluor 700 (clone CBRM1/5), CD11b-APC-Alexa Fluor 750 (clone Bear1), CD305 (LAIR-1)-PE (clone DX26), CD14-eF450 (clone 61D3), CD16-Krome Orange (clone 3G8), CD62L-BV650 (clone DREG 56), CD49d-PECy7 (clone G9F10), CD66b-PerCPCy5.5 (clone G10F5).

### Hormonal measurements

Peripheral blood collected in EDTA tubes was centrifuged 1500 x g for 10 min and the plasma fraction was frozen at -30°C. After 8 days all plasma samples were transferred to a -80°C freezer until further analysis. Plasma ACTH concentrations were measured using an electrochemiluminescence immunoassay on the Cobas E411 (Roche Diagnostics GmbH, D-68298 Mannheim, Germany). The lower limit of detection was 2.0 ng/L and inter-assay variation was 5,7 – 3.5% at 8.5 – 170 ng/L respectively (n = 18). Intra-assay variation was <1.2%. Normative values: morning measurements 5 - 70 ng/L; afternoon measurements < 50 ng/L.

The serum tubes containing blood were directly incubated at 4°C for 30 min to allow clotting. This was followed by centrifugation 1500g for 10 min. The serum fraction was frozen at -30°C within 1 hour after blood collection. After 8 days all serum samples were transferred to a -80°C freezer until further analysis. Serum concentrations of cortisol were determined by a chemiluminescence immunoassay on a Unicel DxI 800 (Beckman Coulter, Bea, CA, USA). Inter-assay variation was 5.4 – 8.2 % at 0.50 – 1.05 µmol/L, respectively. Intra-assay variation 3.4 – 6.2% at 0.50 – 1.05 µmol/L, respectively. Normative values: morning measurements 0.20-0.65 µmol/L; afternoon measurements 0.10-0.35 µmol/L.

### Statistics

Two subjects were excluded from statistical analyses due to missing data on 1 or 2 time points, we used a one-way repeated-measures ANOVA test (GraphPad Prism 6.0). From the ANOVA results, we calculated fold-changes and their 95% CI to determine the statistical significance of the differences observed between time points.

The only exception was the statistical analysis of the MFI's of the three neutrophil subsets over time, for this analysis a two-way repeated-measures ANOVA test (GraphPad Prism 6.0) was used.

### Multi-dimensional flow cytometry analysis by FLOOD

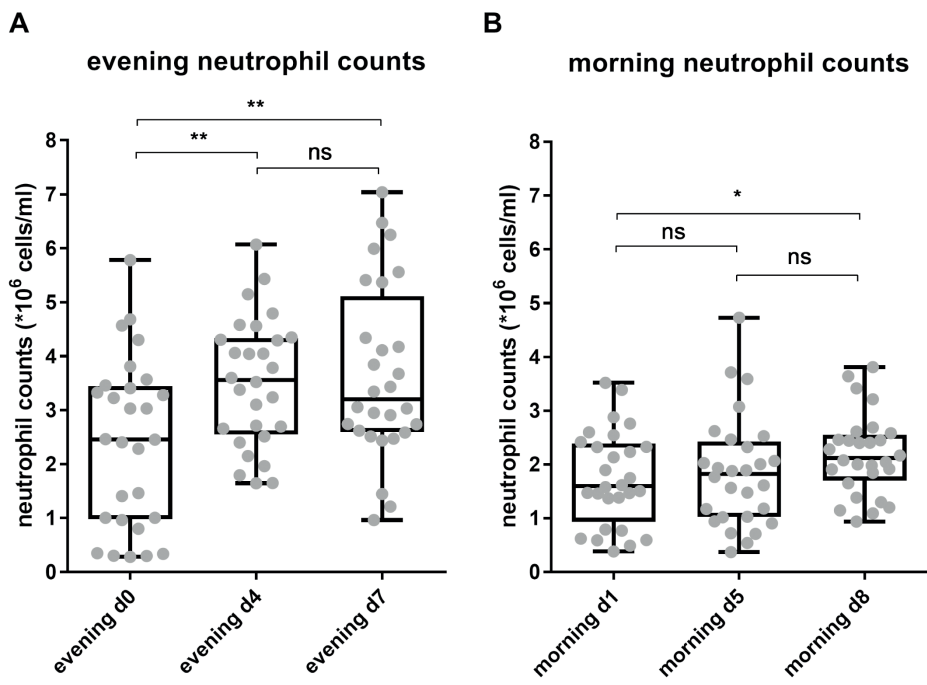
For multi-dimensional analysis of the flow cytometry data the PCA based method FLOOD was used.<sup>25</sup> The neutrophils were identified from the original fcs files based on FSC/SSC and CD16+CD66b+ expression (Figure S1) (FlowJo analysis software, Tree Star Inc., Ashland, Oregon). Only cells within this gate were exported as fcs3.0 files. These files were imported into the FLOOD software.<sup>25</sup> Two FLOOD analyses were performed. For both analyses the control group consisted of all morning day 1 samples. For the first analysis the evening day 0, evening day 4, morning day 5, evening day 7 and morning day 8 time points were labeled as response group. For the second analysis the response group consisted of evening day 0, evening day 4, morning day 5, evening day 7 and morning day 8 time points as well as the LPS samples. A value equal to the most negative intensity -1 was subtracted, resulting in values of >1, to deal with negative intensities as a result of background subtraction by the flow cytometry software and as a result of compensation. Preprocessing steps consisted of log transformation, mean centering based on the control group and scaling based on controls. For further details on the approach we refer the reader to the original manuscript.<sup>25</sup>

## Results

### Neutrophil counts did not recover to baseline after 7 days of intensified endurance exercise

Neutrophil kinetics were studied in 30 healthy amateur cyclists during seven days strenuous cycling (mean daily distance of 160 km and 2300 altimeters). The exercise volume of the contestants increased on average with  $770 \pm 24\%$  during the week of exercise compared to their regular training schedule. We observed an exercise-induced increase in neutrophil counts in the evening of day 4 and day 7 when compared to evening day 0 (Table 3, Figure 2A). Comparing neutrophil counts in the morning during TFL, we found that at day 5 during the TFL the total neutrophil count normalized overnight to baseline levels. After 7 days of cycling, an increased number of cells was detected after overnight recovery (1.26 fold-increase; 95%CI 1.01-1.51  $p=0.0431$  (Table 4, Figure 2B). Thus, neutrophilia did not recover at the morning of day eight after a seven-day strenuous cycling tour.





**Figure 2. Total peripheral blood neutrophil counts increased upon daily endurance exercise and did not recover to baseline overnight after 7 days of cycling.**

Neutrophil counts were measured at (A) 3 evening time points and (B) 3 morning time points during the TFL.

\*  $p=0.0431$ ; \*\*  $p<0.005$ ; ns = not significant.

**Table 3 – Evening peripheral blood neutrophil counts**

	Cell counts		Mean difference	95% CI of difference
	day 0	follow-up		
Evening d0 vs. evening d4	2.40	3.51	1.11	0.38-1.85 <sup>a</sup>
Evening d0 vs. evening d7	2.40	3.67	1.26	0.44-2.09 <sup>a</sup>

Cell counts in  $10^6$  cells/ml  
a  $p<0.001$

**Table 4 - Neutrophil subset counts**

	Cell counts			Mean difference		
	day 1	day 5	day 8	day 5 vs. day 1	day 8 vs. day 1	day 8 vs. day 5
<b>Total neutrophils</b>	1.78	1.89	2.24	0.11	0.46 <sup>a</sup>	0.35
<b>CD16<sup>bright</sup>CD62L<sup>bright</sup></b>	1.49	1.39	1.44	-0.11	-0.05	0.05
<b>CD16<sup>bright</sup>CD62L<sup>dim</sup></b>	0.22	0.37	0.58	0.15 <sup>a</sup>	0.36 <sup>b</sup>	0.21 <sup>c</sup>
<b>CD16<sup>dim</sup>CD62L<sup>bright</sup></b>	0.02	0.05	0.07	25.8	0.04 <sup>c</sup>	0.02

Cell counts in 10<sup>6</sup> cells/ml

a p<0.05

b p<0.0001

c p<0.005

### **The maintained increases of neutrophil subsets counts were independent of changes in peripheral blood ACTH or cortisol levels**

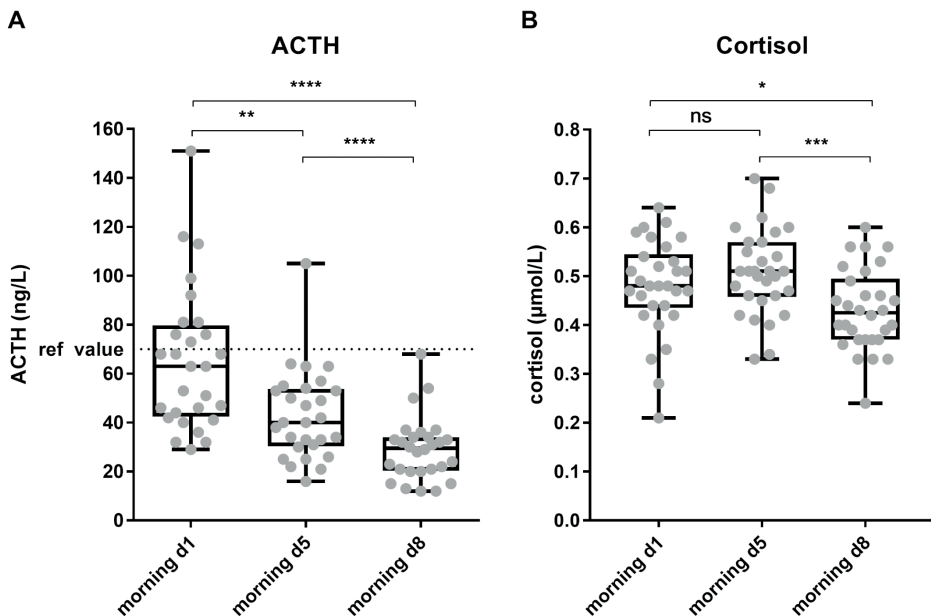
To test the hypothesis that the maintained neutrophilia after prolonged exercise was caused by a rise in plasma ACTH and/or cortisol levels, we studied the relation between the increase in neutrophil counts and the ACTH and cortisol concentrations in the plasma. We measured these hormones at the same time point in the 3 mornings. Surprisingly, we found a clear decrease in ACTH plasma concentration during the TFL (Figure 3A). For cortisol there was no significant change in concentration halfway during the TFL, but levels were also significantly decreased at day 8 compared to day 1 (Figure 3B).

### **Intensified endurance exercise led to mobilization of different neutrophil subsets**

The finding that the impaired recovery of the innate immune response did not coincide with increased ACTH and/or cortisol levels implied that the maintained neutrophilia could not simply be explained by steroid-induced demargination of neutrophils into the circulation. To study whether the neutrophilia was caused by the mobilization of extra neutrophil subsets such as evoked by a systemic LPS challenge in healthy volunteers <sup>16</sup>, we analyzed these cells in the context of their differential expression of CD16 and CD62L. During the TFL we collected blood samples and stored the leukocyte fractions by freezing. Once back in The Netherlands we thawed the cells and performed flow cytometry analysis as described by de Ruiter et al <sup>29</sup>. We confirmed that this method still enabled us to distinguish between the different neutrophil subsets (Figure S2A).

When analyzing the counts of the individual neutrophil subsets we found marked changes over time (Figure 4A). Normal segmented CD16<sup>bright</sup>CD62L<sup>bright</sup> neutrophil counts remained stable over all morning time points (0.93 fold-increase (95% CI 0.65-1.19; p=0.80) at day 5 and 0.96 fold-increase (95% CI 0.69-1.23; p=0.95) at day 8 when compared to day 1),

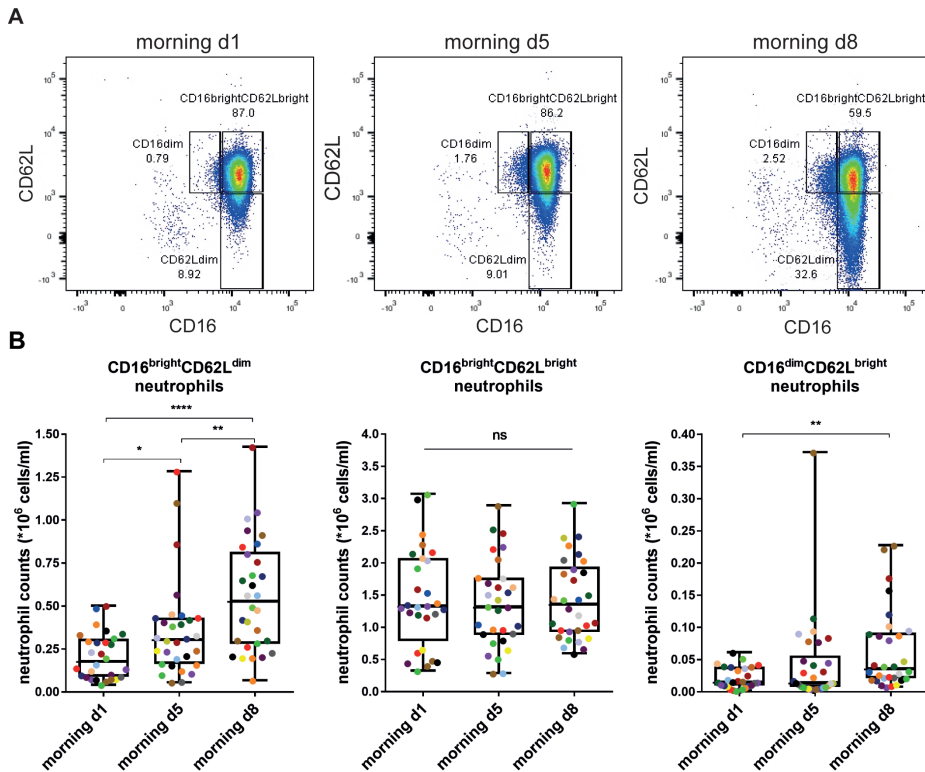
whereas CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophils were significantly increased by 1.68 fold at morning day 5 (95% CI 1.04-2.39;  $p=0.0376$ ) and by 2.64 fold at morning day 8 (95% CI 1.95-3.37;  $p<0.0001$ ) when compared to morning day 1. Also a 3.00 fold-increase (95% CI 1.45-4.55;  $p<0.005$ ) in CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils was found at morning day 8 when compared to morning day 1 (Table 4, Figure 4B). Thus the sustained neutrophilia after exercise can be at least in part explained by the appearance of extra neutrophil subsets into the peripheral blood.



**Figure 3. A decrease in ACTH plasma concentration was found over time, while serum cortisol concentrations showed a delayed decrease.**

A) Plasma ACTH concentrations in ng/L at morning day 1, day 5 and day 8 (n=28). B) Serum cortisol concentrations at morning day 1, day 5 and day 8 are depicted in μmol/L (n=29).

\*  $p=0.0323$ ; \*\*  $p<0.005$ ; \*\*\*  $p<0.0005$ ; \*\*\*\*  $p<0.0001$ .



**Figure 4. Increasing numbers of CD16<sup>bright</sup>CD62L<sup>dim</sup> and CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils are found.**

A) Representative FACS plot of CD16 vs. CD62L expression of the total neutrophil population of one individual at the 3 morning time points during the TFL. B) graphs of the three CD16CD62L neutrophil subset counts at morning day 1, morning day 5 and morning day 8 per individual (n=28). Colors visualize the individual changes over time. ns: not significant; \* p<0.05; \*\* p<0.005; \*\*\*\* p<0.0001.

### Intensified endurance exercise led to complex changes in the phenotype of the different neutrophil subsets

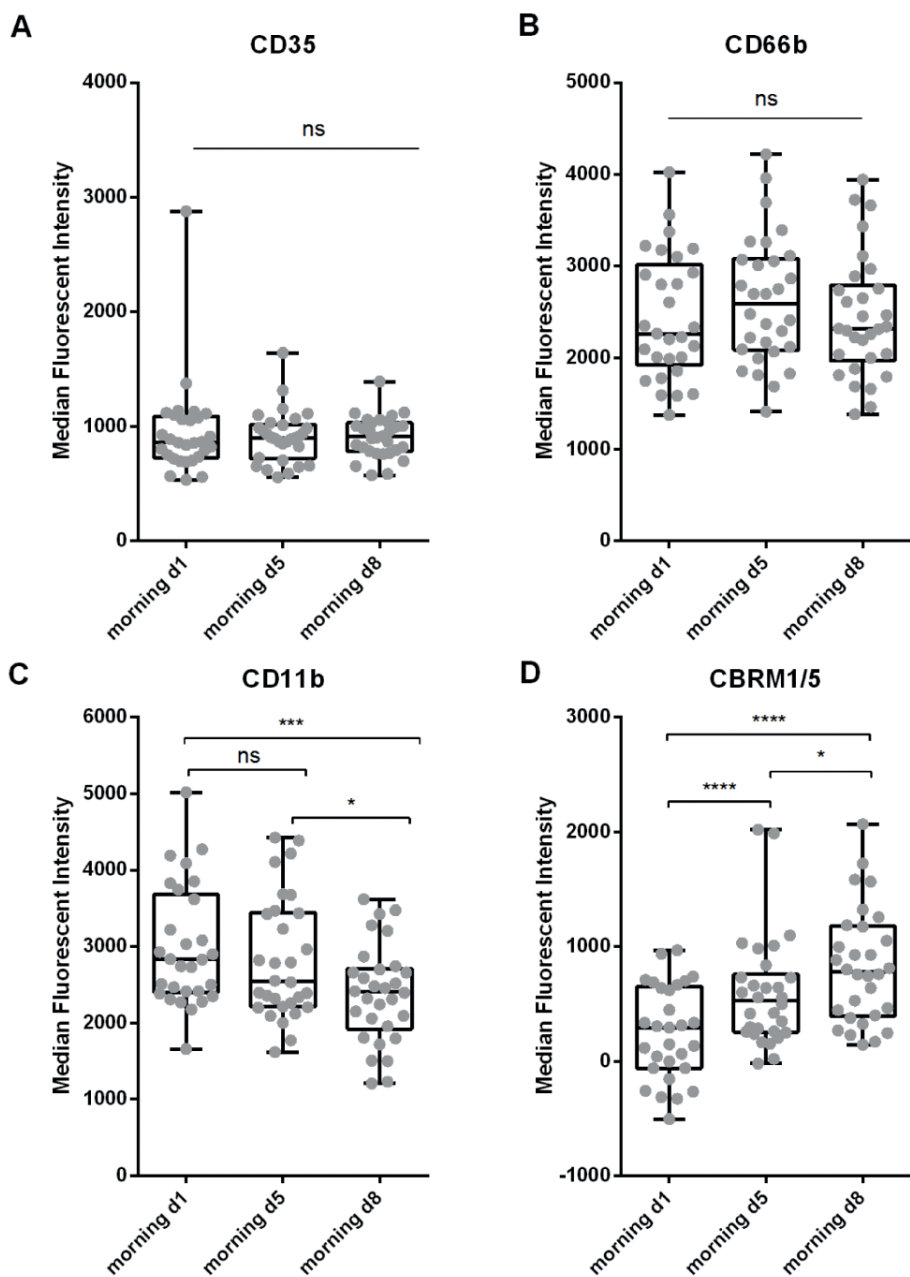
Next to determination of the expression CD16 and CD62L to identify the different subsets, we studied the expression of the classical activation/degranulation markers CD35, CD66b and CD11b by flow cytometry. In addition, CD49d (very late antigen 4, VLA-4) and LAIR-1 (CD305) were included in the panel as markers for immature or alternatively activated neutrophils.<sup>30–33</sup> Of note, our freeze thaw method possibly permeabilizes granulocytes and, therefore, both extracellular and intracellular proteins might be measured. However, activation markers and their responsiveness to stimuli like fMLF could still be clearly measured.<sup>29</sup> Also, when employing this protocol on peripheral blood of LPS challenged

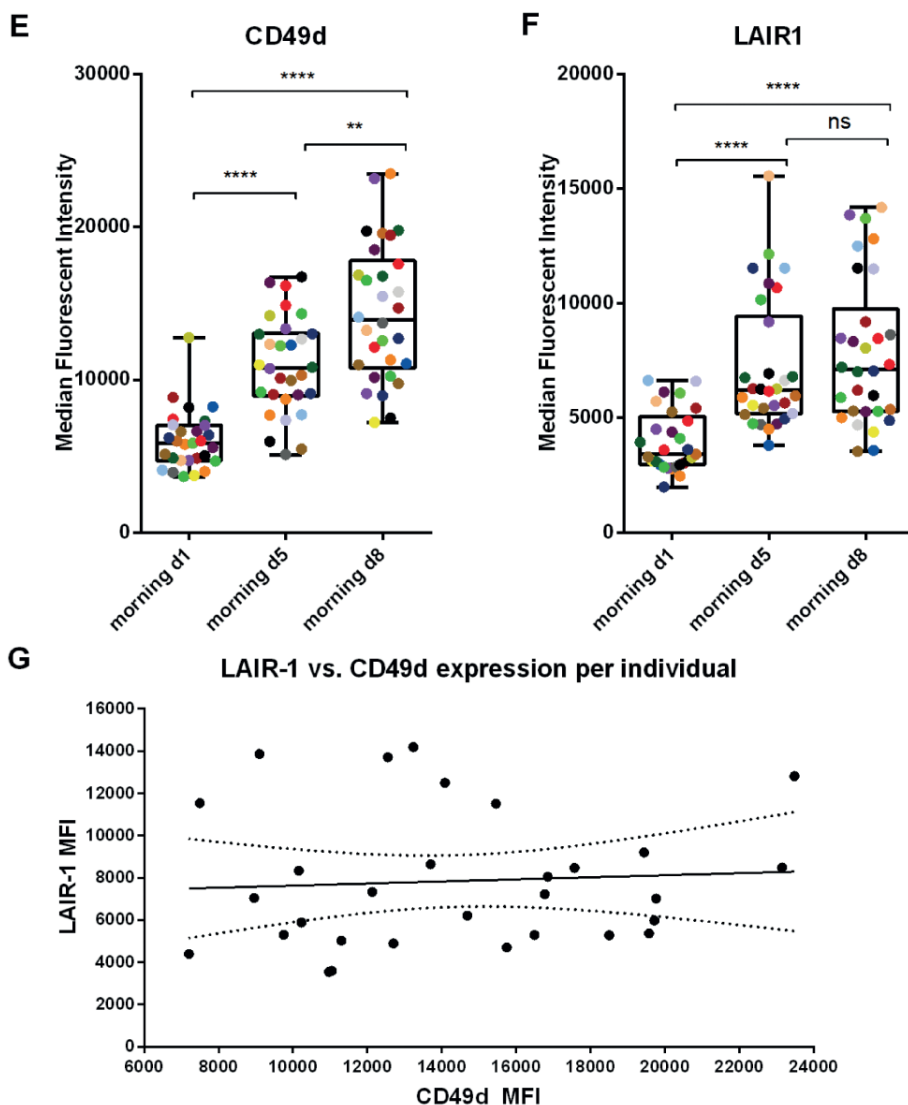
volunteers, we did find differences in expression of the classical degranulation markers CD35 and CD11b between LPS subsets (Figure S2B and C), which was in line with previous results.<sup>16</sup> During the TFL, the classical activation/degranulation markers CD35 and CD66b were unaffected (Figure 5A, B). CD11b expression was lowered (Figure 5C), whereas the expression of CBRM1/5, an epitope expressed by CD11b in its activated state, significantly increased at day 5 (2.4 fold (95% CI 1.80-3.01) and at day 8 (3.3 fold (95% CI 2.57-4.03) (Figure 5D). The expression of CD49d increased markedly over the time course of the TFL from 1.80 fold-increase at day 5 (95% CI 1.57-2.04) to a 2.38 fold-increase at day 8 (95% CI 2.02-2.73) compared to day 1 (Figure 5E). The expression of LAIR-1 was increased by 1.82 fold (95% CI 1.53-2.10) halfway during the TFL and this increase was maintained at 2.00 fold (95% CI 1.68-2.33) at the end of the TFL when compared to morning day 1 (Figure 5F). The results of the statistical analysis are depicted in Table 5. At day 5 two groups are visible when analyzing LAIR-1 expression: part of the participants showed a steep upregulation of LAIR-1 expression, while the other participants upregulated LAIR-1 moderately. Most of the participants with high LAIR-1 expression at day 5 also highly expressed LAIR-1 at day 8. For CD49d we also found that most participants with high expression at day 5 maintained high expressions at day 8. However, no correlation was found between the MFI's of LAIR-1 and CD49d ( $R^2= 0.0046$ ,  $p=0.7218$ , Figure 5G).

**Table 5 – Median Fluorescent Intensities of 6 neutrophil markers**

	MFI			MFI mean difference		
	day 1	day 5	day 8	day 5 vs. day 1	day 8 vs. day 1	day 8 vs. day 5
CD35	951	905	909	-47	-42	4
CD66b	2442	2638	2441	-196	1	198
CD11b	2989	2842	2380	-147	-609 <sup>a</sup>	-462 <sup>b</sup>
CBRM1/5	250	602	827	352 <sup>c</sup>	576 <sup>c</sup>	225 <sup>b</sup>
CD49d	6039	10891	14346	4853 <sup>c</sup>	8308 <sup>c</sup>	3455 <sup>d</sup>
LAIR1	3967	7209	7953	3242 <sup>c</sup>	3986 <sup>c</sup>	744

MFI's are in arbitrary units  
a  $p < 0.0005$   
b  $p < 0.05$   
c  $p < 0.0001$   
d  $p < 0.005$





**Figure 5. Intensified endurance exercise leads to a marked change in the peripheral blood neutrophil phenotype.**

Median fluorescent intensities for A) CD35, B) CD66b, C) CD11b, D) CBRM1/5, E) CD49d and F) LAIR-1 are depicted for all individuals at the 3 morning time points day 1, day 5 and day 8 ( $n=30$ ). Colors in E and F visualize the individual changes over time. G) Scatter plot of the MFI of CD49d at day 8 vs. the MFI of LAIR-1 at day 8 per participant ( $R^2=0.0046$ ,  $p=0.7218$ ), dotted lines represent the 95% CI. ns: not significant; \*  $p<0.05$ ; \*\*  $p<0.005$ ; \*\*\*  $p<0.0005$ ; \*\*\*\*  $p<0.0001$ .

Since three different neutrophil subsets were found in the peripheral blood of the participants, the expression of above mentioned markers per neutrophil subset were also analyzed. Only small differences were found. The expressions expression of CD35 and CD49don CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils were significantly decreased when compared to mature CD16<sup>bright</sup>CD62L<sup>bright</sup> neutrophils at morning day 1 and day 8, respectively. No other differences in expression were found between the subsets (Figure S3).

### **Multi-dimensional FLOOD analysis identified the exercise-induced changes in marker expressions that are found on the total population of peripheral blood neutrophils**

The flow cytometry results so far were obtained by classically analyzing the data in a univariate or bivariate (CD16/CD62L) manner. We next evaluated whether these changes in marker expressions were correlated to each other. Moreover, we aimed to find out whether other neutrophil subsets than the CD16/CD62L subsets were responsible for the changes in MFI's or whether the total neutrophil population exhibited up- or downregulation of the markers. To answer this question, we performed a computational Principal Component Analysis (PCA) based multivariate analysis: FLOW Cytometric Orthogonal Orientation for Diagnosis (FLOOD) on the TFL data <sup>25</sup>. In the analysis the morning day 1 samples (control group) were compared to the evening day 0, evening day 4, morning day 5, evening day 7 and morning day 8 samples (defined as the response groups). In short, a response specific multi-dimensional model ('response model') was created in which the single cell variation within the response group was best explained. The FLOOD output consists of multi-dimensional plots (Figure 6) projected in 2D <sup>25</sup>. Per individual six plots were generated: one plot per time point. The plots contain data of one sample of a representative individual, and the data of all participants at morning day 1 (80% of the morning day 1 cells of all the participants are within the cyan benchmark). The data of the representative individual is encompassed by a blue benchmark at morning day 1 and by a red benchmark at evening day 0, evening day 4, morning day 5 evening day 7 and morning day 8. The relative importance of a marker for the description of the multi-dimensional space is depicted as the length of the vector. The relation in expression of the markers is depicted in the angle of the vectors <sup>25</sup>.

For the TFL dataset 54% of the total response specific variance is shown in the PC1/PC2 plots (explaining 33% and 21% of the response specific variation, respectively). The results of the FLOOD analysis were as follows:

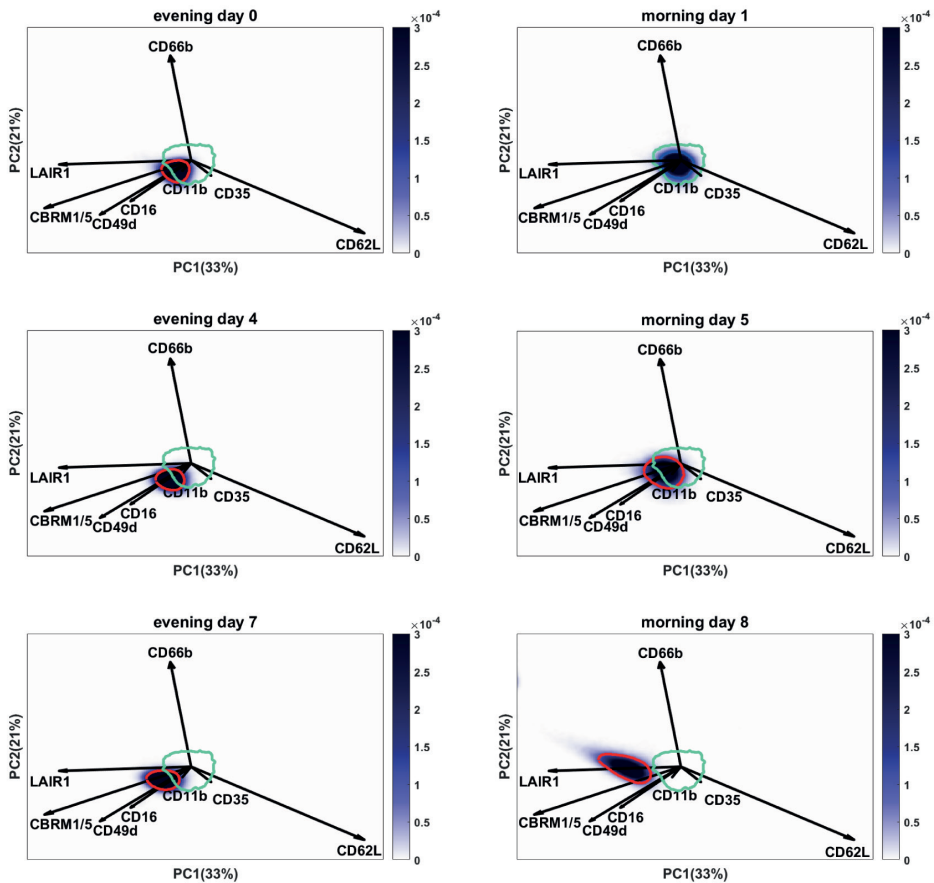
Firstly, an exercise induced shift of the peripheral blood neutrophil population was visible towards the direction of vectors of markers CBRM1/5, CD49d, LAIR-1 and CD16 (these vectors represent the PCA loadings). This was both found at evening day 4 as well as at evening day 7. Strikingly, the shift of the neutrophil population did not recover overnight.



In fact, the shift was even more pronounced at morning day 8. At morning day 8 the shift was directed more clearly in the opposite direction of the CD62L loading vector, indicating that the cells were relatively more CD62L<sup>dim</sup> when compared to the neutrophils at the previous time points. The loading of CD66b is orthogonally oriented to the loading vectors of LAIR-1, CBRM1/5, CD49d, CD16 and loadings CD62L and CD35. This implies there was no correlation between CD66b and the other markers. The CD16 loading vector is aligned with CD49d, CBRM1/5 and LAIR-1, but the length of the loading vector is relatively short, which indicates its minor importance for the variance explained by this model. The same accounts for CD11b and CD35, which have very short loading vectors. Therefore, CD11b, CD35 and CD16 did not contribute to the model. From the FLOOD plots one can appreciate that the shift of the whole cell population is quite dramatic: the cells at day 8 did not overlap with the region where the cells of the day 1 sample were found. This means that all the peripheral blood neutrophils changed their phenotype.

Secondly, besides a shift of the total neutrophil population, the cell populations at the morning time points have a larger surface area than the post-exercise cell population at evening day 4 and evening day 7. This indicates that the neutrophils at evening day 4 and evening day 7 were a more homogeneous cell population with less variance between the cells when compared to the morning time points. In addition, the total neutrophil population was more elongated in a diagonal direction at day 8. The elongation aligns to the CD62L loading, meaning that at day 8 the variance in CD62L expression was larger when compared to the previous time points. Together these findings corresponded well with what we found by univariate and bivariate analysis.

Thirdly, this analysis shows that the increase in CBRM1/5, LAIR-1 and CD49d expression was found on all neutrophils irrespective of their expression of CD16 and CD62L. After 4-7 days of intensive endurance exercise all neutrophils in the peripheral blood relatively express more LAIR-1, CBRM1/5 and CD49d. Overnight the neutrophil population did not recover and the cells were even more CD62L<sup>dim</sup> at morning day 5 and day 8 (which in the example subject of Figure 6 is more prominent at day 8 when compared to day 5).



**Figure 6. Multi-dimensional FLOOD analysis shows results found by conventional gating analyses in one biplot.**

FLOOD output: neutrophil deep phenotyping results at evening day 0, morning day 1, evening day 4, morning day 5, evening day 7 and morning day 8 in the multivariate response space. Results of six time point of one representative subject are shown here. Cells are depicted in a blue density gradient. The length of the vector of each marker correlates to the relative importance of that marker for the description of the multivariate space. The correlations between the different markers is defined by the angle between the vectors. Vectors pointing in the same direction are positively correlated; vectors in an angle  $90^\circ$  are not correlated with each other; vectors positioned in an angle of  $180^\circ$  are negatively correlated. The cyan circle depicts the region of 80% of the cells of all the control samples (day 1). The dark blue benchmark in the morning day 1 plot encompasses 80% of the cells of this specific sample. The neutrophils of this control sample fall within the cyan control benchmark. In the graphs of the other time points the red benchmark encompasses 80% of the cells of this individual. The cells shifted in the multivariate space towards CBRM1/5, LAIR-1, CD49d and CD16. Overnight from evening day 4 to morning day 5 the shift was maintained, but part of the cells still overlapped with the control benchmark. A total shift of the cells was seen at morning day 8, in the opposite direction of CD62L and more towards LAIR-1, CBRM1/5 and CD49d. Also a larger spread of the cells was found in the negative direction of CD62L.

### **The exercise induced neutrophil phenotype differs from the neutrophil subsets in LPS challenged volunteers**

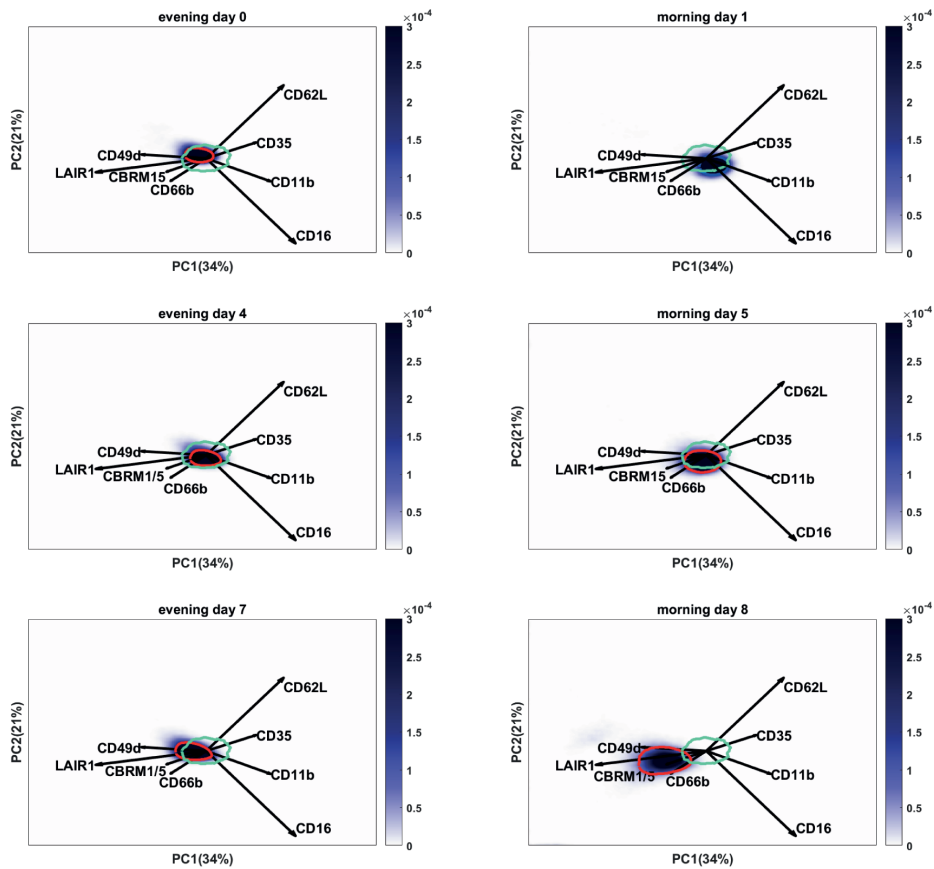
The flow cytometry results show both similarities and differences when compared to peripheral blood neutrophil phenotypes which arise during acute inflammation such as a evoked by LPS challenge <sup>16</sup>. We hypothesized that a long term exercise-induced neutrophil response, possibly driven by reperfusion problems and damage associated molecular patterns (DAMPs), would differ from a short term response to pathogen associated molecular patterns (PAMPs) such as LPS. To test this, we performed a second FLOOD analysis on both the TFL data as well as on data of samples obtained from LPS challenged volunteers at time point t=180 minutes post LPS infusion. For both datasets the same multicolor flow cytometry panel was measured and all samples were processed using the same freeze thaw procedure <sup>29</sup>. In this analysis the TFL and LPS datasets were used to generate the response specific model to ensure that both the exercise response as well as the LPS response were described in the same response specific multi-dimensional space.

Fifty-five percent of the total response specific variance is described in the plots of Figure 7. For the TFL samples a shift is visible mainly towards LAIR-1 and to a lesser extent to CD49d, CBRM1/5 and CD66b (since the loading vectors are shorter). The shift is in the opposite direction of the loadings vectors of CD11b, CD62L and CD35, meaning these markers were downregulated. The shift of the cell population is orthogonally oriented to the loading vector for CD16, which indicates that CD16 did not contribute to the variance of the exercise induced neutrophil response.

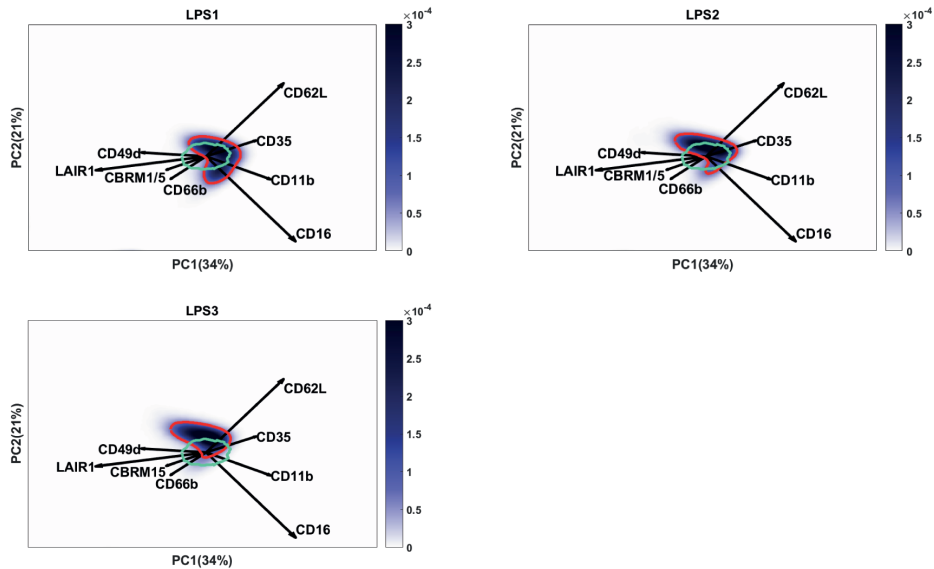
The neutrophil phenotypes found in the peripheral blood 180 min after LPS infusion were clearly different from the exercise induced neutrophil phenotype, when comparing the location of the cells in the FLOOD plots. LPS samples showed the typical bean-shaped neutrophil population representing the different neutrophil subsets. The CD16<sup>dim</sup>CD62L<sup>bright</sup> cells were found in the opposite direction of the CD16 loading; whereas the CD16<sup>bright</sup>CD62L<sup>dim</sup> cells were found in the opposite direction of the CD62L loading and for this subset a higher expression of CD11b was found when compared to the other subsets. Compared to the TFL samples, the total LPS challenged neutrophil population upregulated CD35 and CD11b. Also a downregulation of CD16 and LAIR-1 on the total neutrophil population was found.

The TFL neutrophils at morning day 8 did not overlap with the region where the LPS neutrophils were found, meaning that the exercise-induced peripheral blood neutrophils exhibited a clearly different phenotype from LPS neutrophils.

A



B



**Figure 7. Exercise induced changes of neutrophil marker expressions differ greatly from peripheral blood neutrophil phenotypes found after LPS challenge.**

FLOOD analysis on TFL and LPS datasets. Cells are depicted in a blue density gradient. The cyan circle depicts the region of 80% of the cells of all the control samples (day 1). The dark blue benchmark in the morning day 1 plot encompasses 80% of the cells of this specific sample. The neutrophils of this control sample fall within the cyan control benchmark. In the graphs of the other time points the red benchmark encompasses 80% of the cells of this individual. A) During the TFL at the evening time points a shift towards LAIR-1, CD49d, CBRM1/5 and CD66b was found, negatively correlating with CD62L, CD35 and CD11b. B) The LPS samples showed a heterogeneous cell population with a CD16<sup>dim</sup> subset and a CD62L<sup>dim</sup> subset. A shift of the neutrophil cell population was found in the direction of loading vectors CD16, CD62L, CD35 and CD11b.

**Discussion**

In contrast to the numerous studies<sup>24,34,35</sup> investigating the innate immune response to a single exercise bout, our study examined the recovery of the neutrophil compartment at the morning of day 5 and 8 after daily intensified endurance exercise. We aimed to gain more insight into the mechanisms underlying the increased neutrophil counts in the peripheral blood during a prolonged and increased training load. We found that intensified endurance exercise led to impaired recovery of the innate immune response characterized by differences in activation markers and mobilization of different neutrophil subsets normally absent in the peripheral blood.

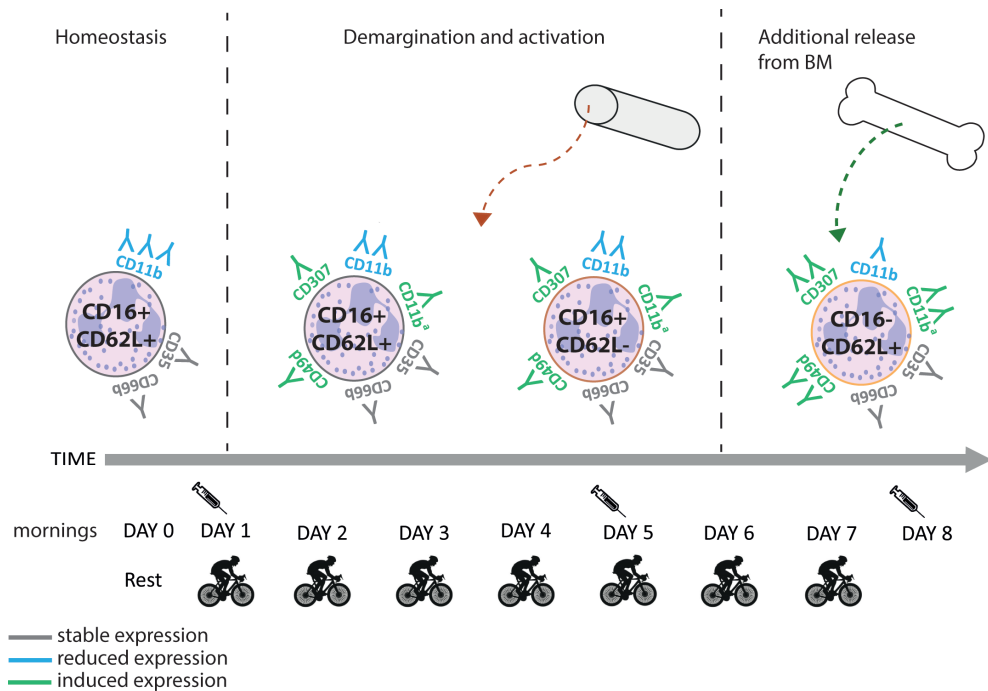
During the week of the TFL, amateur sports men and women increased their training volume on average with  $770 \pm 24\%$ . Exercise induced neutrophilia recovered to baseline at the morning of day 5, but was significantly higher than baseline at the morning of day 8. This neutrophilia could not be explained by increases in serum cortisol and/ or plasma ACTH since these hormone levels were either stable or even decreasing in the mornings of the TFL. Other studies have reported stable resting cortisol levels after a period of intensified training<sup>11–14</sup> or even an increase in cortisol levels.<sup>12</sup> However, these studies enrolled professionally trained athletes instead of amateurs and increased the training load with 2.5 to 6 fold, that was less compared to our study.

Interestingly, in our study the ACTH levels did not relate to the cortisol levels nor to leukocyte numbers. This suggests that cortisol levels were either not controlled by ACTH or half-lives of both hormones were completely different. Similar dissociation of ACTH and cortisol has been described before in inflammatory mouse models.<sup>36,37</sup>

In order to study the characteristics of the neutrophilia and changes in the neutrophil compartment during recovery in more detail we performed multi-color flow cytometry. We applied a field applicable method that has recently been reported by us.<sup>29</sup> This method

employs fixing and freezing and, therefore, the reported MFI's in this article cannot be directly compared to studies employing fresh cells. However, irrespective of this limitation, our protocol allowed the identification of the two extra neutrophil subsets mobilized by LPS challenge (Figure S2A). Also, upregulation of CD11b and CD35 on CD16<sup>bright</sup>CD62L<sup>dim</sup> cells when compared to CD16<sup>dim</sup>CD62L<sup>bright</sup> cells was found similar (Figure S2B and C). This demonstrates the validity of the new protocol to identify subtle changes in phenotypes of white blood cells obtained from blood samples collected during field work. Similarly as found during acute inflammation evoked by LPS, two extra neutrophil subsets were mobilized during intensified endurance exercise. However, the phenotype of these two extra neutrophil populations is clearly different between induction by LPS and exercise. This is illustrated by the fact that no differences in CD11b and CD35 expression were found after 4 and 7 days of cycling between CD16<sup>bright</sup>CD62L<sup>dim</sup> and CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils. Also the overall expression of markers on the complete neutrophil compartment was different as CBRM1/5, CD49d and LAIR-1 were clearly enhanced on all cells irrespective of the expression of CD62L and CD16 (see Figure 8 for a schematic overview of the different PMN phenotypes emerging in the peripheral blood over time). From this we can conclude that DAMPS liberated under conditions of exercise evoke an alternative response when compared to the PAMP LPS. Interestingly, the neutrophilia did not recover overnight to normal values after 7 days of cycling. This was particularly caused by the continued presence of the different neutrophil subsets in the peripheral blood and is in marked contrast to a single LPS challenge after which the blood neutrophil compartment completely normalized overnight.<sup>38</sup> Apparently, repeated intensified endurance exercise during the TFL evoked persistent signals leading to chronic mobilization of the different neutrophil subsets.

Importantly, the CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophil subset has been described as immunosuppressive in an acute inflammation model by inhibiting T-cells.<sup>16</sup> In addition, these cells exhibit a low propensity to contain and or kill bacteria.<sup>21</sup> The occurrence of this subset in the blood and presumably also in the tissues during a period of intensified training might be part of an immune response to mild ischemia reperfusion injury.<sup>39</sup> The increase in neutrophil counts might be necessary for the repair of tissue damage<sup>40</sup> and the specific recruitment of this immunosuppressive neutrophil subset might be important for the inhibition of inflammatory responses to prevent collateral tissue damage, a key feature of reperfusion injury.<sup>41</sup> On the downside, these neutrophils would then respond inadequately to a potential infectious trigger by suppressing the function of other immune cells. This might contribute to the finding that profound exercise is associated with enhanced susceptibility for infections.



**Figure 8. Schematic overview of neutrophil phenotypes emerging in the peripheral blood in the analysed time window.**

The various neutrophil subsets present in the peripheral blood at day 0, day 5 and day 8 are depicted based on their CD16 and CD62L expression. The downregulation of CD11b (in blue) and the upregulation of LAIR1, CD49d and CD11b<sup>a</sup> (CBRM1/5, the activated epitope of CD11b) are visualised (in green) on the neutrophil cell membrane. The expression of CD35 and CD66b remained stable over time (in grey). At day 5 the rise in CD16+CD62L- neutrophils might be caused by demargination from the blood vessel wall, activation or mobilization from outside the vasculature. At day 8 an additional neutrophil subset with a CD16-CD62L+ phenotype was recruited from the bone marrow.

CD62L<sup>dim</sup> neutrophils in the LPS model were characterized by a high expression of CD11b whereas the cells during the TFL showed lowered expressions, indicating important differences in these cell phenotypes between the two situations. In addition to the increase in CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophils, we found both absolute and relative increases in CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils, at the morning of day 8. These cells were likely recruited from the bone marrow, mediated by an accelerated flux of cells through the neutrophil compartment such as proposed for inflammatory diseases.<sup>42</sup> Previous studies have reported increases in banded neutrophils upon endurance exercise.<sup>18–20</sup> The peripheral mature neutrophil pool (including the marginated and bone marrow pool) might become partly exhausted during 7 days of intensified endurance exercise as has also been described during leukocytosis induced by sepsis<sup>43–45</sup>, leading to an increased release of banded cells from the bone marrow. Interestingly, we only found a clear increase in

CD16<sup>dim</sup>CD62L<sup>bright</sup> cells in 13 out of the 28 contestants at day 8. This finding cannot be explained by differences in age, training load or sex. We also found a heterogeneous response in expression of LAIR-1 and CD49d at day 8. This was found on all neutrophils irrespective of expression of CD16 and or CD62L. The unique heterogeneity in response to prolonged exercise may define novel leads to healthy training or even be used to individually monitor athletes in their pre-competitive training.

Next we analyzed MFI's of several markers associated with neutrophil activation. Surprisingly, classical degranulation markers such as CD35 and CD66b did not show significant changes in any of the neutrophil phenotypes over time. The expression of the integrin CD11b was even downregulated over time. These data are in strong contrast to our findings during LPS challenge and clearly imply that neutrophils were not degranulating under these conditions in the peripheral blood. LAIR-1, CD49d, CBRM1/5 were found to be upregulated at day 5 and day 8 during the TFL. LAIR-1 is an immune inhibitory receptor binding to collagens and transducing negative signals into the cellular interior.<sup>30,46,47</sup> Therefore, it is tempting to speculate that the cells belong to a phenotype associated with immune suppression. The observed increase in CD49d and LAIR-1 might also be interpreted as the recruitment of (meta)myelocytes. However, this was unlikely as no distinct CD16<sup>low</sup>CD62L<sup>low</sup>CD66b<sup>low</sup>CD35<sup>low</sup> subset was found during the TFL. Also the more immature CD16<sup>dim</sup>CD62L<sup>bright</sup> cells did not express these markers differently from the CD16<sup>bright</sup>CD62L<sup>bright</sup> cells. In fact, all three neutrophil subtypes similarly changed their expression profile of the aforementioned markers at day 8 of intensive endurance exercise, except from CD49d, which was downregulated in CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils at morning day 8 when compared to CD16<sup>bright</sup>CD62L<sup>bright</sup> cells. However, even on the CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils CD49d expression was increased compared to morning day 1 and day 5. This supports the idea of a non-classical activation phenotype, rather than the rise of an immature neutrophil subset in the blood such as also has been described in sepsis patients.<sup>48</sup> Recently, also a circulating CD49d<sup>bright</sup> pro-angiogenic neutrophil phenotype has been described in mice in response to hypoxia-induced VEGF-A signaling.<sup>33</sup> It is possible that the neutrophils present at day 5-8 after cycling upregulate their CD49d because they are exposed to both DAMPs originating from limited reperfusion as well as hypoxia signals.

The increased metabolic demand and anaerobic respiration caused by exercise lead to an increase in oxidative stress.<sup>49</sup> Generally this would result into the recruitment of activated neutrophils, potentially causing tissue damage such as seen in ischemia-reperfusion injury.<sup>41</sup> However, there were no clear indicators of reperfusion-associated damage during the prolonged period of intensified exercise during the TFL. Therefore, the increase of the CD49d<sup>bright</sup>LAIR1<sup>bright</sup>CBRM1/5<sup>bright</sup>CD11b<sup>dim</sup> neutrophil subset is suggestive for an immunosuppressive neutrophil phenotype. It would be interesting



to study the presence of such cells in patients who experience reperfusion injury such as found during coronary bypass surgeries, during ischemic stroke or in claudication patients.<sup>50,51</sup> In animal studies it has been proven that targeting the CD11b/CD18 (Mac-1) receptor on neutrophils is protective against reperfusion injury.<sup>52,53</sup> However, clinical trials in humans have been disappointing.<sup>54,55</sup> Future functional experiments focusing on the characteristics of exercise-induced neutrophils might tell us more about the pathways involved in neutrophil immunosuppression.

Besides this conventional flow cytometry analysis to identify cell populations and analyze MFI's of several markers associated with neutrophil subsets and neutrophil activation, we also analyzed the data using the unsupervised multi-dimensional method FLOOD<sup>25</sup> to perform deep phenotyping of the data set. One of the benefits of this analysis is the 2D representation of the multiple correlations between all markers. This analysis is both more robust (no subjective gating) and much faster than the currently available multi-gating strategies. The FLOOD analysis for the TFL data clearly visualized an exercise-induced shift. The exercise-induced upregulation of CD49d, CBRM1/5 and LAIR-1 did not recover overnight. In fact, the shift was maintained at morning day 5 and even more pronounced at morning day 8 than directly post-exercise. From this we can conclude that the total neutrophil population did not normalize at day 5 and 8 after exercise.

Next to this, we compared the exercise induced neutrophil phenotype to the neutrophil subsets that arise upon LPS administration by analyzing both datasets together using the FLOOD method. From this analysis we can conclude an important finding: the exercise-induced activation of neutrophils differs greatly from the LPS (PAMP)-induced activation of neutrophils in the peripheral blood. The position of the TFL neutrophils at day 8 and the position of the LPS challenged neutrophils did not overlap with each other in the FLOOD response specific space, meaning none of these cells were the same.

This study is descriptive, but it provides us with new insights into the role of neutrophils during the changed innate immune response in response to periods of increased intensive endurance training. These changes might affect the normal immune response towards invading micro-organisms, which has been suggested by several studies (as reviewed in<sup>56,57</sup>). Regular physical exercise is correlated with fewer upper respiratory tract infections (URTI) when compared to a sedentary lifestyle.<sup>58,59</sup> On the other hand, periods of increased intensive endurance training in athletes are associated with an increased risk of URTI by yet poorly understood immunological mechanisms.<sup>60</sup> A single low-intensity exercise bout induces a biphasic mobilization of CD8+ T-cells and NK-cells. First the peripheral blood counts increase followed by a decrease resulting in lymphocytopenia.<sup>61-63</sup> High-intensity exercise has been found to lead to a diminished increase as well as a decrease of CD8+ T-cell counts,<sup>64</sup> which is hypothesized to create an "open window"<sup>60</sup> for pathogens like viruses to infect the host. Regardless, a systemic neutrophil response has been reported to

precede the peak in CD8+ T-cell activation during a viral lung infection by 2-8 days<sup>65</sup> and might also be implicated in URTI after intensified exercise. These latter studies support the hypothesis of an important role of neutrophils in antiviral responses and could explain the increased risk of URTI during periods of intensified training in athletes. Our study is the first to show the maintained changes in neutrophil phenotypes upon intensified training even after overnight recovery. These changes may result in a more chronically impaired innate immune system, at least during periods of intensified training, supporting the “open window” hypothesis.<sup>60</sup>

In conclusion, at day 5 and 8 after intensified endurance exercise, the neutrophil compartment did not recover overnight. In fact, the aberrant neutrophil subtypes increased during the TFL. These neutrophils did not upregulate classical activation markers arguing against a degranulated phenotype. The observed changes in neutrophil phenotypes are in line with the hypothesis of immune suppression by neutrophils and could explain the increased risk of URTI during periods of intensified training in athletes. Our findings therefore provide a strong basis to design functional experiments to test this hypothesis.

## Acknowledgements

This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area COAST.

We are very grateful to all participants who voluntarily took part in all measurements. Also we would like to thank Tour for Life organizer Emolife, especially Nienke Wynia and Kelian Lankester, for their collaboration.

We thank Julia Drylewicz for her help with statistics and André van den Doel for his early work for the development of the FLOOD analysis method. Not the least thanks to Danilo Iannetta, Erik Oudenhoven and Hans de Koning for their enthusiastic help with sample collection.

## References

1. Robson PJ, Blannin a K, Walsh NP, Castell LM, Gleeson M. Effects of exercise intensity, duration and recovery on in vitro neutrophil function in male athletes. *Int J Sports Med.* 1999;20(2):128-135. doi:10.1055/s-2007-971106
2. Vejls G. The distribution of leukocytes in the vascular system. *Acta Pathol Microbiol Scand.* 1938;Suppl 33:1-239.
3. Athens JW, Raab SO, Haab OP, et al. Leukokinetic studies. III. The distribution of granulocytes in the blood of normal subjects. *J Clin Invest.* 1960;40:159-164. doi:10.1172/JCI104230
4. Athens JW, Haab OP, Raab SO, et al. Leukokinetic studies. IV. The total blood, circulating and marginal granulocyte pools and the granulocyte turnover rate in normal subjects. *J Clin Invest.* 1961;40(2):989-995. doi:10.1172/JCI104338
5. French EB, Steel CM, Aitchison WRC. Studies on Adrenaline-induced Leucocytosis in Normal Man. *Br J Haematol.* 1971;21(4):423-431. doi:10.1111/j.1365-2141.1971.tb02702.x
6. Joyce RA, Boggs DR, Hasiba U, Srodes CH. Marginal neutrophil pool size in normal subjects and neutropenic patients as measured by epinephrine infusion. *J Lab Clin Med.* 1976;88(4):614-620.
7. Muir AL, Cruz M, Martin BA, Thommasen H, Belzberg A, Hogg JC. Leukocyte kinetics in the human lung: role of exercise and catecholamines. *J Appl Physiol.* 1984;57(3):711-719. doi:10.1152/jappl.1984.57.3.711
8. Farrell P a, Garthwaite TL, Gustafson a B. Plasma adrenocorticotropin and cortisol responses to submaximal and exhaustive exercise. *J Appl Physiol.* 1983;55(5):1441-1444.
9. Kraemer WJ, Fleck SJ, Callister R, et al. Training responses of plasma beta-endorphin, adrenocorticotropin, and cortisol. *MedSciSportExerc.* 1989;21(2):146-153. doi:10.1249/00005768-198904000-00006
10. Dale DC, Fauci AS, Guerry IV D, Wolff SM. Comparison of agents producing a neutrophilic leukocytosis in man. Hydrocortisone, prednisone, endotoxin, and etiocholanolone. *J Clin Invest.* 1975;56(4):808-813. doi:10.1172/JCI108159
11. Urhausen A, Gabriel HHW, Kindermann W. Impaired pituitary hormonal response to exhaustive exercise in overtrained endurance athletes. *Med Sci Sport Exerc.* 1998;30(3):407-414. doi:10.1097/00005768-199803000-00011
12. Svendsen IS, Killer SC, Carter JM, Randell RK, Jeukendrup AE, Gleeson M. Impact of intensified training and carbohydrate supplementation on immunity and markers of overreaching in highly trained cyclists. *Eur J Appl Physiol.* 2016;116:867-877. doi:DOI 10.1007/s00421-016-3340-z
13. Hooper SL, Mackinnon LT, Gordon RD, Bachmann AW. Hormonal responses of elite swimmers to overtraining. *Med Sci Sport Exerc.* 1993;25(6):741-747. doi:10.1249/00005768-199306000-00015
14. Mackinnon LT, Hooper SL, Jones S, Gordon RD, Bachmann AW. Hormonal, immunological, and hematological responses to intensified training in elite swimmers. *Med Sci Sport Exerc.* 1997;29(12):1637-1645. doi:10.1097/00005768-199712000-00014

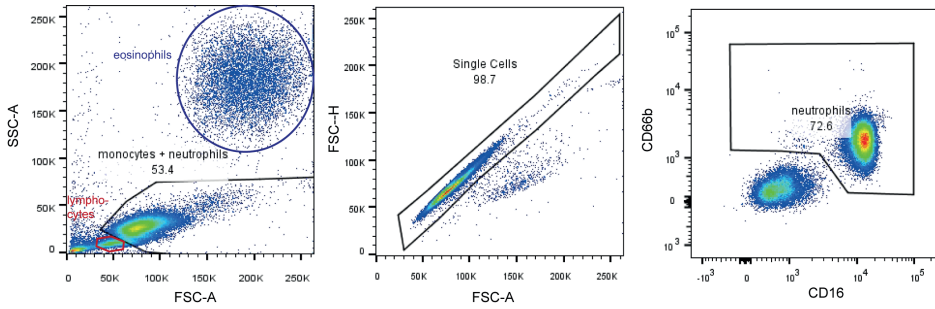
15. Pillay J, Tak T, Kamp VM, Koenderman L. Immune suppression by neutrophils and granulocytic myeloid-derived suppressor cells: Similarities and differences. *Cell Mol Life Sci*. 2013;70(20):3813-3827. doi:10.1007/s00018-013-1286-4
16. Pillay J, Kamp VM, van Hoffen E, et al. A subset of neutrophils in human systemic inflammation inhibits T cell responses through Mac-1. *J Clin Invest*. 2012;122(1):327-336. doi:10.1172/JCI57990
17. Tak T, Wijten P, Heeres M, et al. Human CD62Ldim neutrophils identified as a separate subset by proteome profiling and in vivo pulse-chase labeling. *Blood*. 2017;129(26). doi:10.1182/blood-2016-07-727669
18. Suzuki K, Naganuma S, Totsuka M, et al. Effects of exhaustive endurance exercise and its one-week daily repetition on neutrophil count and functional status in untrained men. *Int J Sports Med*. 1996. doi:10.1055/s-2007-972833
19. Suzuki K, Totsuka M, Nakaji S, et al. Endurance exercise causes interaction among stress hormones, cytokines, neutrophil dynamics, and muscle damage. *J Appl Physiol*. 1999;87(4):1360-1367. doi:10.1001/jama.1994.03510270072043
20. Suzuki K, Nakaji S, Yamada M, et al. Impact of a competitive marathon race on systemic cytokine and neutrophil responses. *Med Sci Sports Exerc*. 2003;35(2):348-355. doi:10.1249/01.MSS.0000048861.57899.04
21. Leliefeld PHC, Pillay J, Vriskoop N, et al. Differential antibacterial control by neutrophil subsets. *Blood Adv*. 2018;2(11):1344-1355. doi:10.1182/BLOODADVANCES.2017015578
22. Peake J, Wilson G, Hordern M, et al. Changes in neutrophil surface receptor expression, degranulation, and respiratory burst activity after moderate- and high-intensity exercise. *J Appl Physiol*. 2004;97(2):612-618. doi:10.1152/japplphysiol.01331.2003
23. Peake JM, Suzuki K, Wilson G, et al. Exercise-induced muscle damage, plasma cytokines, and markers of neutrophil activation. *Med Sci Sports Exerc*. 2005;37(5):737-745. doi:10.1249/01.MSS.0000161804.05399.3B
24. Wang J-S, Chiu Y-T. Systemic hypoxia enhances exercise-mediated bactericidal and subsequent apoptotic responses in human neutrophils. *J Appl Physiol*. 2009;107(4):1213-1222. doi:10.1152/japplphysiol.00316.2009
25. Jansen JJ, & Hilvering B, van den Doel A, et al. FLOOD: FLOW cytometric Orthogonal Orientation for Diagnosis. *Chemom Intell Lab Syst*. 2016;151(December):126-135. doi:10.1016/j.chemolab.2015.12.001
26. Pauw K De, Roelands B, Geus B De, Meeusen R. Guidelines to classify subject groups in sport-science research . Guidelines to Classify Subject Groups in Sport-Science Research. *Int J Sports Physiol Perform*. 2013;8(March 2013):111-122. doi:10.1123/ijspp.8.2.111
27. Decroix L, De Pauw K, Foster C, Meeusen R. Guidelines to classify female subject groups in sport-science research. *Int J Sports Physiol Perform*. 2016;11(2):204-213. doi:10.1123/ijspp.2015-0153
28. Kox M, van Eijk LT, Zwaag J, et al. Voluntary activation of the sympathetic nervous system and attenuation of the innate immune response in humans. *Proc Natl Acad Sci*. 2014;111(20):7379-7384. doi:10.1073/pnas.1322174111

29. de Ruiter K, van Staveren S, Hilvering B, et al. A field-applicable method for flow cytometric analysis of granulocyte activation: Cryopreservation of fixed granulocytes. *Cytom Part A*. 2018. doi:10.1002/cyto.a.23354
30. Verbrugge A, Ruiter T de, Geest C, Coffe PJ, Meyaard L. Differential expression of leukocyte-associated Ig-like receptor-1 during neutrophil differentiation and activation. *J Leukoc Biol*. 2006;79(4):828-836. doi:10.1189/jlb.0705370
31. Lund-Johansen F, Terstappen LW. Differential surface expression of cell adhesion molecules during granulocyte maturation. *J Leukoc Biol*. 1993;54(1):47-55. doi:10.1002/jlb.54.1.47
32. Liesveld JL, Winslow JM, Frediani KE, Ryan DH, Abboud CN. Expression of integrins and examination of their adhesive function in normal and leukemic hematopoietic cells. *Blood*. 1993;81(1):112-121.
33. Massena S, Christoffersson G, Evelina V, et al. Identification and characterization of VEGF-A-responsive neutrophils expressing CD49d, VEGFR1, and CXCR4 in mice and humans. *Blood*. 2015;126(17):2016-2027. doi:10.1182/blood-2015-03-631572.
34. Peake J. Changes in neutrophil surface receptor expression, degranulation, and respiratory burst activity after moderate- and high-intensity exercise. *J Appl Physiol*. 2004;97(2):612-618. doi:10.1152/jappphysiol.01331.2003
35. Gavrieli R, Ashlagi-Amiri T, Eliakim A, et al. The effect of aerobic exercise on neutrophil functions. *Med Sci Sports Exerc*. 2008;40(9):1623-1628. doi:10.1249/MSS.0b013e318176b963
36. Franchimont D, Bouma G, Galon J, et al. Adrenal cortical activation in murine colitis. *Gastroenterology*. 2000;119(6):1560-1568. doi:10.1053/gast.2000.20235
37. Silverman MN, Miller AH, Biron CA, Pearce BD. Characterization of an Interleukin-6- and Adrenocorticotropin-Dependent, Immune-to-Adrenal Pathway during Viral Infection. *Endocrinology*. 2004;145(8):3580-3589. doi:10.1210/en.2003-1421
38. Kamp VM, Leentjens J, Pillay J, et al. Modulation of granulocyte kinetics by GM-CSF/IFN- $\gamma$  in a human LPS rechallenge model. *J Leukoc Biol*. 2013;94(3):513-520. doi:10.1189/jlb.0213066
39. Carden DL, Granger DN. Pathophysiology of ischaemia-reperfusion injury. *J Pathol*. 2000;190(3):255-266. doi:10.1002/(SICI)1096-9896(200002)190:3<255::AID-PATH526>3.0.CO;2-6
40. Bastian OW, Koenderman L, Alblas J, Leenen LPH, Blokhuis TJ. Neutrophils contribute to fracture healing by synthesizing fibronectin+ extracellular matrix rapidly after injury. *Clin Immunol*. 2016;164. doi:10.1016/j.clim.2016.02.001
41. Schofield ZV, Woodruff TM, Halai R, Wu MCL, Cooper MA. Neutrophils - A key component of ischemia-reperfusion injury. *Shock*. 2013;40(6):463-470. doi:10.1097/SHK.0000000000000044
42. Smith JA. Neutrophils, host defense, and inflammation: a double-edged sword. *J Leukoc Biol*. 1994;56(6):672-686. doi:10.1002/jlb.56.6.672
43. Bone RC, Balk RA, Cerra FB, et al. Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis. *Chest*. 1992;101(6):1644-1655. doi:10.1378/chest.101.6.1644

44. Orr Y, Taylor JM, Bannon PG, Geczy C, Kritharides L. Circulating CD10-/CD16low neutrophils provide a quantitative index of active bone marrow neutrophil release. *Br J Haematol*. 2005;131(4):508-519. doi:10.1111/j.1365-2141.2005.05794.x
45. Taneja R, Sharma AP, Hallett MB, Findlay GP, Morris MR. Immature Circulating Neutrophils in Sepsis Have Impaired Phagocytosis and Calcium Signaling. *Shock*. 2008;30(6):618-622. doi:10.1097/SHK.0b013e318173ef9c
46. Lebbink RJ, Ruiter T de, Adelmeijer J, et al. Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. *J Exp Med*. 2006;203(6):1419-1425. doi:10.1084/jem.20052554
47. Meyaard L. The inhibitory collagen receptor LAIR-1 (CD305). *J Leukoc Biol*. 2008;83(4):799-803. doi:10.1189/jlb.0907609
48. Ibbotson GC, Doig C, Kaur J, et al. Functional  $\alpha 4$ -integrin: A newly identified pathway of neutrophil recruitment in critically ill septic patients. *Nat Med*. 2001;7(4):465-470. doi:10.1038/86539
49. Pingitore A, Lima GPP, Mastorci F, Quinones A, Iervasi G, Vassalle C. Exercise and oxidative stress: Potential effects of antioxidant dietary strategies in sports. *Nutrition*. 2015;31(7-8):916-922. doi:10.1016/j.nut.2015.02.005
50. Turton EPL, Coughlin PA, Kester RC, Scott DJA. Exercise training reduces the acute inflammatory response associated with claudication. *Eur J Vasc Endovasc Surg*. 2002;23(4):309-316. doi:10.1053/ejvs.2002.1599
51. Jin R, Yang G, Li G. Inflammatory mechanisms in ischemic stroke: role of inflammatory cells. *J Leukoc Biol*. 2010;87(5):779-789. doi:10.1189/jlb.1109766
52. Zhang ZG, Chopp M, Tang WX, Jiang N, Zhang RL. Postischemic treatment (2-4 h) with anti-CD11b and anti-CD18 monoclonal antibodies are neuroprotective after transient (2 h) focal cerebral ischemia in the rat. *Brain Res*. 1995;698(1-2):79-85. doi:10.1016/0006-8993(95)00830-J
53. Tajra LCF, Martin X, Margonari J, et al. In Vivo Effects of Monoclonal Antibodies against Rat  $\beta 2$  Integrins on Kidney Ischemia-Reperfusion Injury. *J Surg Res*. 1999;87(1):32-38. doi:10.1006/jsre.1999.5724
54. Faxon DP, Gibbons RJ, Chronos NAF, Gurbel PA, Sheehan F. The effect of blockade of the CD11/CD18 integrin receptor on infarct size in patients with acute myocardial infarction treated with direct angioplasty: The results of the HALT-MI study. *J Am Coll Cardiol*. 2002;40(7):1199-1204. doi:10.1016/S0735-1097(02)02136-8
55. Rusnak JM, Kopecky SL, Clements IP, et al. An anti-CD11/CD18 monoclonal antibody in patients with acute myocardial infarction having percutaneous transluminal coronary angioplasty (the FESTIVAL study). *Am J Cardiol*. 2001;88(5):482-487. doi:10.1016/S0002-9149(01)01723-4
56. Pedersen BK, Rohde T, Ostrowski K. Recovery of the immune system after exercise. *Acta Physiol Scand*. 1998;162(3):325-332. doi:10.1046/j.1365-201X.1998.0325e.x
57. Peake JM, Neubauer O, Walsh NP, Simpson RJ. Recovery of the immune system after exercise. *J Appl Physiol*. December 2016:jap.00622.2016. doi:10.1152/japphysiol.00622.2016

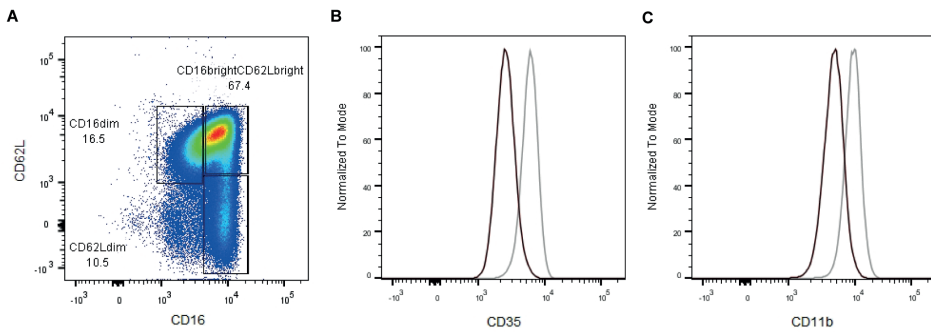
58. Matthews CE, Ockene IR a S, Freedson PS, Rosal MC, Merriam P a, Hebert JR. Moderate to vigorous physical activity and risk of upper-respiratory tract infection. *Med Sci Sport Exerc.* 2002;34(8):1242-1248. doi:10.1097/00005768-200208000-00003
59. Nieman DC, Henson D a, Austin MD, Sha W. Upper respiratory tract infection is reduced in physically fit and active adults. *Br J Sports Med.* 2011;45(12):987-992. doi:10.1136/bjsm.2010.077875
60. Nieman DC. Special feature for the Olympics: effects of exercise on the immune system: exercise effects on systemic immunity. *Immunol Cell Biol.* 2000;78(5):496-501. doi:10.1111/j.1440-1711.2000.t01-5-x
61. LaVoy EC, Bosch JA, Lowder TW, Simpson RJ. Acute aerobic exercise in humans increases cytokine expression in CD27– but not CD27+ CD8+ T-cells. *Brain Behav Immun.* 2013;27:54-62. doi:10.1016/j.bbi.2012.09.006
62. Spielmann G, Bollard CM, Bigley AB, et al. The effects of age and latent cytomegalovirus infection on the redeployment of CD8+ T cell subsets in response to acute exercise in humans. *Brain Behav Immun.* 2014;39:142-151. doi:10.1016/j.bbi.2013.05.003
63. Simpson RJ, Bigley AB, Spielmann G, LaVoy ECP, Kunz H, Bollard CM. Human cytomegalovirus infection and the immune response to exercise. *Exerc Immunol Rev.* 2016;22:8-27.
64. Witard OC, Turner JE, Jackman SR, et al. High-intensity training reduces CD8+ T-cell redistribution in response to exercise. *Med Sci Sports Exerc.* 2012;44(9):1689-1697. doi:10.1249/MSS.0b013e318257d2db
65. Lukens M V, van de Pol AC, Coenjaerts FEJ, et al. A systemic neutrophil response precedes robust CD8(+) T-cell activation during natural respiratory syncytial virus infection in infants. *J Virol.* 2010;84(5):2374-2383. doi:10.1128/JVI.01807-09

## Supplementary Material



**Figure S1. Neutrophil gating strategy.**

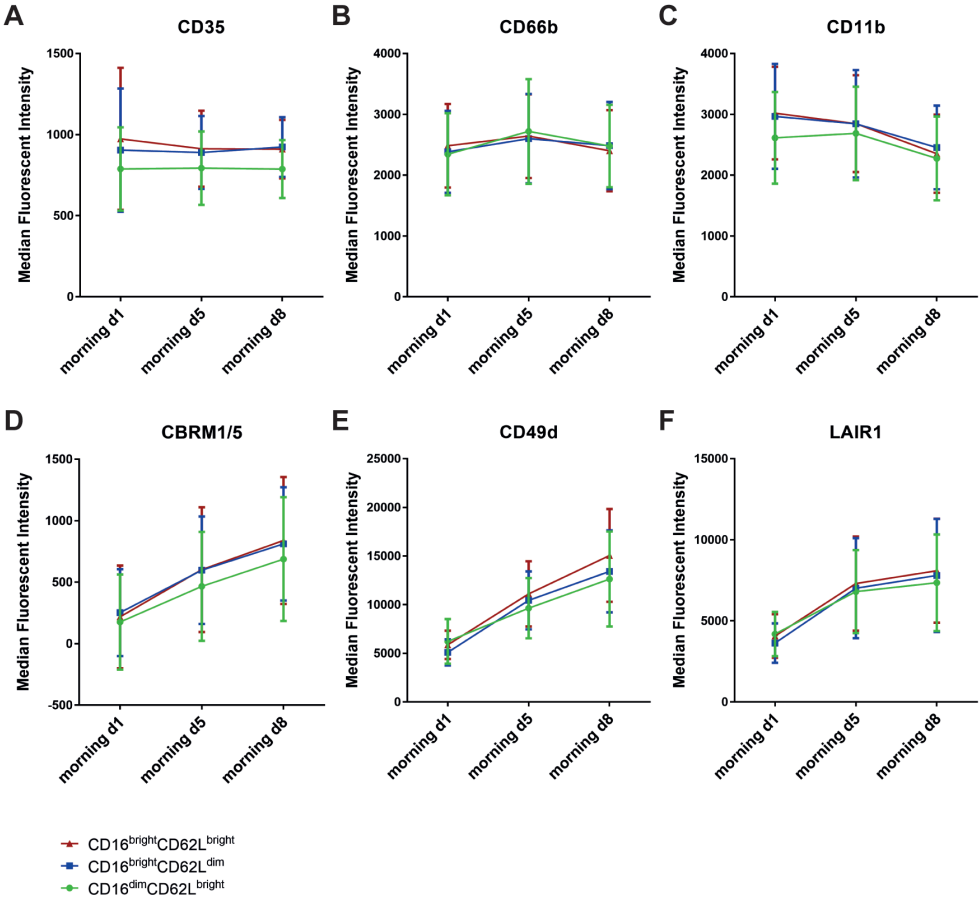
Neutrophils and monocytes were gated based on FSC and SSC, excluding eosinophils (large FSC/SSC), lymphocytes and debris. Single cells were gated. Monocytes were excluded from the neutrophil gate by gating the CD66b+ cells.



**Figure S2. The neutrophil freezing protocol is sufficient for identification of neutrophil subsets based on CD16/CD62L expression.**

A) FACS plot of CD16 vs. CD62L expression of the total neutrophil population during acute inflammation (LPS model). Cells were frozen and thawed before staining the neutrophils. B) histogram of CD35 and CD11b expression of CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils (black) and CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophils (gray) during acute inflammation C) histogram of CD11b expression of CD16<sup>dim</sup>CD62L<sup>bright</sup> neutrophils (black) and CD16<sup>bright</sup>CD62L<sup>dim</sup> neutrophils (gray) during acute inflammation.





**Figure S3. Expressions of CD35, CD66b, CD11b, CBRM1/5, CD49s and LAIR-1 on the different CD16/CD62L neutrophil subsets.**

The median fluorescent intensities  $\pm$  SD of A) CD35, B) CD66b, C) CD11b, D) CBRM1/5, E) CD49d and F) LAIR-1 are shown for the different neutrophil subsets at morning day 1, morning day 5 and morning day 8.

\*  $p < 0.05$ .

**CHAPTER 4**



# Transformation of flow cytometry data with OTflow:

## A necessity for multivariate analysis

Selma van Staveren<sup>1,2\*</sup>, Rita Folcarelli<sup>3\*</sup>, Gerjen Tinnevelt<sup>2,3</sup>, Emily Cadot<sup>3</sup>, Nienke Vrisekoop<sup>1</sup>, Lutgarde Buydens<sup>3</sup>, Leo Koenderman<sup>1</sup>, Jeroen Jansen<sup>3\*</sup>, Oscar van den Brink<sup>2\*</sup>

1 Department of Respiratory Medicine, Laboratory of Translational Immunology, UMC Utrecht, The Netherlands

2 TI-COAST, Amsterdam, The Netherlands

3 Analytical Chemistry, Institute for Molecules and Materials, Radboud University Nijmegen, The Netherlands

\*authors contributed equally to this work

**Introduction:** The flow cytometry field has evolved rapidly, allowing the measurement of 30-50 parameters per cell. This has led to a tremendous increase in multivariate information. Manual gating is insufficient to extract all this information. Therefore, multivariate analysis (MVA) methods have been developed to extract information and efficiently analyse the high-density multicolour flow cytometry (MFC) data. To aid interpretation, MFC data are often logarithmically transformed before MVA.

**Rationale:** We studied the consequences of different transformations of flow cytometry data in datasets containing negative intensities caused by background subtractions and spreading error, as logarithmic transformation of negative data is mathematically impossible. Alternative transformations such as logicle or hyperbolic arcsine transformations allow linearity around zero, whereas higher (positive and negative) intensities are logarithmically transformed.

**Results:** To define the linear range, a parameter (or cofactor) must be chosen. We show how the chosen transformation parameter has great impact on the MVA results. In some cases, peak-splitting is observed, producing two distributions in an actual homogeneous population around 0. This may be misinterpreted as the presence of multiple cell populations. Moreover, when performing arbitrary transformation before MVA analysis biologically relevant and statistically significant information might be missed. We present a new algorithm, Optimal Transformation for flow cytometry data (OTflow), which uses various statistical methods to optimally choose the parameter of the transformation and prevent artifacts such as peak splitting.

**Conclusion:** Arbitrary or unconsidered transformation can lead to wrong conclusions for the MVA cluster methods, dimensionality reduction methods and classification methods. We recommend transformation of flow cytometry data by using OTflow-defined parameters estimated per channel, in order to prevent peak-splitting and other artifacts in the data.

## Introduction

The flow cytometry field has evolved rapidly. Nowadays commonly used flow cytometers measure up to 17 parameters per cell<sup>1</sup> and the most recently developed flow cytometers can potentially measure 30 to 50 parameters.<sup>2</sup> With this increase in measured variables per single cell, the information in the resulting datasets also increases tremendously.

Analysis of flow cytometry datasets by manual gating has essential drawbacks due to the lack of reproducibility, the subjectivity and bias, the inefficiency for large panels and the fact that it is time-consuming for large datasets.<sup>3</sup> To overcome these drawbacks various multivariate analysis (MVA) methods have been developed for automated analysis of multicolour flow cytometry (MFC) data in the past decade. Methods which use dimensionality reduction like viSNE,<sup>4</sup> FLOOD,<sup>5</sup> DAMACY,<sup>6</sup> and ECLIPSE<sup>7</sup> and clustering methods like SPADE<sup>8</sup>, FlowSOM<sup>9</sup> and Citrus<sup>10</sup> are frequently used by researchers to identify and characterize new (disease-related) cell subsets in a fast and unbiased way. Before one can use these methods, the MFC data should be adequately pre-processed. There are many pre-processing options, and the pre-processing steps can have considerable effects on the results of the analysis. For an overview of optional pre-processing steps, we refer to the review of Saeys et al.<sup>3</sup>

Flow cytometry datasets contain a fair amount of negative intensities. Negative fluorescence intensities are physically and biologically meaningless. However, negative values in MFC data can occur in the process of data acquisition due to adjustment of original measurements by the instrument hardware and software. Negative values might emerge from background subtractions performed by the flow cytometry hardware. PMT signals can contain high levels of background signal from fluorescent light from unbound fluorophores, PMT dark current and ambient light.<sup>11</sup> The background signal is measured in between 2 events and this signal intensity is subtracted from the next event measured. In the case of a negative or dim cell, the measured fluorescence can be smaller than the background signal, due to the measurement errors. This then leads to negative values. In BD FACS Diva software, the user of the flow cytometer cannot influence the background subtraction. Additionally, compensation of spectral overlap among fluorochromes can lead to fluorescent intensities below 0.<sup>11,12</sup> Another source of negative values is caused by data spreading. At a somewhat deeper level, statistical variation originates from measurement errors such as photon counting errors and binning errors.<sup>13</sup> Photon counting error is associated with the quantum mechanical nature of the emitted light. The photons that are emitted by the fluorophore, will arrive at the detector at different time points, in a Poisson-distributed manner, as it is a stochastic process. This causes the signal to be heteroscedastic, having higher variance when fluorescence intensity is higher, and

leads to a non-linear spreading of properly compensated MFC data. This causes cells that are very high for marker B and negative for marker A to become more negative.

Up until recently, when analysing marker expressions from MFC data using conventional gating analysis software, logarithmically transformed axes were used. Log transformations perform a non-linear conversion of the output of the analog-to-digital converter. It is useful to correct for heteroscedasticity and to change skewed distributions into more symmetric, Gaussian distributed peaks.<sup>14</sup> In MFC, logarithmic transformation has been very useful in coping with the wide dynamic range of emissions between fluorophores. The log scale provides an informative and proper display of populations in the higher intensity range as well as of populations in the lower intensity regions. Populations with low fluorescent intensities which may be hardly discerned on a linear scale are well visible and separated on a logarithmic scale. Negative intensities cannot be properly displayed on a logarithmic scale, since calculating the logarithm of a negative value is mathematically impossible. Attempts of representing the data on a logarithmic scale lead to an apparent mean that is too high, and negative data points that are squeezed onto the axes. Therefore, alternative transformation methods have been developed. The logicle<sup>15</sup> and hyperbolic arcsine (arcsinh)<sup>16</sup> transformations are the most commonly used in modern flow cytometry. Both transformation methods use a combination of linear transformation for values close to zero and a logarithmic scale for larger (negative and positive) values. The transition of the linear to the logarithmic part is smoothed out. Importantly, next to visualisation of MFC data in bi-plots, logarithmic or bi-exponential transformations are performed as a pre-processing step for the previously mentioned MVA methods. In the viSNE and SPADE algorithms hyperbolic arcsine transformation is used with a standard parameter (cofactor set at 150) for the entire dataset,<sup>17</sup> despite various reports which have shown the importance of data-driven transformation of MFC data.<sup>16,18</sup>

Here we present an algorithm for Optimal Transformation of flow cytometry data, called OTflow. OTflow is an automated and validated algorithm for optimization of transformation parameters for both visualisation and MVA of flow cytometry data. The algorithm combines properties of normality of the signal and stabilization of the variances among the peaks to best represent MFC data on a bi-exponential scale. Improving the flowVS method by Azad et al<sup>18</sup> we combined variance stabilization<sup>19</sup> together with Bartlett's test<sup>20</sup> to define the optimal parameter for transformation based on the flow cytometry dataset itself. Variance stabilization dissociates the existing correlation between the mean intensity of a cell population and the variance, which is typical for MFC measurements. Bartlett's test<sup>20</sup> is used as statistics to select the arcsinh transformation cofactors which leads to homogenous variances per fluorescent channel measured. As a result, cell populations with different intensities are better discerned and the corresponding peaks resemble more normal distributions. This enables comparison of phenotypical similar cell

populations across multiple samples.<sup>21</sup> However, variance stabilization by flowVS can be performed only when two or more peaks are present per channel. In the case of a single peak per channel, the flowVS algorithm may not be suitable to find optimal cofactor transformation. In the novel algorithm OTflow, in addition to Bartlett's statistics for multiple peaks, we integrated the possibility to estimate optimal parameters for transformations also when a single peak per channel is present. In this case, Jarque-Bera statistics test<sup>22</sup> for normality of the peak is applied. Moreover, OTflow specifically prevents peak-splitting in its parameter optimization process. This is very important in datasets which contain negative values, which are increasingly common when samples are measured on modern flow cytometers. As far as our knowledge goes, no other transformation method for MFC data takes this last feature into account. In this paper we show the versatility of OTflow by applying the algorithm to complementary datasets and performing PCA,<sup>23</sup> viSNE,<sup>4</sup> flowSOM<sup>9</sup> and Citrus<sup>10</sup> analyses. Additionally, we compare these MVA results to the results from the same datasets transformed using hyperbolic arcsine with the default cofactor 150, or using the recently published flowVS algorithm.<sup>18</sup> Also, in the supplementary material, we compare the MVA results after logicle transformation with the optimized W parameters from OTflow to those obtained after logicle transformation with a calculated W value as proposed by Parks et al.<sup>15</sup> We demonstrate that suboptimal transformations by a poorly chosen or calculated parameter can lead to misleading MVA results that may subsequently invoke incorrect immunological conclusions.

## Methods

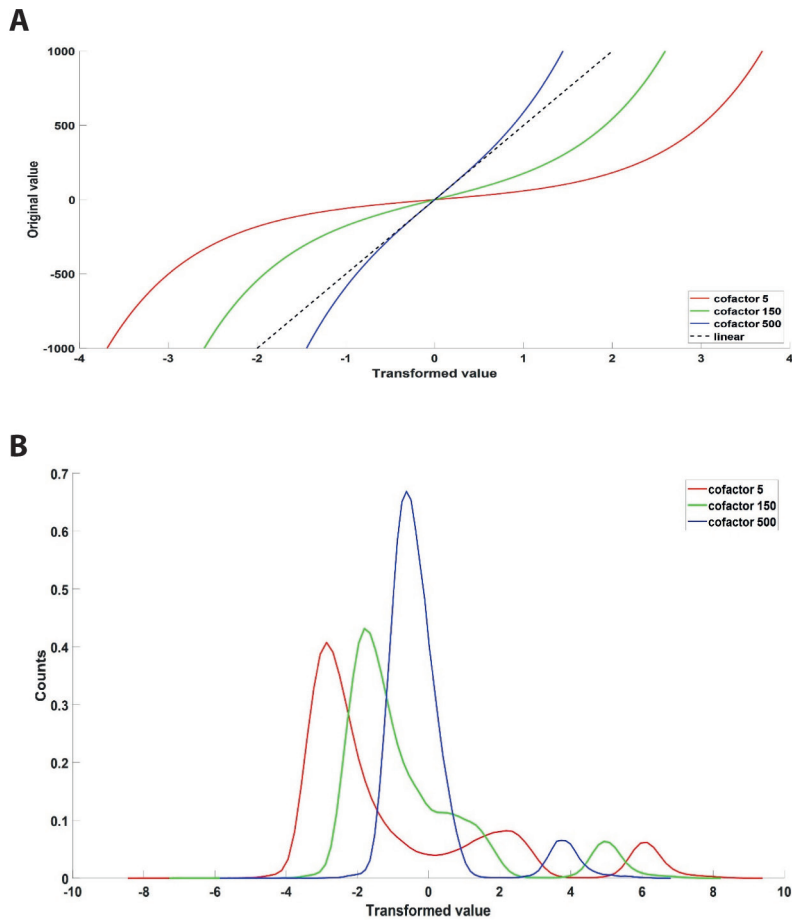
### Transformations

We considered two of the most widespread transformations used to process flow cytometry data: inverse hyperbolic sine<sup>16</sup> and logicle function.<sup>15</sup> They both belong to the class of bi-exponential functions and have the characteristics of being linear and symmetric near zero and becoming exponential for higher values, with a smooth transition between the linear and exponential regions.<sup>24</sup> Note that both transformations deal with negative numbers.

### Inverse hyperbolic sine

The inverse hyperbolic sine (arcsinh) function is defined as follows:

$$\operatorname{arcsinh}(x) = \log \left( \frac{x}{c} + \sqrt{\left( \left( \frac{x}{c} \right)^2 + 1 \right)} \right) \quad \text{Eq. 1}$$



**Figure 1. Arcsinh transformed data using different values for the cofactor.**

A) Original values vs Transformed values plot. Arcsinh transforms characterized by 3 different cofactors (5, 150 and 500) are displayed together with linear transform (dashed line). The higher the cofactor is, the more predominant the linear region around zero is; B) Cell counts vs transformed value plot. Histogram representation of data transformed with arcsinh functions with different cofactors (5, 150 and 500). Low cofactors cause splitting of the negative populations around zero (red and green line). As the cofactor increases the splitting is removed and the variances become smaller.

The  $c (c > 0)$  term is a scaling factor generally called cofactor which modulates the strength and the range of the linearization around zero. Higher values of the cofactor will enlarge the linear range. The linear range should be big enough to properly display signals with low means and high standard deviation. However, too high cofactors will result in too strong linear transformations. A cofactor of 150 has been promoted for flow cytometry data, while a value of 5 is suggested for mass cytometry experiments.<sup>25</sup> These two values



are set as default in Cytobank, bioconductor packages such as flowCore and in many multivariate analysis methods such as flowSOM, SPADE and viSNE. Arcsinh functions with 3 different cofactors are plotted in Figure 1A. In this Figure the linear function (dotted line) is closest to the arcsinh with cofactor 500 (blue line) where the linearly transformed region around zero is more predominant compared to arcsinh functions with lower cofactors (green and red line). The effect of the three different cofactors in the arcsinh-transformation of the same (2-peaks) data distribution is shown in Figure 1B. Low cofactors (5 and 150) cause splitting of the negative population in two peaks around zero (red and green line). As the cofactor increases, this artificial splitting is removed, the variances become smaller and two populations are correctly visualized.

### Logicle transformation

The logicle transformation is a bi-exponential function on which a 'logicle condition' is imposed.<sup>15,26</sup> The condition consists of having a null second derivative at the zero point so that the zero value represents the inflection point. The mathematical formula of the logicle function is as follows:

$B(x, T, W, M, A) = \begin{cases} T10^{-(M-W-A)} \left( 10^{x-W-A} - p^2 10^{-\frac{x-W-A}{p}} + p^2 - 1 \right), y \geq W + A \\ -T10^{-(M-W-A)} \left( 10^{W+A-y} - p^2 10^{-\frac{W+A-y}{p}} + p^2 - 1 \right), y < W + A \end{cases}$	Eq. 2a
---	--------

Where  $T$  is the maximum value in the displayed scale range;  $M$  is the width of the displayed scale range in natural logarithm unit, defined as  $M = \log(T)$ ;  $W$  is the width parameter defining the linear range of the function;  $A$  describes the range of the negative values that should be included in the scale. For standard logicle transformation  $A = 0$ , which means that the most negative value displayed is equal, in absolute value, to the top value of the linear range. While  $M$ ,  $T$ ,  $A$  are set by experimental circumstances or fixed to a default value,  $W$  should be adjusted. The value for the logicle parameter can be derived as follow:<sup>15,26</sup>

$W = \frac{M - \left( \log \left( \frac{T}{ r } \right) \right)}{2}$	Eq. 2b
--	--------

Where  $r$  is a reference value in the negative region, which can be chosen as the 5<sup>th</sup> percentile of all the data values below zero. This is an arbitrary choice and it should be applied only after removing extreme negative data points.<sup>15,26</sup>

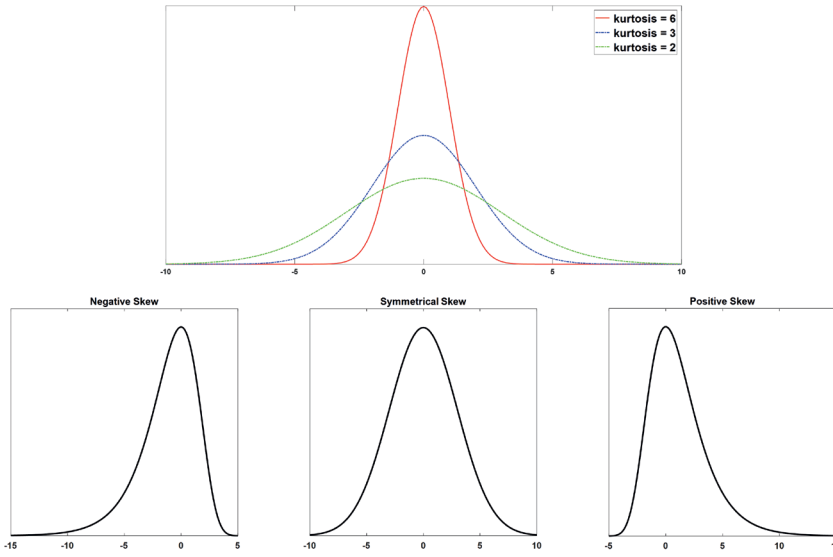
### Variance stabilization and normality of flow cytometry peaks

Bi-exponential function-based transformations perform variance stabilization of the signals, as they remove the correlation between data variability and mean, which is typically present in MFC data. If the variance between the signals is not stabilized, cell populations with higher signal intensity will have a larger variance which does not necessarily reflect the true marker variability. For accuracy in MVA, optimally transformed MFC data is characterized by peaks resembling normal distributions with homogenous variances (homoscedasticity), which are not dependent on the fluorescence intensity. The variance between distributions, after optimal transformation, will therefore reflect the true variability in protein expression of the cell subsets. This facilitates the comparison of cell populations with different marker expression levels. Checking for variance stabilization between multiple peaks for one marker can be used as measure of how well (bi-exponential) functions transform the MFC data. This may be evaluated by using Bartlett's likelihood-ratio test, commonly chosen in statistics to check for homoscedasticity among multiple groups. The Bartlett's test is adopted in the flowVS algorithm.<sup>18,20</sup> Also in OTflow, we use Bartlett's test to assess the homogeneity of variance between peaks per measured marker across all the individuals. However, when a single cell population is present per marker, variance stabilization cannot be applied to peaks within the same sample, but only to peaks across all the individuals. With a single peak per sample and a low number of samples present, optimizing Bartlett's statistics is not preferred, since biologically relevant variance between different samples might be removed. Alternatively, optimal transformation of the single peak can be evaluated by how well the transformed peak resembles a normal distribution. The Jarque-Bera test is used to determine the normality of the single peaks. The test is based on estimation of kurtosis and skewness of the peak, which are schematically represented in Figure 2. These are properties related to the shape of distributions and they are commonly used to check the deviation from normality.

In particular, the kurtosis is an indicator of the sharpness of the peak distribution (Figure 2, top panel). A normal distribution is characterized by a kurtosis equal to 3; kurtosis smaller than 3 (platykurtic) indicates a flatter distribution with more data points in the tails of the curve; kurtosis bigger than 3 (leptokurtic) indicates a sharper distribution with less data points in the tails. The skewness corresponds to the spread of the data around the mean value. Skewness of a symmetric distribution is zero; a negative skewness indicates a distribution leaning to the left of the mean value; a positive skewness indicates a distribution leaning to the right of the mean value.

### The algorithm step by step

We describe below the process to choose an optimal cofactor for arcsinh transformation (Eq.1) for each channel measured. The same process can be applied to the optimization problem of the parameter  $W$  for logicle transformation (Eq.2).



**Figure 2. Skewness and kurtosis.**

Top panel: Distributions having different values of kurtosis are displayed. Normal distribution (blue) has a kurtosis equal to 3; distribution with a kurtosis larger than 3 (red, kurtosis equal to 6) is sharper than a normal peak; distribution with a kurtosis smaller than 3 (green, kurtosis equal to 2) is flatter than a normal peak. Bottom panel: Representation of distribution with negative skewness (left-skewed peak), symmetrical skewness (normal peak), positive skewness (right-skewed peak).

**Step 0:** *multiset structure of MFC data.*

Single MFC samples can be arranged in the 'multiset' matrix  $X$ , of size  $(\sum_1^I N_i \times J)$ , where  $N_i$  is the number of cells of the  $i^{\text{th}}$ -individual and  $J$  corresponds to the markers measured,  $1 \dots j \dots J$ .

**Step 1:** *data transformation.*

The matrix  $X$  is transformed by the arcsinh function as described in Eq. 1. We define  $X_{\log_i} = \text{arsinh}(\frac{X_i}{c})$  the arcsinh transformed matrix of the  $i$ -th sample, of size  $(N_i \times J)$ . The value of cofactor  $c$  is progressively increased by assuming values  $e^a$ , with  $a$  ranging from 0 to 10, in steps of 0.05, corresponding to 201 cofactor values.

**Step 2:** *each marker is analyzed per individual.*

Each sample  $X_{\log_i}$  is randomly subsampled (without replacement) to 1000 cells. The subsampling is repeated 100 times by a Monte-Carlo cross-validation. If the sample size is smaller than 1000 cells, then the non-subsampled set is used. For each subsample, the algorithm is applied to each marker individually, included in the column vector  $\mathbf{x}_{\log_i, j'}$  representing the arcsinh-transformed  $j^{\text{th}}$ -channel of the  $i^{\text{th}}$ -individual.

**Step 3: peak detection.**

Peaks or cell subpopulations are determined for each separate marker per measurement. First univariate probability density is estimated with kernel density estimate <sup>27</sup> by using Gaussian function as kernel. Then a peak finding algorithm (*findpeak* function in Matlab) is applied to the resulting density estimate. Peaks are identified by using a minimal peak prominence of 0.1% of the density estimate. The prominence of a peak measures how much the peak emerges due to its absolute height and its location relative to other peaks. If this is bigger than the chosen threshold, separate peaks will be identified. Peaks containing less than 5 cells will be disregarded from further analysis. This will prevent inclusion of very small peaks in the analysis which might be associated with outliers or noise in the data. Additionally, the mean of each leftmost peak is estimated. A negative mean suggests the presence of *peak splitting* due to the transformation and, in this case, a penalty is given to the measurement.

**Step 4: peak counting and penalties.**

Subsequently, the number of peaks is checked per measurement for each marker. Depending on the number of peaks found, different statistics or penalties are applied.

**a)** If no peaks are present (which means all the detected peaks contained less than 5 cells), a penalty is given; **b)** if only one peak is present, *Jarque-Bera* statistics are applied; **c)** Samples containing multiple (two or more) peaks are grouped together; *Bartlett* statistics are then applied to check variance stability of all the peaks from the merged samples.

**Step 5a: Jarque-Bera statistics.**

When a single peak per marker is found, the Jarque-Bera (*JB*) statistical method <sup>28</sup> is applied to optimize the normality of the peak. For each value of the cofactor *c*, the score  $JB(c)$  is estimated as shown in Equation 3.

Where  $S(c)$  and  $K(c)$  represent skewness and kurtosis of the distribution of the peak, respectively. The smaller the value of the  $JB(c)$  statistics, the more likely the peak resembles a normal distribution.

(a)	$S(c) = \left( \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}} \right)$	Eq. 3
(b)	$K(c) = \left( \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} \right)$	
(c)	$JB(c) = \frac{n}{6} \left( S^2 + \frac{(K - 3)^2}{4} \right)$	

**Step 5b: Bartlett statistics.**

Subsamples with multiple peaks per marker are bundled together. Bartlett statistics test for equal variance across all the peaks from all the merged subsamples against the alternative hypothesis that variances are unequal across the peaks. The Bartlett statistics are calculated for each of the 201 values of  $c$  (see Eq. 1), as follow:

	$\text{Bartlett}(c) = \frac{(N - k) \ln(\sigma_p^2) - \sum_{i=1}^k (n_i - 1) \ln(\sigma_i^2)}{1 + \frac{1}{3(k-1)} (\sum_{i=1}^k (\frac{1}{n_i - 1}) - \frac{1}{N - k})}$	<b>Eq. 4</b>
--	---	--------------

where  $N$  is the number of cells included in all the found peaks,  $k$  is the number of peaks,  $\sigma_p^2$  the pooled variance of peaks,  $\sigma_i^2$  is the variance of the  $i^{\text{th}}$ -peak. When the variances between the peaks are equal, the Bartlett ( $c$ ) assumes the lowest value.

**Step 6-7: final score and optimal cofactor.**

For each Monte-Carlo iteration a score is calculated as contribution of penalties, *Jarque-Bera* and *Bartlett* statistics. The two statistical tests are weighted to assure that both statistics have similar impact in determining the final score. Mean and standard deviation of the scores of the 100 Monte-Carlo iterations are estimated. The optimum cofactor is the value corresponding to the lowest mean scores for each marker.

**Data set****HIV data**

The HIV dataset consists of 7 healthy control samples and 18 patient samples, 14 patients were treated with anti-retroviral therapy and 4 untreated. Peripheral blood samples were obtained from individuals in the context of standard diagnostic care. Residual blood not used for the standard diagnostic tests was used for research, with Informed Consent from patient according to protocols of the UMC Utrecht.

The MFC panel comprised 10 fluorescently labeled antibodies: CD123, CD14, CD16, CD3, CD4, CD8, CD56, CD20, CD193, CD62L for identification of different leukocyte subsets in the peripheral blood. Per sample between 250,000 – 500,000 cells were measured on a BD LSR-II Fortessa. Prior to MVA, the single cells were selected, and debris was excluded from the analysis (Suppl Figure S1) using manual gating (Flowjo analysis software, Tree Star Inc., Ashland, Oregon). The gated fcs files were exported.

## **Multivariate Analysis (MVA)**

Here the way different multivariate analysis methods were applied is shortly explained. For background information about the methods we refer to the supplementary material or the original papers about the methods (see references).

### **Principal Component Analysis (PCA)**

Principal Component Analysis<sup>23</sup> (PCA) is a dimensionality reduction technique which has been widely used for analysis of MFC data.<sup>6,7,29,30</sup> After arcsinh transformation, the data were mean centered and scaled over all the samples, prior to the PCA analyses.

### **viSNE**

viSNE analyses were performed in Cytobank<sup>17</sup> with the following parameters: perplexity 30, random seed, # Iterations 1000, Theta 0.5. To perform the viSNE analysis on the HIV dataset, each individual was randomly down-sampled to 2000 events. This enabled decrease of the running time and thereby computational burden and crowding effects were avoided.

### **flowSOM**

The flowSOM<sup>9</sup> algorithm was performed by running the R code available at Bioconductor. The algorithm was trained on the MFC data using suggested (default) parameters: 100 as number of nodes, Euclidean distance to find nearest neighbor, and a training length of 10 epochs. A minimum spanning tree was then built to visualize the clusters detected by the algorithm.

### **Citrus**

Citrus<sup>10</sup> was run using the online platform Cytobank<sup>17</sup> with the following default parameters: minimum cluster size of 5% and a cross-validation fold of 5.

## Results

### Flow cytometry datasets contain a considerable amount of negative values

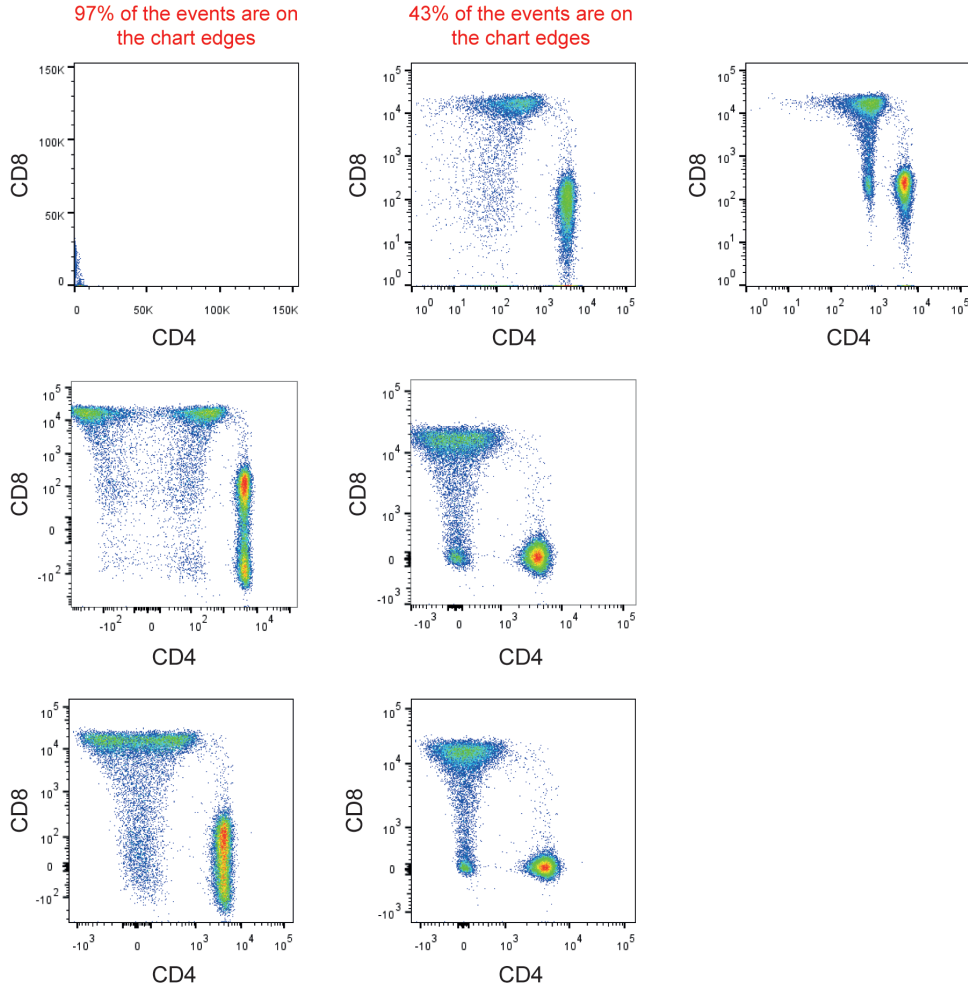
In the datasets used in this article the percentage of cells displaying one or more negative intensities ranged between 39% and 96%, and these negative intensities vary by 8 or 9 orders of magnitude (Supplementary Table 1).

As explained in the introduction, background signal may be one of the causes of negative values in a dataset. If a lot of unbound fluorophores are present in the solution, the background signal is higher than the signal of a particle or cell to which no such antibody has bound, again leading to negative values (Suppl Figure 2).

Another cause of the appearance of negative intensities in a flow cytometry dataset, is data spreading. Cell populations that are negative for a certain marker, but positive for another marker with spectral overlap in the channel of the negative marker, can cause data spreading for the negative marker. For the positive cell population, a wide range of values for the negative marker may be found. When compensation is applied, the intensities for the negative marker will be decreased and part of the lowest values are shifted into the negative range (Suppl Figure 3).<sup>31</sup>

For flow cytometry data, the log scale provides an informative and proper display of populations in the higher intensity range as well as of populations in the lower intensity regions. Populations with positive low fluorescent intensities which may be hardly discerned on a linear scale are well visible and separated on a logarithmic scale. (Suppl Figure 4) However, negative values cannot be logarithmically transformed. If representing data containing negative intensities on a logarithmic scale, events are plotted onto the axes. (Figure 3A+B). To overcome this problem, it might be suggested to shift the whole dataset to positive values and then apply log transformation. Doing so, events in the lower region are overly dispersed and the ratio between values changes. This would lead to incorrect estimation of signal means, which is not desirable, as shown in Figure 3 C. As a solution bi-exponential transformations are used, e.g. arcsinh or logicle transformation. For both functions, the width of the linear region is determined by a parameter. When this parameter is not carefully chosen, artifacts can arise at the transition point of the linear part to the logarithmic part.<sup>32</sup> Figure 3D-G shows logicle and arcsinh transformed flow cytometry data with different parameters. Homogeneous populations with a mean intensity close to zero can be split into two distinct populations when the value for the parameter is too low. We will refer to this phenomenon as peak splitting. In Figure 3D and 3F suboptimal parameters were used for arcsinh and logicle transformation, respectively. In these plots 6 and 4 or 5 cell populations can be discerned, respectively, while in the

Figures with an optimally defined parameter for transformation (3E and 3G) 3 cell populations are found.

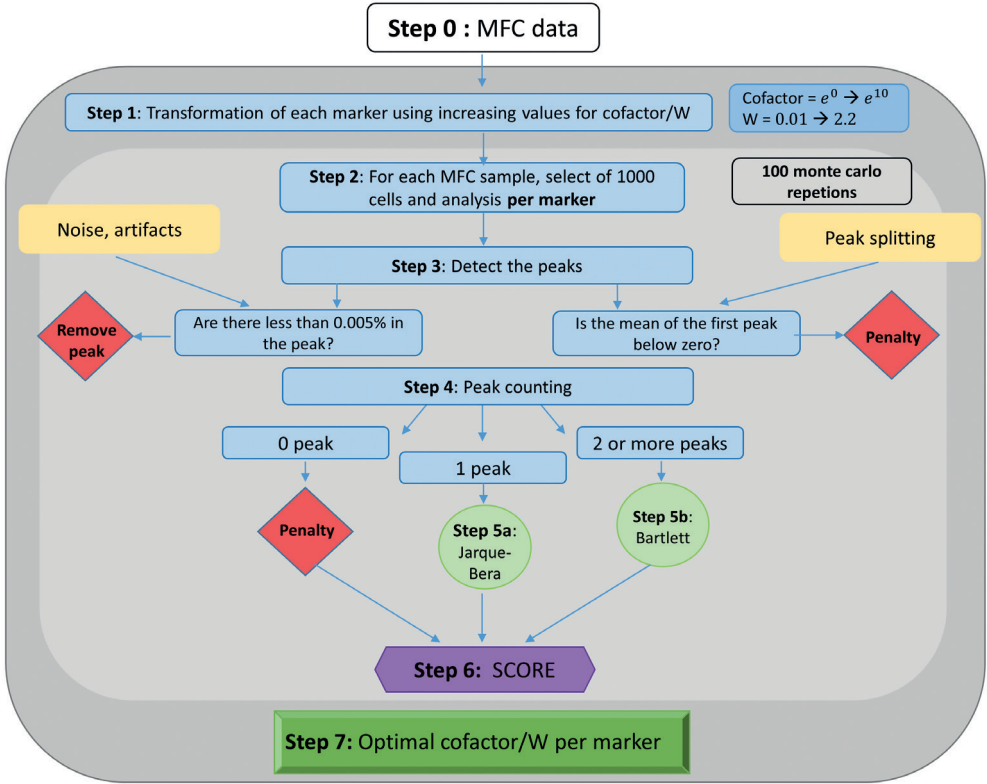


**Figure 3. Effect of transformation parameter choice on visualization of MFC data.**

CD4 CD8 biplots of gated CD3+ lymphocytes. All plots show the same original data set, with different transformation of the x and y axes. A) Linear axes (no transformation); B) Logarithmically transformed axes; C) Data was shifted to positive by subtracting the most negative values for both markers from the intensities measured and displayed on logarithmically transformed axes; D) Hyperbolic-arc sine transformed axes with a low cofactor; E) Hyperbolic-arc sine transformed axes with a high cofactor; F) Logicle transformed axes with a low W; G) Logicle transformed axes with a high W.



We developed a new method, named OTflow, which defines the optimal co-factor or W value for arcsinh or logicle transformation, respectively. The method has been described in the methods section step by step. A schematic overview of the algorithm is shown in Figure 4.



**Figure 4. Schematic overview of OTflow algorithm.**

**Steps 1-6**, included in the light grey sector, are repeated through the 100 monte-carlo iterations. The validated final score is calculated as average of the score obtained each repetition. The score is a result of Jarque-Bera statistics (estimated for single peak), Bartlett statistics (estimated for 2 or more peaks) and Penalty. A penalty is given when a negative peak is found (peak splitting) or no peak is detected. The optimal cofactor or W parameter (**Step 7**) is the one that produces the lowest mean score per marker.

The parameter optimization algorithm OTflow was tested on several different flow cytometry data. The number of cells subpopulations varied per dataset. In fact, manual gating strategy was applied to select or exclude certain cell subpopulations, depending on the specific research question or diagnostic study, as described in detail below.

HIV dataset

The HIV dataset comprises blood samples obtained from healthy controls and HIV patients. The patient group is heterogeneous, since it contains treated and untreated patients. The MFC panel used includes 10 surface markers, mostly differentiation markers, aimed to identify the most common leukocyte subsets in the peripheral blood. OTflow was used to identify the optimal cofactor for the arcsinh transformation for each of the 10 fluorescence channels in the HIV dataset. Bartlett and Jarque-Bera statistics were estimated by the OTflow algorithm depending on the amount of peaks found per cofactor tested, ranging between  $e^0$  and  $e^{10}$  (equal to 0 and approximately 22,026). Additionally, a penalty was given when no peaks were present or when negative peaks were found in the transformed data. The average final score and standard deviation obtained per marker for varying cofactors are shown in Figure 5. For most of the markers, higher scores are obtained when very low and very high values of the cofactor are used for the arcsinh transformation, while for the medium values a dip in the curves is present between  $e^4$  and  $e^7$  ( $\approx 54.6 - 1,097$ ). An exception to this trend is observed for marker CD20 having maximum score value for central cofactors. By investigating the contribution of the different statistics (Bartlett's and Jarque-Bera) to the average score (Figure S5), we observed that such trend is due mainly to Bartlett's statistics, meaning that variances between peaks in this range of cofactors are not well stabilized.

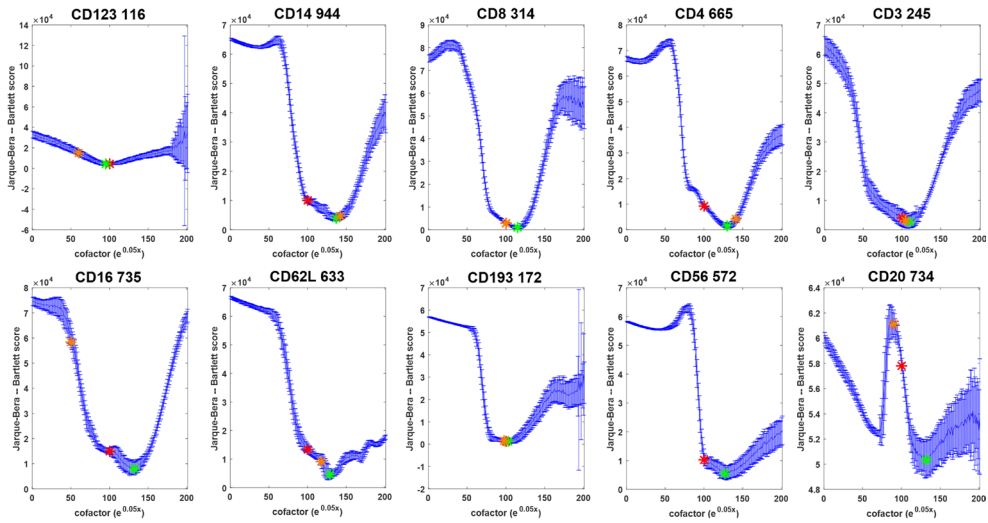
Table 1 compares the optimal cofactors found by OTflow with the results obtained by applying flowVS on the same dataset (Table 1).

**Table 1. Cofactors for arcsinh transformation calculated per marker by flowVS and OTflow algorithms.**

	CD123	CD14	CD8	CD4	CD3	CD16	CD62L	CD193	CD56	CD20
flowVS	20	1233	150	1164	186	12	359	132	0.3	90
OTflow	116	944	314	665	245	735	632	172	572	734

Running the flowVS algorithm on the dataset led to significantly lower cofactors for marker CD123, CD16, CD56 and CD20 compared to the results of OTflow and to the default cofactor 150. Figure 6 A-C shows the distribution of these four representative surface markers on leukocytes of healthy controls (blue) and HIV patients (red), after transformation with a default cofactor 150, or with the cofactors determined by flowVS or OTflow, respectively. Arcsinh transformation with default cofactor of 150 for all the markers produced spurious splits of the leftmost population for CD56 and CD20 (Figure 6A, marked with green rectangles). When using the cofactors calculated by flowVS, negative intensities, resulting from peak splitting, were present for all the four selected markers

(Figure 6B, marked in green). The emergence of the extra peaks in the data could lead to the conclusion that there are 3 populations present: a cell population not expressing the marker; a cell population with a weak expression of the marker and a cell population which strongly expresses the marker. Transformation with OTflow-defined cofactors did not lead to peak splitting (Figure 6C). For the remaining markers no peak splitting was present after transformation with default cofactor 150 and cofactors determined by flowVS, as shown in Figure S5, Supplementary Material II. When using the optimal cofactors from OTflow for the transformation, the variance between the peaks is stabilized more evenly for all the measured markers, especially compared to the transformation with the default cofactor.



**Figure 5: Scores of the OTflow algorithm for the HIV dataset.**

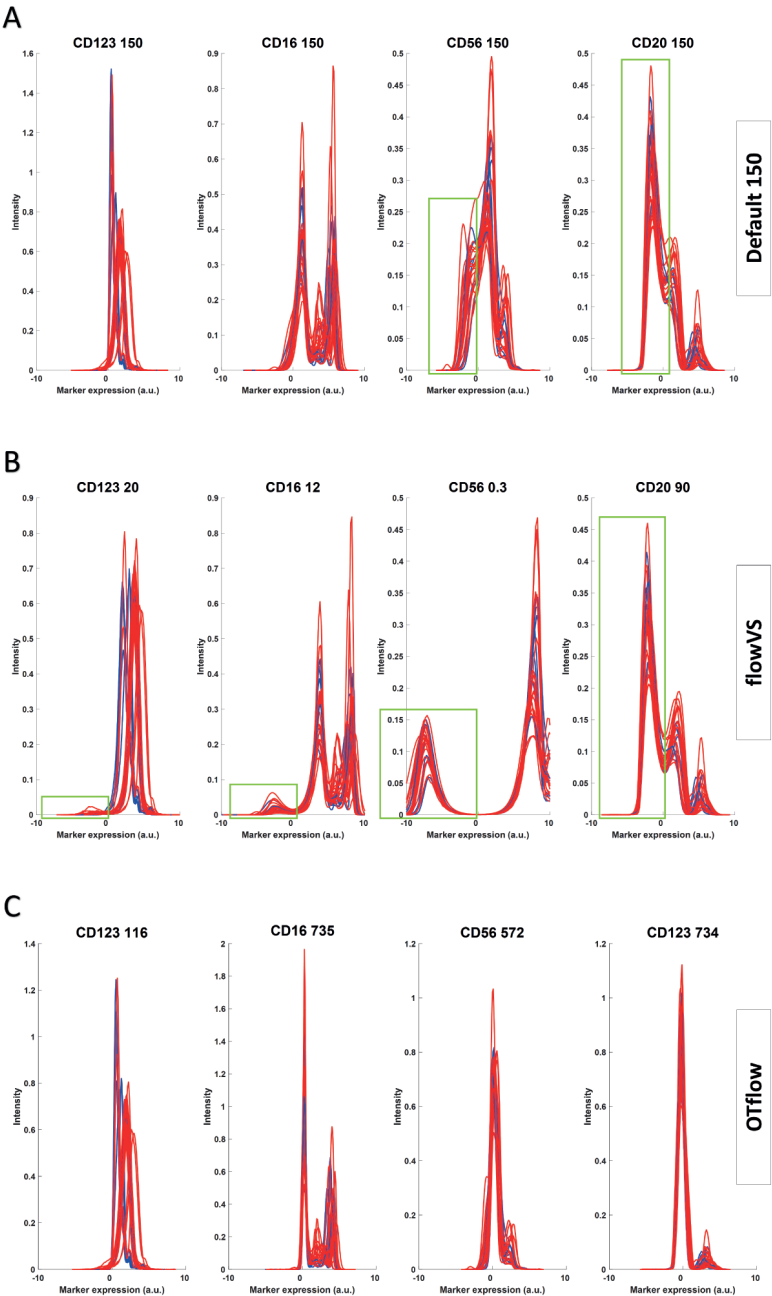
Graphs show the OTflow average total scores and standard deviation per measured marker for cofactors ranging between  $e^0$  and  $e^{10}$  ( $x=0 - x \approx 20000$ ). The green star indicates the lowest score which defines the optimal cofactor. Its value is shown in the title of each plot. The red star indicates the value of the score corresponding to the default cofactor 150 and the orange star corresponds to the optimal value found by flowVS algorithm. No orange star is present in the plot of CD56 because the value estimated by flowVS (0.26) is too low to be inside the range considered by OTflow.

## Multivariate Analysis

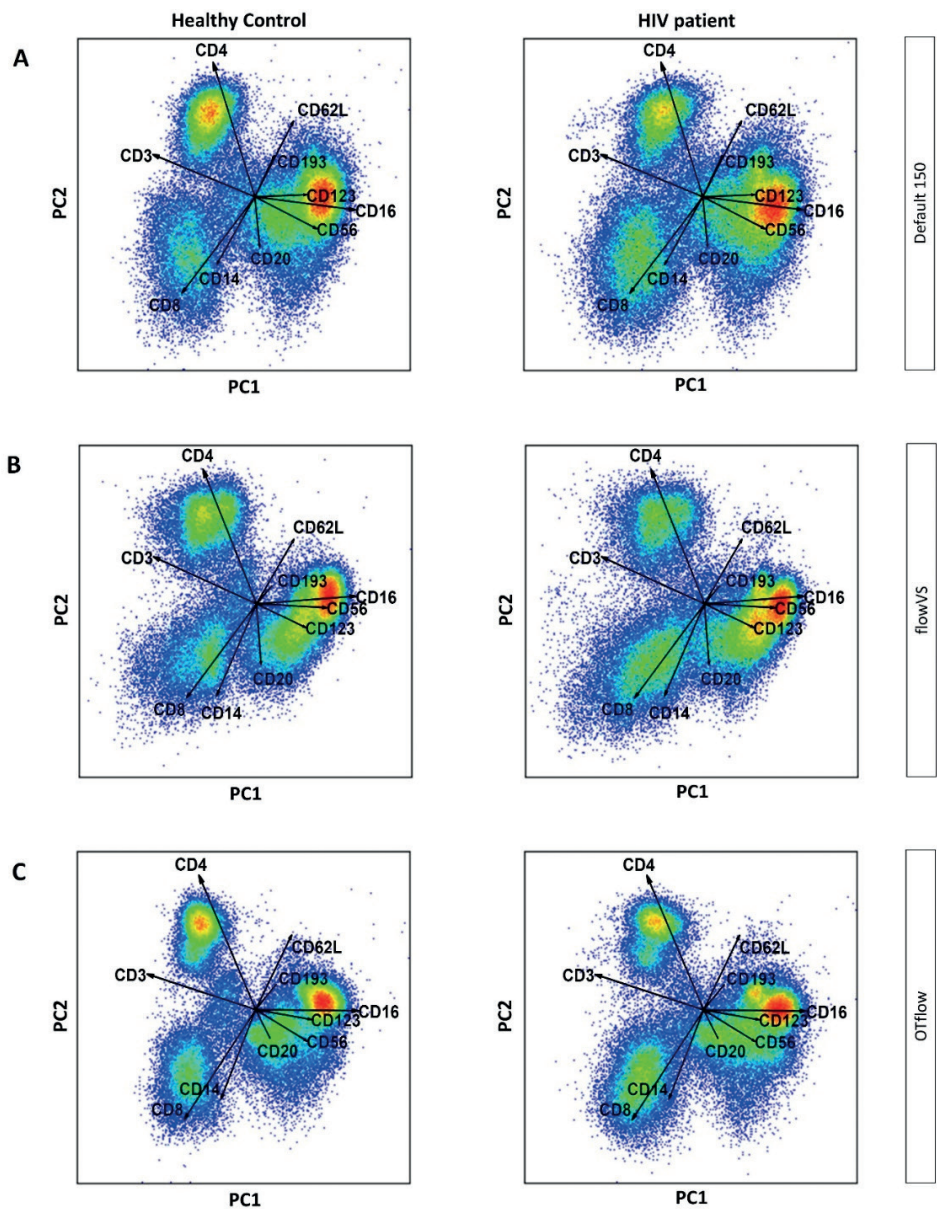
We evaluated the effect of the arcsinh transformation using the default cofactor 150 and cofactors determined by OTflow and flowVS algorithms by applying various multivariate analysis methods which enable visualization of MFC data. These methods include the dimension reduction techniques Principal Component Analysis (PCA) 23 and viSNE 4; the clustering method flowSOM 9 and the classification method Citrus 10.

### Principal Component Analysis on HIV dataset

When analyzing the PCA results of the differently transformed data, differences were observed in the cellular scores distributions (Figure 7). In particular, in the PCA model of the flowVS-transformed data, two putatively CD3<sup>bright</sup>CD4<sup>bright</sup> and two CD3<sup>bright</sup>CD8<sup>bright</sup> cell subsets were identified in healthy controls by using the vectors of the markers as a compass for marker co-expression level (Figure 7B). The two CD3<sup>bright</sup>CD4<sup>bright</sup> and two CD3<sup>bright</sup>CD8<sup>bright</sup> populations were gated and shown in more detail in Figure 8.



**Figure 6.** Histograms of the leukocyte expressions of CD123, CD16, CD20 and CD56 after transformation with default cofactor 150 (A), flowVS cofactors (B) and with OTflow defined cofactors (C). Negative peaks originated due to suboptimal transformation, resulting into peak splitting, are marked with a green box.



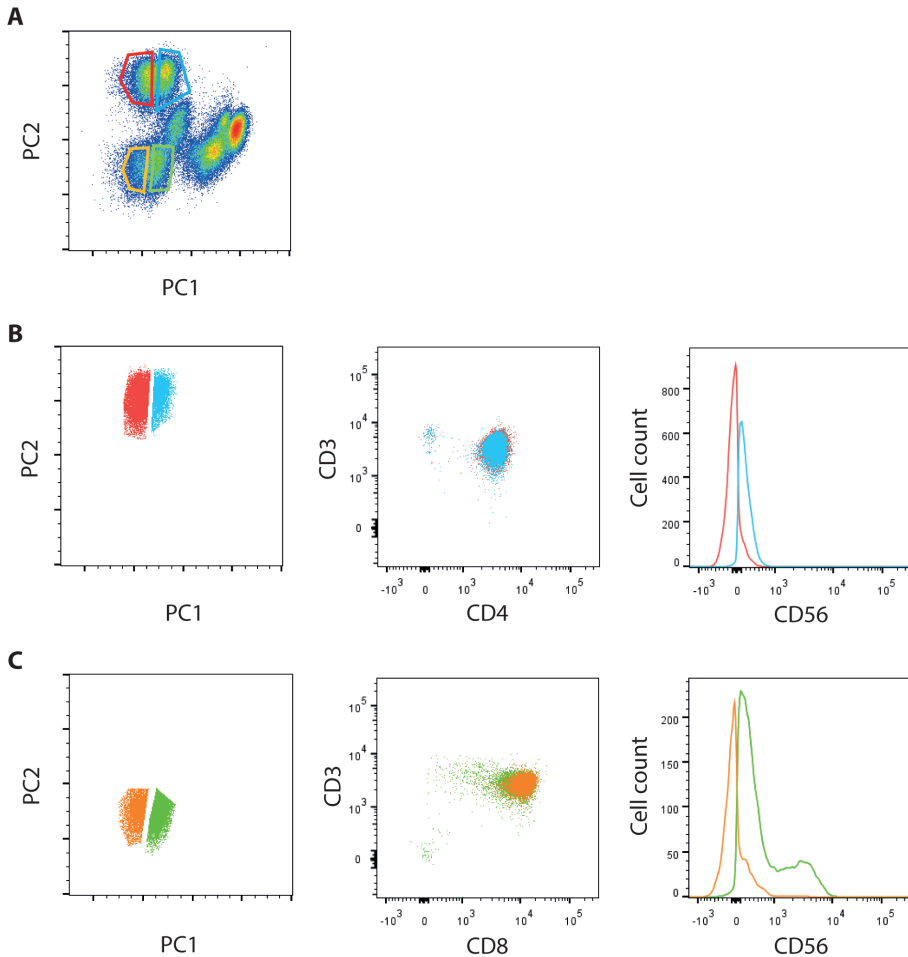
**Figure 7. PCA of differently transformed HIV datasets.**

PCA plots of PC1 vs PC2 of the HIV data after arcsinh transformation with default cofactor 150 (A), flowVS determined cofactors per channel (B), and channel-specific cofactors as calculated by OTflow (C). The cells are plotted based on the PC1 and PC2 scores and the loadings (vectors), representing the marker expressions, are projected within the PC models. The left panels represent all healthy controls and the right panels represent the HIV patients (100.000 cells are shown per plot).

By backgating we confirmed that the two CD3<sup>bright</sup>CD4<sup>bright</sup> subsets differed only based on CD56 expression levels: one of the subsets having almost only negative intensities for CD56 while the other subset mostly consisted of positive intensities for CD56 (Figure 8A+B). The same difference accounts for the two CD3<sup>bright</sup>CD8<sup>bright</sup> subsets (Figure 8A+C). In contrast, the PCA results of the datasets transformed with default cofactor 150 only contains a single CD3<sup>bright</sup>CD4<sup>bright</sup> population and one CD3<sup>bright</sup>CD8<sup>bright</sup> population (Figure 7A), both with low expressions for CD56 (data not shown).

These have similar expression levels for CD3 and CD4, but they are separated due to differences in CD56 expression, showing positive and negative expression because of peak splitting. The presence of CD3<sup>bright</sup>CD4<sup>bright</sup>CD56<sup>bright</sup> and CD3<sup>bright</sup>CD8<sup>bright</sup>CD56<sup>bright</sup> NKT-like cells in the peripheral blood of healthy subjects has been described,<sup>33</sup> but not in these high proportions (38% of total CD3<sup>bright</sup>CD4<sup>bright</sup> cells and 55% of total CD3<sup>bright</sup>CD8<sup>bright</sup> cells, respectively). MVA results from these inadequately transformed data could result into misleading interpretations of the findings, when drawing conclusions about percentages of CD3<sup>bright</sup>CD4<sup>bright</sup>CD56<sup>bright</sup> and CD3<sup>bright</sup>CD8<sup>bright</sup>CD56<sup>bright</sup> populations in healthy controls vs HIV patients.

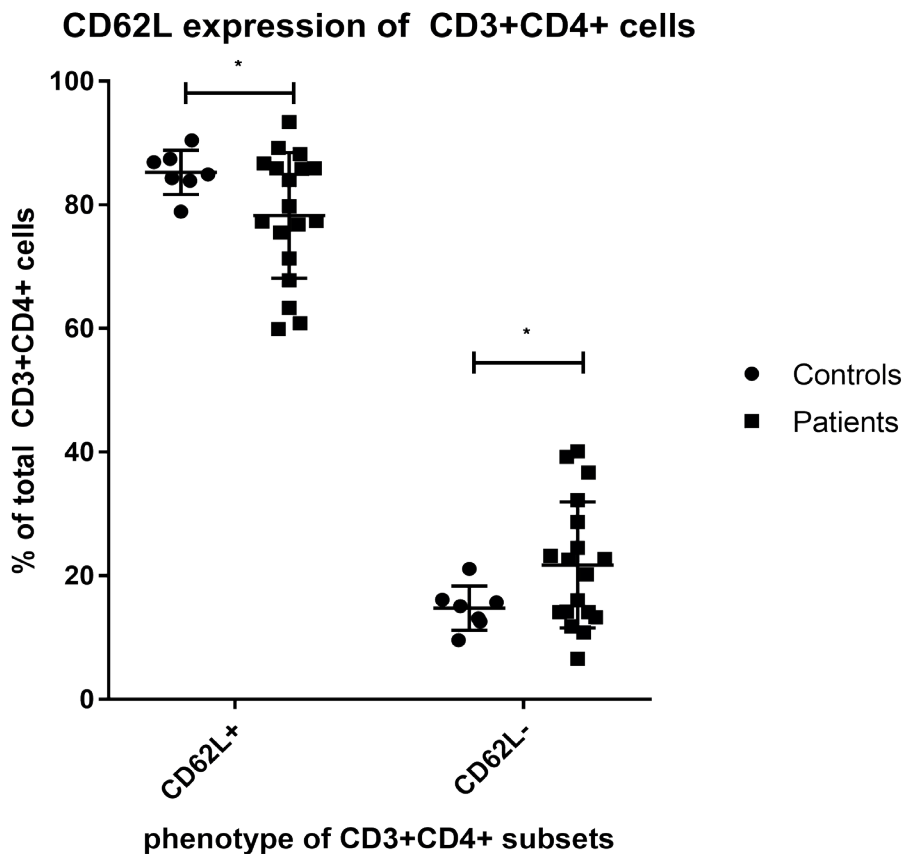
Importantly, besides the appearance of artificial cell populations, suboptimal transformation can also lead to loss of biologically relevant information. When analyzing the PCA model of the optimally transformed data by OTflow, also 2 types of CD3<sup>bright</sup>CD4<sup>bright</sup> populations were found (Figure 7C), one major population and one minor population attached to it. In contrast to the flowVS transformed data, the two CD3<sup>bright</sup>CD4<sup>bright</sup> populations in the OTflow transformed analysis expressed CD56 to the same level (CD56<sup>low</sup>, data not shown). The two CD3<sup>bright</sup>CD4<sup>bright</sup> populations differed from each other based on expression levels of CD62L. Naïve T cells express CD62L at high levels, while memory T cells display low levels of CD62L expressions. And HIV patients had significantly higher percentages of CD3<sup>bright</sup>CD4<sup>bright</sup>CD62L<sup>low</sup> cells compared to healthy controls (Figure 9,  $p=0.018$ ), as has been reported.<sup>34–36</sup> This difference was missed in the PCA model of the flowVS transformed data because of the more notable but artificial difference in CD56 expression for the CD3<sup>bright</sup>CD4<sup>bright</sup> cells caused by peak splitting. Also in the PCA model of the transformed data with the default cofactor 150 this information was not represented, since we did not find two separate CD3<sup>bright</sup>CD4<sup>bright</sup> cell populations.



**Figure 8. The  $CD3^{\text{bright}}CD4^{\text{bright}}$  and  $CD3^{\text{bright}}CD8^{\text{bright}}$  populations in the flowVS transformed data are both split into two subpopulations based on CD56 positive and negative expressions.**

A) Representative PCA plot of a HIV patient after transformation with cofactors defined by flowVS. The two  $CD3^{\text{bright}}CD4^{\text{bright}}$  were gated in red and blue and the two  $CD3^{\text{bright}}CD8^{\text{bright}}$  populations were gated in orange and green. B) Left panel: PC plot of the gated  $CD3^{\text{bright}}CD4^{\text{bright}}$  cells. Middle panel: CD4/CD3 bi-plot of these gated cells. Right panel: histograms of the CD56 expression of the two  $CD3^{\text{bright}}CD4^{\text{bright}}$  subpopulations. The red  $CD3^{\text{bright}}CD4^{\text{bright}}$  cells almost only have negative CD56 expression levels; the blue population mainly consists of cells with positive CD56 intensities. C) Left panel: PC plot of the gated  $CD3^{\text{bright}}CD8^{\text{bright}}$  cells. Middle panel: CD8/CD3 bi-plot of the gated cells. Right panel: histograms of the CD56 expression of the two subpopulations. The orange  $CD3^{\text{bright}}CD8^{\text{bright}}$  cells almost only have negative CD56 expression levels; the green population mainly consists of cells with positive intensities for CD56.



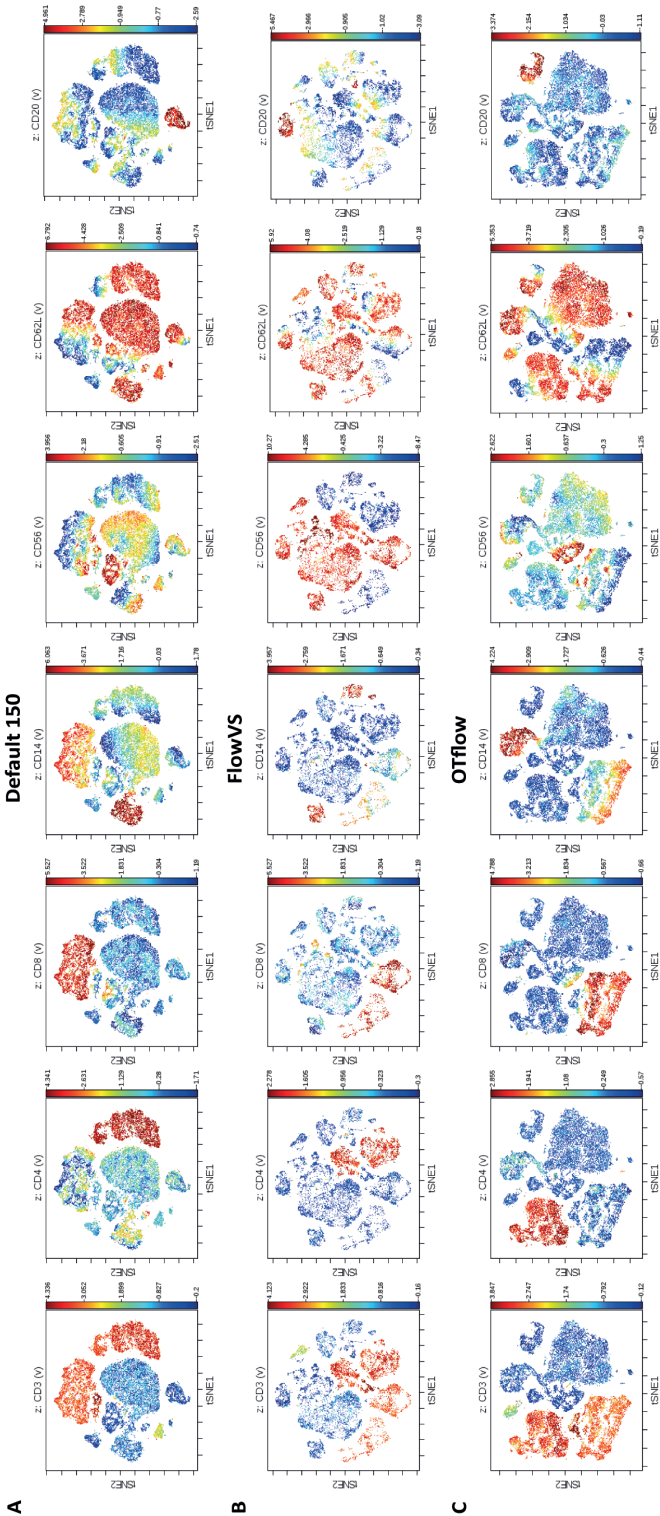


**Figure 9. Scatter plot showing the percentage of CD3<sup>bright</sup>CD4<sup>bright</sup>CD62L<sup>bright</sup> and CD3<sup>bright</sup>CD4<sup>bright</sup>CD62L<sup>low</sup> cells of the total CD3<sup>bright</sup>CD4<sup>bright</sup> cell population per patient and control.**

The mean per group is depicted in red. (\* p = 0.018). The PCA results of the OTflow transformed data reveal a significant increase in the percentage of CD3<sup>bright</sup>CD4<sup>bright</sup>CD62L<sup>low</sup> cells of the total CD3<sup>bright</sup>CD4<sup>bright</sup> cells in the peripheral blood of HIV patients compared to healthy controls.

**viSNE analysis on HIV dataset**

The viSNE algorithm <sup>4</sup> was also applied to the differentially transformed HIV dataset. This showed that the deceptive results were inherent to suboptimal transformed data, and not specific for a certain type of MVA technique. The resulting viSNE maps, colored based on the expression levels of CD3, CD4, CD8, CD62L, CD14 and CD56, are shown in Figure 10. When comparing the viSNE maps of the differently transformed data, the flowVS transformed dataset contains more ‘cell populations’ than the datasets transformed with the default cofactor or with the cofactors defined by OTflow. Comparable to the PCA



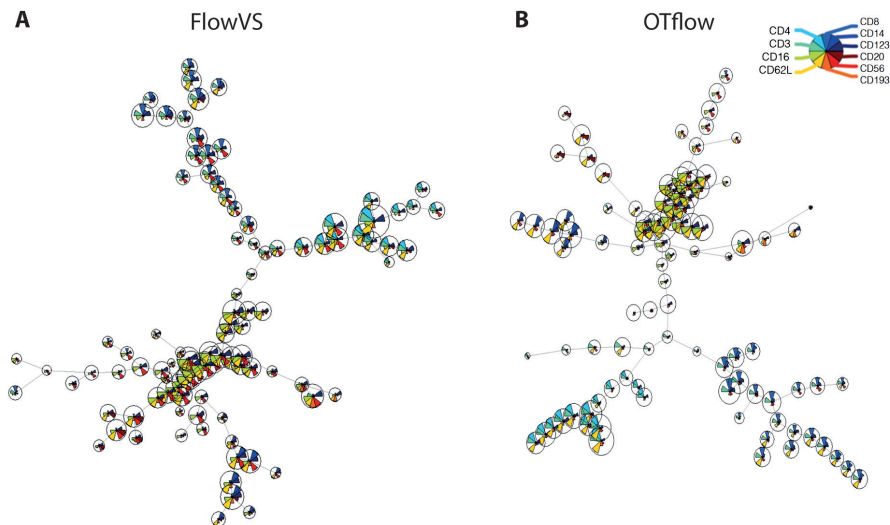
**Figure 10. visNE maps of default transformed cofactor 150 (A), flowVS transformed (B) and OTflow transformed (C) HIV datasets.**

The color coding of the cells represent the expression levels for the specific marker named in the title of the plots. The visNE maps of the variously transformed datasets differ greatly from each other.

results, the viSNE map of the flowVS transformed data shows CD3<sup>bright</sup>CD4<sup>bright</sup> cells and CD3<sup>bright</sup>CD8<sup>bright</sup> cells divided over 2 clusters, of which one seems to be CD56<sup>bright</sup> and the other CD56<sup>low</sup> (Figure 10B). In addition to these putative T-cell sub-clusters, also monocytes (CD14<sup>bright</sup> expression) appear to be split in two separate clusters, based on high and low CD56 expression levels. Additionally, the CD56 heat map shows a clear pattern with only CD56<sup>low</sup> or CD56<sup>high</sup> cells, while for the transformation with both the default cofactor 150 and OTflow-based cofactors, the CD56 expression levels gradually range from low to high. Both for the default 150 and OTflow transformed viSNE results the CD4<sup>bright</sup> and CD8<sup>bright</sup> T-cell populations and monocyte population are represented by a more homogeneous cell cluster, not subdivided based on CD56 diverse expression levels (Figures 10A+C). In summary, viSNE analysis of poorly transformed data results into a more complicated viSNE map with cell populations that manifest due to peak splitting.

### flowSOM on the HIV dataset

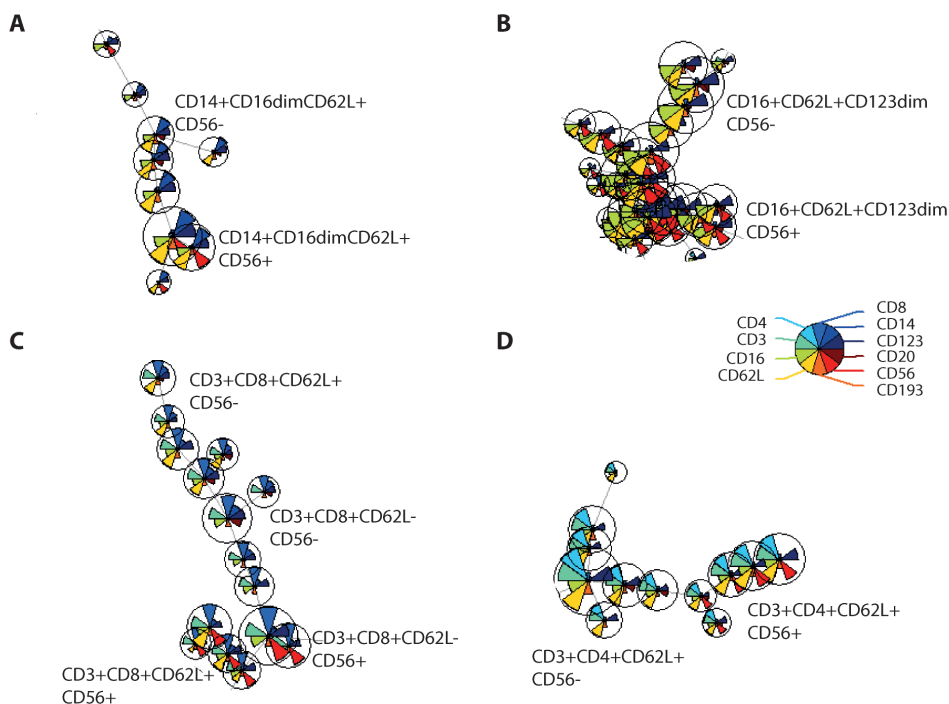
Next to PCA and viSNE analyses, which both adopt dimension reduction techniques to visualize MFC data, we performed cluster analysis using the flowSOM algorithm.<sup>9</sup> The algorithm produces self-organizing maps to simultaneously visualize MFC data in clusters/nodes represented either in a minimum spanning tree or in a two dimensional grid. Cells with similar phenotypes are clustered together in the same node. For conciseness reasons we only show the result of the flowVS and OTflow transformed HIV datasets, which differed the most from each other (Figures 11A+B). Figure 12 and 13 will be used to explain the differences between the two flowSOM maps.



**Figure 11: FlowSOM results of flowVS and OTflow transformed HIV data.**

A) flowSOM results HIV data transformed with flowVS-defined cofactors; B) flowSOM minimum spanning tree built on the OTflow transformed HIV data.

The mean marker expressions are visualized in each node by a pie chart. The size of the nodes corresponds to the relative amount of cells in each node.



**Figure 12. Leukocyte populations in flowVS transformed flowSOM map are divided over CD56<sup>bright</sup> and CD56<sup>low</sup> nodes..**

FlowVS transformed flowSOM branches containing different leukocyte clusters are enlarged and pie charts depicting the phenotypes are shown: A) monocyte (CD14<sup>bright</sup>) subpopulations; B) neutrophils (CD16<sup>bright</sup>CD62L<sup>bright</sup>); C) CD8 positive T cell subpopulations (CD3<sup>bright</sup>CD8<sup>bright</sup>); D) CD4 positive T cell subpopulations (CD3<sup>bright</sup>CD4<sup>bright</sup>)

In the FlowSOM results of the flowVS transformed HIV dataset we identified branches containing nodes with similar phenotypes, only differing based on CD56 expression. Monocyte (CD14<sup>bright</sup>), neutrophil (CD16<sup>bright</sup>) and CD4<sup>bright</sup> and CD8<sup>bright</sup> T-cell subpopulations were divided into CD56<sup>low</sup> and CD56<sup>bright</sup> nodes (Figure 12 A-D). As an example, the putative CD4<sup>bright</sup> and CD8<sup>bright</sup> T-cell subpopulations that show differential expression of (mainly) marker CD56 due to peak splitting, are found in multiple nodes of which the phenotypes are shown in Figure 12 A-D. This is in contrast to the OTflow transformed flowSOM map, where these cell populations all have low CD56 expressions. The 7 nodes with a CD56<sup>bright</sup> phenotype in the OTflow map are also CD16<sup>dim</sup>, fitting with the NK cell phenotype (Figure 11b, top branch). When peak splitting occurs due to suboptimal transformation with poorly-defined cofactors, the appearing negative cell populations may occupy unique nodes at the expense of biologically distinct cell populations which

may be clustered together in the same node. This is in contrast to PCA and viSNE methods, which do not require the user to define the number of cell populations beforehand. As expected, cell subsets which are defined as individual nodes in the OTflow flowSOM map are missing in the flowVS transformed results, like eosinophils (CD193+CD123+) and B-cells (CD20+CD62L+).

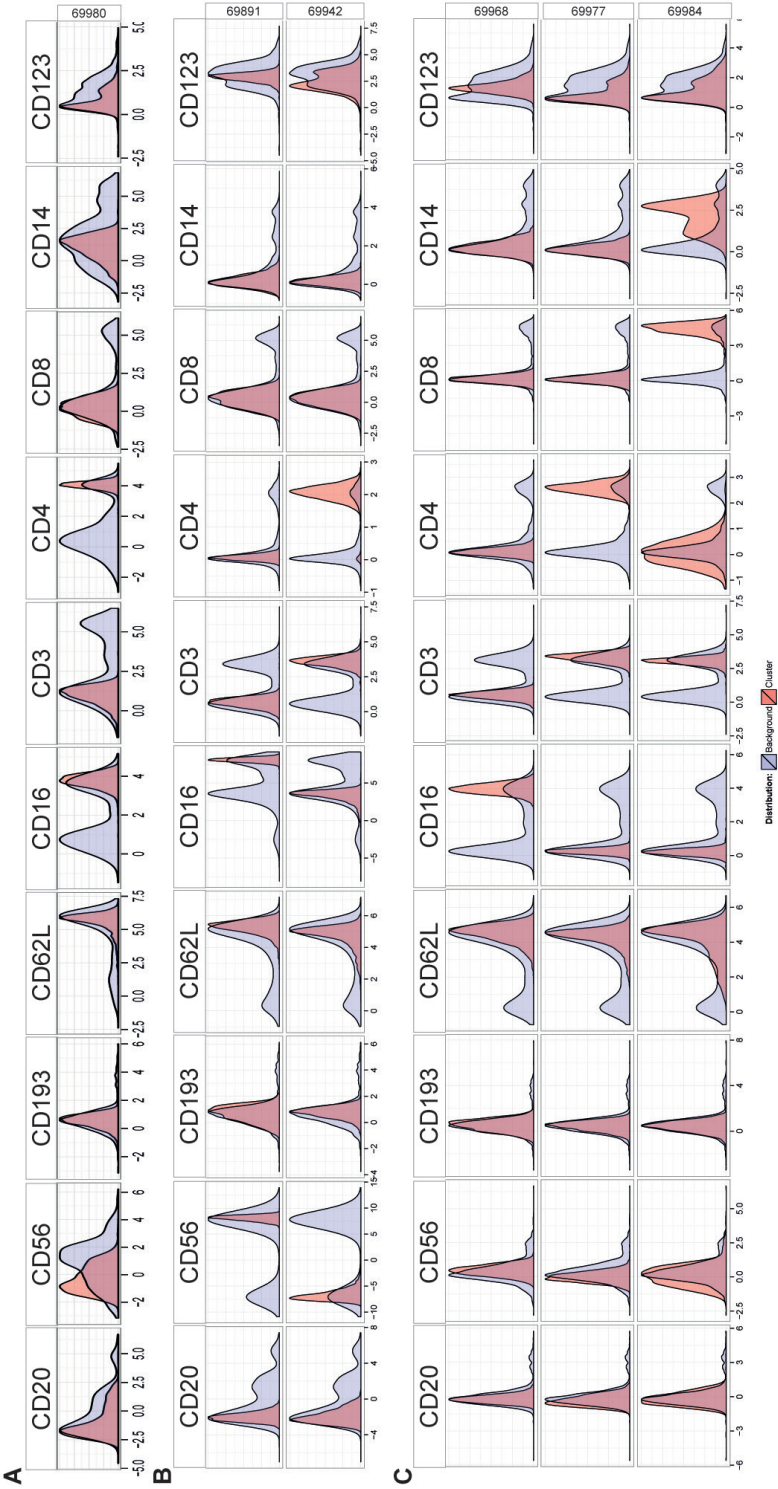
### Citrus on the HIV dataset

Citrus analysis<sup>10</sup> was performed to check whether suboptimal transformations of HIV data could affect the discrimination between healthy control individuals and HIV patients.

Citrus was performed on the HIV dataset transformed using the default cofactor 150, and the cofactors determined by flowVS and OTflow algorithms. For a balanced classification analysis, 7 healthy controls and 7 HIV patients treated with anti-retroviral therapy were selected. The method was run on the online Cytobank platform.<sup>17</sup> The accuracy of the models constructed on the differentially transformed data is shown in Figure S5 Supplementary Material II. Low accuracies are obtained in all the three analyses for varying numbers of features, with a range of 20%-30% misclassified samples. A considerable overlap in cellular phenotype between healthy controls and treated HIV patients, of which the immune system may resemble the homeostasis status, might be the cause of the low accuracies achieved.

However, when analyzing the clusters produced as the most discriminant between the two groups, misleading conclusions of the findings might be drawn. The phenotypes of the clusters selected by Citrus for the different analyses are shown in Figure 13.

In the case of data transformed with default cofactor 150, the discriminant model can discriminate between the two groups by the abundance of Cluster A in the healthy control group. The phenotype of such cluster is associated to CD3<sup>low</sup>CD4<sup>bright</sup>CD16<sup>bright</sup>CD62L<sup>bright</sup> cells with apparent gradual expression (both low and dim) of markers CD20 and CD56. It is difficult to conclude which cell subset matches with this phenotype found, as neutrophils (CD16<sup>bright</sup>CD62L<sup>bright</sup>) normally do not express CD4 and CD4<sup>bright</sup> T cells are expected to express CD3. Unfortunately, the Citrus algorithm does not enable further analysis of clusters found by the algorithm. Negative intensities, typical of peak splitting, are clearly present in the phenotypes of Clusters B and C, selected by the Citrus analysis performed on the flowVS transformed data and more present in the healthy control group. Cluster B might be identified as neutrophils based on the CD16<sup>bright</sup>CD62L<sup>bright</sup> phenotype, and seem to be CD56<sup>bright</sup>, which is probably due to peak splitting. Cluster C has a CD3<sup>bright</sup>CD4<sup>bright</sup>CD62L<sup>bright</sup>CD16<sup>dim</sup> phenotype. Caused by splitting of the CD16<sup>low</sup> peak into a low and dim peak. A more realistic representation of discriminant cell clusters is offered by the Citrus analysis on the OTflow transformed data.



**Figure 13. Citrus analysis results on the diversely transformed HIV data.**

The histograms show the phenotype of the cells in the red clusters selected by the Citrus model against a background cluster (blue) containing the remaining cells. Red clusters are either cell populations that are found more in controls or in HIV patients. (A) A single cluster is selected as the most discriminant between controls and HIV patients when data are transformed with cofactor 150. The phenotype of this cluster is associated with CD3<sup>low</sup>CD4<sup>bright</sup>CD16<sup>bright</sup>CD62L<sup>bright</sup> cells; (B) Two discriminant clusters are identified after flowVS-defined cofactors. Cells in cluster B are characterized by CD16<sup>+</sup>CD62L<sup>+</sup>CD56<sup>+</sup> phenotype; cells in cluster C show CD16<sup>dim</sup>CD62L<sup>+</sup>CD3<sup>+</sup>CD4<sup>+</sup> phenotype; (C) Three discriminant clusters are identified when transforming the data with OTflow-defined cofactors.



Here Clusters D and E, which are both more abundant in the healthy-control group, are represented by neutrophils ( $CD16^{\text{bright}}CD62L^{\text{bright}}$ ) and  $CD3^{\text{bright}}CD4^{\text{bright}}CD62L^{\text{bright}}$  T cells, respectively. The latter phenotype we also found more abundant in healthy controls compared to patients in the OTflow transformed PCA results and by manual gating, as shown in Figure 8. Additionally, Citrus detected Cluster F which may be associated to CD8+ T cells, and these cells are more represented in HIV patients.

## Discussion

MFC data transformation, required before multivariate methods such as PCA, viSNE and flowSOM can be applied, is often performed by bi-exponential transformations, such as arcsinh and logicle. We demonstrate bi-exponential transformations can introduce artifacts in multivariate analysis results. An optimal display of the MFC data in both univariate and multivariate data representation greatly relies on the choice of the transformation parameter.

A cofactor of 150 for the arcsinh function, suggested by default in various methods and in Cytobank, rarely leads to proper representation of the cell populations in the data for all the markers. A disadvantage of the use of a single co-factor (of 150) consists in the fact that different markers may necessitate different cofactors because the fluorescence intensities depend on the dye used, on marker expression, and on the compensation used to correct for spill-over and background fluorescence. Notably, all of these factors may vary per channel. Visual inspection of the transformed data per channel (not performed here) would be highly subjective, time-consuming and inefficient, especially with large panels.

The novel OTflow algorithm enables an automated, data-driven and validated estimation of channel-specific optimal parameters for both arcsinh and logicle transformations. It avoids artifacts such as merging two phenotypically different cell populations into one population by condensing two separate peaks into one spiky peak. OTflow also accurately prevents artifacts at low intensities by disregarding cofactors which lead to over-dispersion of the leftmost peak by splitting it in two cell populations (i.e. 'peak splitting'). This key feature of the algorithm represents one of the innovative and essential steps of OTflow compared to other methods developed to estimate transformation parameters such as flowVS and logicle calculation.<sup>15,18</sup>

As a result, OTflow outperforms methods such as flowVS and default-cofactor arcsinh transformation in properly displaying cell subpopulations and subsequent multivariate analyses. We showed with various MFC datasets (Supplementary Material) that

unintentional peak splitting in one or more channels caused by flowVS defined cofactors or default cofactors is dramatically propagated to higher dimensions when applying multivariate analysis methods. For instance, in the HIV data, homogenous CD4<sup>bright</sup>CD56<sup>low</sup> and CD8<sup>bright</sup>CD56<sup>low</sup> T-cell populations showed differential expression in CD56 in the PCA space and they were over-fragmented in the viSNE map, because of peak splitting present in CD56, as confirmed by backgating. Suboptimal transformation greatly complicates multivariate analysis because the phenotypically identical populations are separated into multiple seemingly phenotypically different populations. This generates confusion and it is highly detrimental to subsequent immunological interpretation and conclusions. Also, we showed how the additional variability introduced by peak splitting may overshadow the more subtle but relevant biological information in the data. This was the case in PCA as well as flowSOM. In the flowSOM map peak splitting caused homogeneous cell populations to be divided over multiple nodes at the expense of 'real' phenotypes that were clustered together in one node. Citrus also recognized the negative peaks as most discriminant between the two groups considered.

We demonstrated that OTflow may aid also in finding the optimal value for the W parameter for logicle transformation (Supplementary Material). The calculation of the optimal W as proposed by Parks et al 15 may not always lead to proper transformation of all the channels. Also here artifacts may be introduced, because the W is highly dependent on the negative intensities in the data. Also because focusing only on the first population with the lowest expression, logicle transformation may not stabilize variance between all the other cell populations with higher expressions.

For a widespread application of OTflow, some assumptions considered in the algorithm's steps are here discussed. The algorithm works by identifying uni-dimensional peaks in the iteratively transformed channels. By using a very sensitive function, OTflow is able to detect also very small peaks. However some threshold in 'finding' peaks is imposed. This is done to circumvent the detection of too large numbers of peaks not necessarily having biological meaning but associated to noise. OTflow thus removes peaks containing less than 5 cells, corresponding to 0.5% of the subsample considered. Especially in the extreme ends of the fluorescence range these are likely to be associated to noise or machine artifacts and therefore not to be introduced in the variance stabilization step. In some studies, however, very small cell populations may be expected. In this case it may be advisable to reduce this limit to 0.01 or 0.02% to avoid loss of relevant information, with the drawback of reducing the computational power of the algorithm.

The number of parameters per cell measured continues to increase with the development of new flow cytometers.<sup>37</sup> Larger panels will have more spill over per channel and appropriate compensation becomes more difficult. Overcompensation may occur more



easily. Overcompensation may shift the mean of a peak into the negative range. In OTflow a peak with a negative mean is considered an artifact, as a result of peak splitting. Therefore a penalty is given and the parameter(s) resulting into a peak with a negative mean will not be used. It is no news that a proper compensation is important, for the quality of the dataset and for interpretability of the results, independent of the type of analysis method used. Also for optimal transformation overcompensation should be avoided.

## Conclusion

Transformation of MFC data is an essential step before applying any type of univariate and multidimensional data analysis method. Suboptimal transformation due to poorly defined parameters can highly complicate the analysis results by introducing artifacts in the data and thus lead to misleading interpretations. The algorithm OTflow, here introduced, is a robust method for automated data-driven bi-exponential transformation of flow cytometry data. Most importantly, OTflow prevents split peaks to arise in the data, thereby preventing directly the false interpretation of distinct cell populations and/or the indirect failure (due to domination) to observe phenotypically different cell subsets. The versatility of the algorithm allows its application to any type of MFC dataset for the estimation of parameters for both arcsinh and logicle transformation. Integration of the algorithm in MFC analysis methods will facilitate rapid optimized transformation of the data and minimize the risk of misinterpretation.

## References

1. Perfetto SP, Chattopadhyay PK, Roederer M. Innovation: Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol.* 2004;4(8):648-655. doi:10.1038/nri1416
2. Chattopadhyay P, Perfetto S, Gaylord B, et al. Toward 40+ Parameter Fluorescence Flow Cytometry. In: *XXIX Congress of the International Society for Advancement of Cytometry.* ; 2014.
3. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: Helping to make sense of high-dimensional immunology data. *Nat Rev Immunol.* 2016;16(7):449-462. doi:10.1038/nri.2016.56
4. Amir ED, Davis KL, Tadmor MD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol.* 2013;31(6):545-552. doi:10.1038/nbt.2594
5. Jansen JJ, & Hilvering B, van den Doel A, et al. FLOOD: FLOW cytometric Orthogonal Orientation for Diagnosis. *Chemom Intell Lab Syst.* 2016;151(December):126-135. doi:10.1016/j.chemolab.2015.12.001
6. Tinnevelt GH, Kokla M, Hilvering B, et al. Novel data analysis method for multicolour flow cytometry links variability of multiple markers on single cells to a clinical phenotype. *Sci Rep.* 2017;7(1):1-11. doi:10.1038/s41598-017-05714-1
7. Folcarelli R, van Staveren S, Bouman R, et al. Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm. *Sci Rep.* 2018;8(1):1-18. doi:10.1038/s41598-018-29367-w
8. Qiu P, Simonds EF, Bendall SC, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol.* 2011;29(10):886-893. doi:10.1038/nbt.1991
9. Van Gassen S, Callebaut B, Van Helden MJ, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytom Part A.* 2015;87(7):636-645. doi:10.1002/cyto.a.22625
10. Bruggner R V., Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci.* 2014. doi:10.1073/pnas.1408792111
11. Verwer B. *BD FACSDiVa Option (White Paper)*. Becton, Dickinson and Company; 2002. <http://www.bdbiosciences.com/ds/is/others/23-6579.pdf>.
12. Herzenberg LA, Tung J, Moore WA, Herzenberg LA, Parks DR. Interpreting flow cytometry data: A guide for the perplexed. *Nat Immunol.* 2006;7(7):681-685. doi:10.1038/ni0706-681
13. Roederer M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. *Cytometry.* 2001;45(3):194-205. <http://www.ncbi.nlm.nih.gov/pubmed/11746088>.
14. van den Berg RA, Hoefsloot H CJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics.* 2006. doi:10.1186/1471-2164-7-142

15. Parks DR, Roederer M, Moore WA. A new "logicle" display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. *Cytom Part A*. 2006. doi:10.1002/cyto.a.20258
16. Finak G, Perez J-M, Weng A, Gottardo R. Optimizing transformations for automated, high throughput analysis of flow cytometry data. *BMC Bioinformatics*. 2010;11(1):546. doi:10.1186/1471-2105-11-546
17. Kotecha N, Krutzik PO, Irish JM. Web-based analysis and publication of flow cytometry experiments. *Curr Protoc Cytom*. 2010. doi:10.1002/0471142956.cy1017s53
18. Azad A, Rajwa B, Pothén A. flowVS: Channel-specific variance stabilization in flow cytometry. *BMC Bioinformatics*. 2016;17(1):1-14. doi:10.1186/s12859-016-1083-9
19. Tibshirani R. Estimating transformations for regression via additivity and variance stabilization. *J Am Stat Assoc*. 1988. doi:10.1080/01621459.1988.10478610
20. Bartlett MS. The Square Root Transformation in Analysis of Variance. *Suppl to J R Stat Soc*. 1936;3(1):68-78. doi:10.2307/2983678
21. Ray S, Pyne S. A computational framework to emulate the human perspective in flow cytometric data analysis. *PLoS One*. 2012. doi:10.1371/journal.pone.0035693
22. Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ Lett*. 1980. doi:10.1016/0165-1765(80)90024-5
23. Jolliffe IT. *Principal Component Analysis. Second Edition.*; 2002. doi:10.2307/1270093
24. Box GEP, Cox DR. An analysis of transformations (with discussion). *J R Stat Soc B*. 1964. doi:10.2307/2287791
25. Bendall SC, Simonds EF, Qiu P, et al. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science (80-)*. 2011;332(6030):687-696. doi:10.1126/science.1198704
26. Moore WA, Parks DR. Update for the logicle data scale including operational code implementations. *Cytom Part A*. 2012;81 A(4):273-277. doi:10.1002/cyto.a.22030
27. Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. *Ann Stat*. 2010. doi:10.1214/10-AOS799
28. Jarque CM, Bera AK. A Test for Normality of Observations and Regression Residuals. *Int Stat Rev / Rev Int Stat*. 1987. doi:10.2307/1403192
29. Lugli E, Pinti M, Nasi M, et al. Subject classification obtained by cluster analysis and principal component analysis applied to flow cytometric data. *Cytom Part A*. 2007. doi:10.1002/cyto.a.20387
30. Costa ES, Pedreira CE, Barrena S, et al. Automated pattern-guided principal component analysis vs expert-based immunophenotypic classification of B-cell chronic lymphoproliferative disorders: A step forward in the standardization of clinical immunophenotyping. *Leukemia*. 2010;24(11):1927-1933. doi:10.1038/leu.2010.160
31. Tung JW, Parks DR, Moore WA, Herzenberg LA, Herzenberg LA. New approaches to fluorescence compensation and visualization of FACS data. *Clin Immunol*. 2004. doi:10.1016/j.clim.2003.11.016

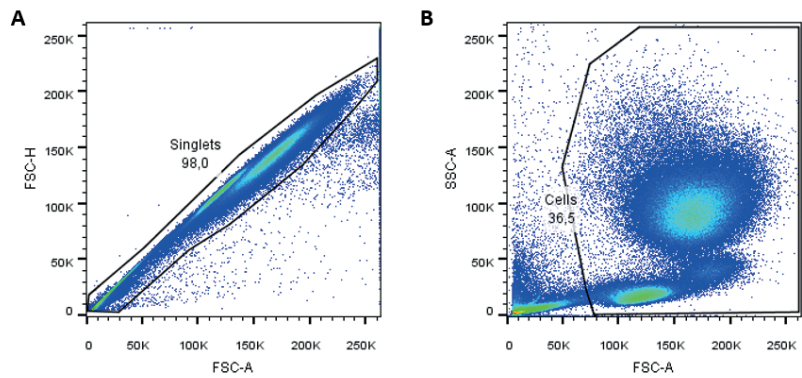
32. Bagwell CB. HyperLog - A flexible log-like transform for negative, zero, and positive valued data. *Cytom Part A*. 2005;64(1):34-42. doi:10.1002/cyto.a.20114
33. Lemster BH, Michel JJ, Montag DT, et al. Induction of CD56 and TCR-independent activation of T cells with aging. *J Immunol*. 2008;180(3):1979-1990. doi:10.4049/jimmunol.180.3.1979
34. Potter SJ, Lacabartz C, Lambotte O, et al. Preserved Central Memory and Activated Effector Memory CD4+ T-Cell Subsets in Human Immunodeficiency Virus Controllers: an ANRS EP36 Study. *J Virol*. 2007;81(24):13904-13915. doi:10.1128/JVI.01401-07
35. Vassena L, Giuliani E, Buonomini AR, Malagnino V, Andreoni M, Doria M. Brief report: L-selectin (CD62L) is downregulated on CD4+and CD8+T lymphocytes of HIV-1-infected individuals naive for ART. *J Acquir Immune Defic Syndr*. 2016;72(5):492-497. doi:10.1097/QAI.0000000000000999
36. Kononchik J, Ireland J, Zou Z, et al. HIV-1 targets L-selectin for adhesion and induces its shedding for viral release. *Nat Commun*. 2018;9(1):1-15. doi:10.1038/s41467-018-05197-2
37. Robinson JP, Roederer M. Flow cytometry strikes gold. *Science (80- )*. 2015;350(6262):739-740. doi:10.1126/science.aad6770

## Supplementary material

### Contents

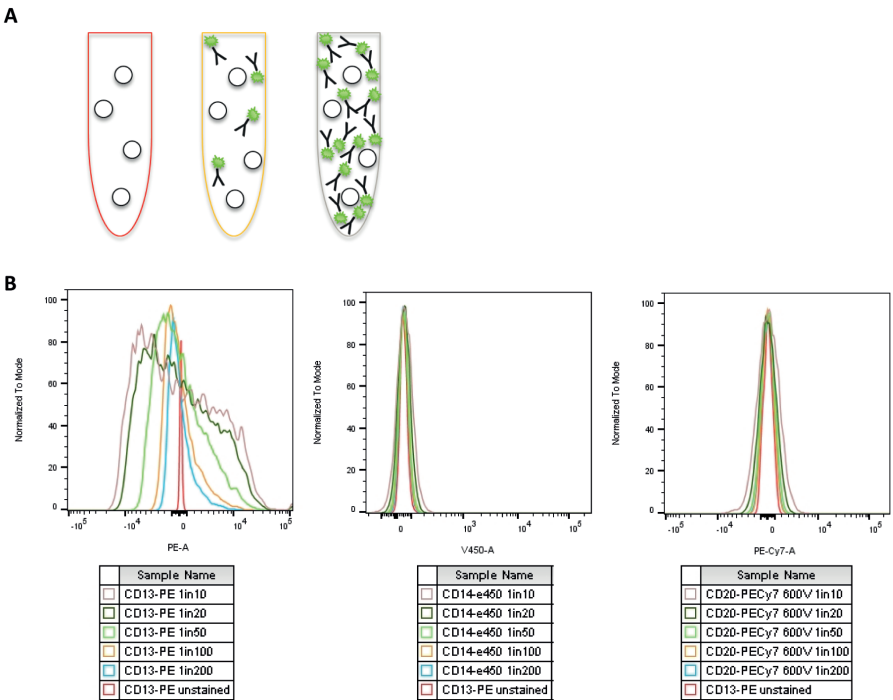
Supplementary Material I	122
Supplementary Information II	124
Background information about the MVA methods	124
PCA	124
viSNE	125
flowSOM	125
Citrus	125
Supplementary Material III	126
HIV analysis supplementary material	126
OTflow scores plot explained	128
Additional datasets	129
Lean vs Obese data	129
Tuberculosis data	129
Lean vs Obese – arcsinh transformation	130
Principal Component Analysis applied to arcsinh transformed lean vs obese data	133
Tuberculosis data – arcsinh transformation	137
viSNE applied to arcsinh transformed TBC data	139
Supplementary Material IV	144
Logicle transformation	144
References	146

Supplementary Material I



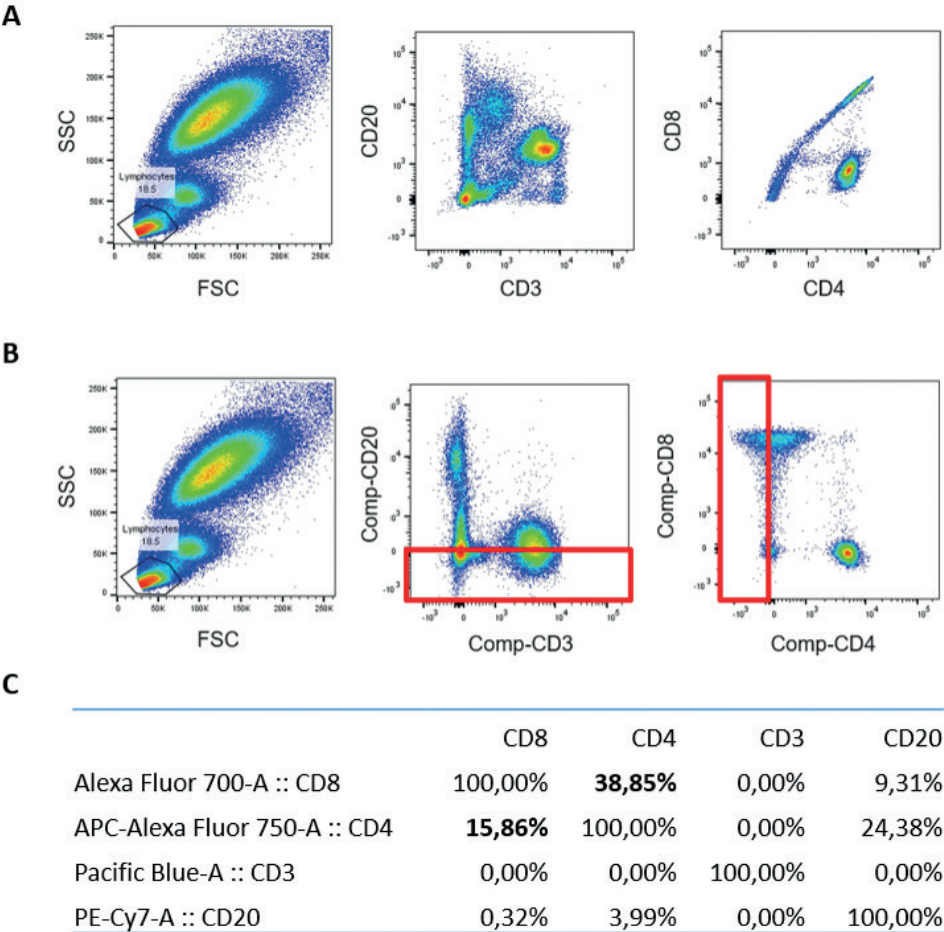
**Figure S1. gating single cells from peripheral blood samples of the HIV dataset.**

A) Single cells were manually gated by first selecting singlets based on FSC-Area and FSC-Hight. B) Subsequently dead cells and debris (with low FSC and SSC) were excluded and cells were gated in a FSC/ SSC plot.



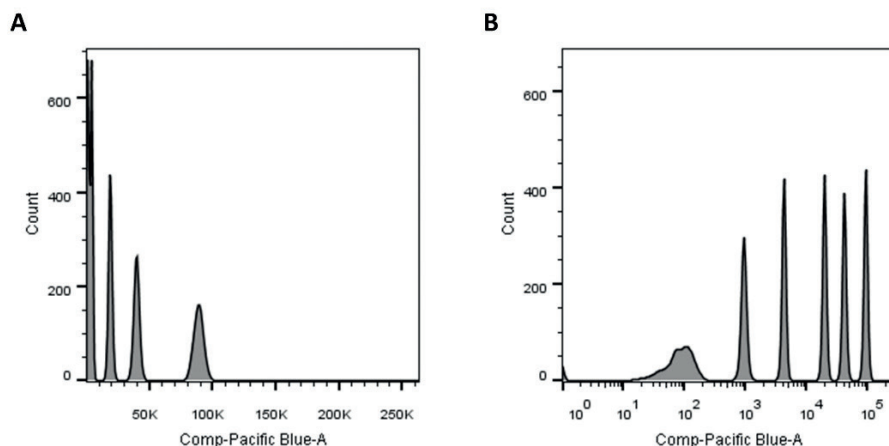
**Figure S2. Background subtraction resulting in negative intensities**

A) Schematic representation of titration samples containing various concentrations of unbound fluorescently labelled antibodies and stable concentrations of non-fluorescent beads. B) Fluorescent intensities of the samples containing non-fluorescent beads and different concentrations of unbound fluorescently labelled antibodies in the solution.



**Figure S3. Compensation leading to negative values**

A+B) FSC/SSC of leukocytes from a peripheral blood sample (pre-gated on Single Cells and debris excluded); lymphocytes were gated. Centre panels show CD3/CD20 biplots of the gated lymphocytes. Of the CD3 positive cells the expression of CD4 and CD8 was examined in the last panel. A) shows the uncompensated data, while B) shows the compensated data after using the compensation matrix shown in C). The boxes in B contain the cells with negative intensities for CD20 and CD4 emerging due to compensation.



**Figure S4. Populations with low fluorescent intensities are better visible on a log scale**

BD Cyto-Cal beads with 6 different fluorescent intensities for Pacific Blue were measured on a BD Fortessa flow cytometer and the Pacific Blue intensities were plotted on a linear scale (A) or a log scale (B). Using the linear scale only 4 types of beads are visible; on the log scale all 6 types of beads with various Pacific Blue intensities are visualised.

## Supplementary Information II

### Background information about the MVA methods

#### PCA

Principal Component Analysis <sup>1</sup> (PCA) is a dimensionality reduction technique which is able to condense the information of the original data into a lower-dimensional space by retaining relevant variability and disregarding minor variance (such as noise). The resulting PCA space generally consists of two or three dimensions, called principal components (PCs). The first PC always explains most of the original variance. The property of mutual orthogonality between the components implies that the successive PCs explain the highest percentage of remaining variance. The outcome of a PCA is constituted by scores and loadings. The scores are the new coordinates of the data in the built PCA space, while the loadings represent the contribution of each feature (original dimension) to the PCs. A PCA scores plot shows the distribution of all cells describing them in terms of marker co-expression. The loadings, generally represented as arrows in PCA plots, indicate the orientation of the markers in the PCA space.

Centering is generally needed to remove marker expression (or offsets) consistently present across all the samples. As a result, the model focuses on deviation from the mean



rather than on global mean by creating a common point of reference to quantify variation in cellular marker expression between the cells. Scaling is then applied to assure that the markers equally contribute to the model regardless of the intensity of the label used.

### **viSNE**

Another dimensionality reduction technique is viSNE.<sup>2</sup> The method employs the non-linear t-Stochastic Neighbor Embedding (t-SNE) to reduce dimensionality and thus enables visualization of MFC data in a bi-axial viSNE map.

### **flowSOM**

In flowSOM MFC data are clustered by using a Self-Organizing Map algorithm.<sup>3</sup> The number of nodes is defined prior to the analysis. Higher numbers of nodes correspond to higher cluster 'purities' but might also lead to over-clustering and thus spreading of similar cell populations over multiple nodes. When applying flowSOM we adopted the default value of 100 nodes. Nodes connected to each other are most similar, based on their multidimensional marker expressions. Information about the marker expression can be visualized through pie chart in each node representing the average intensities of the markers for cells assigned to the same node.

### **Citrus**

In the case of intergroup analysis, discrimination between two groups can be performed with the Citrus algorithm.<sup>4</sup> Citrus applies a clustering algorithm (hierarchical clustering) to extract information (features) about similar cell populations clustered together. For intergroup analysis, group belonging is defined prior the analysis. Citrus then employs a supervised regularized classification model<sup>5</sup> to identify the descriptive cell cluster-specific features which best predict the group any of the used MFC samples belongs to. These features may include, for example, the proportion of cells in the detected clusters which is more prominent in one of the two groups. The fit of each discriminant model provided by Citrus is estimated through (double) cross-validation. An optimal model has a low cross-validation error rate, which corresponds to a low percentage of misclassified samples. When the error remains constant within a range of increasing number of features, the lowest number of features is chosen by the algorithm in order to select the most informative features that differ between the two groups.

Supplementary Material III

HIV analysis supplementary material

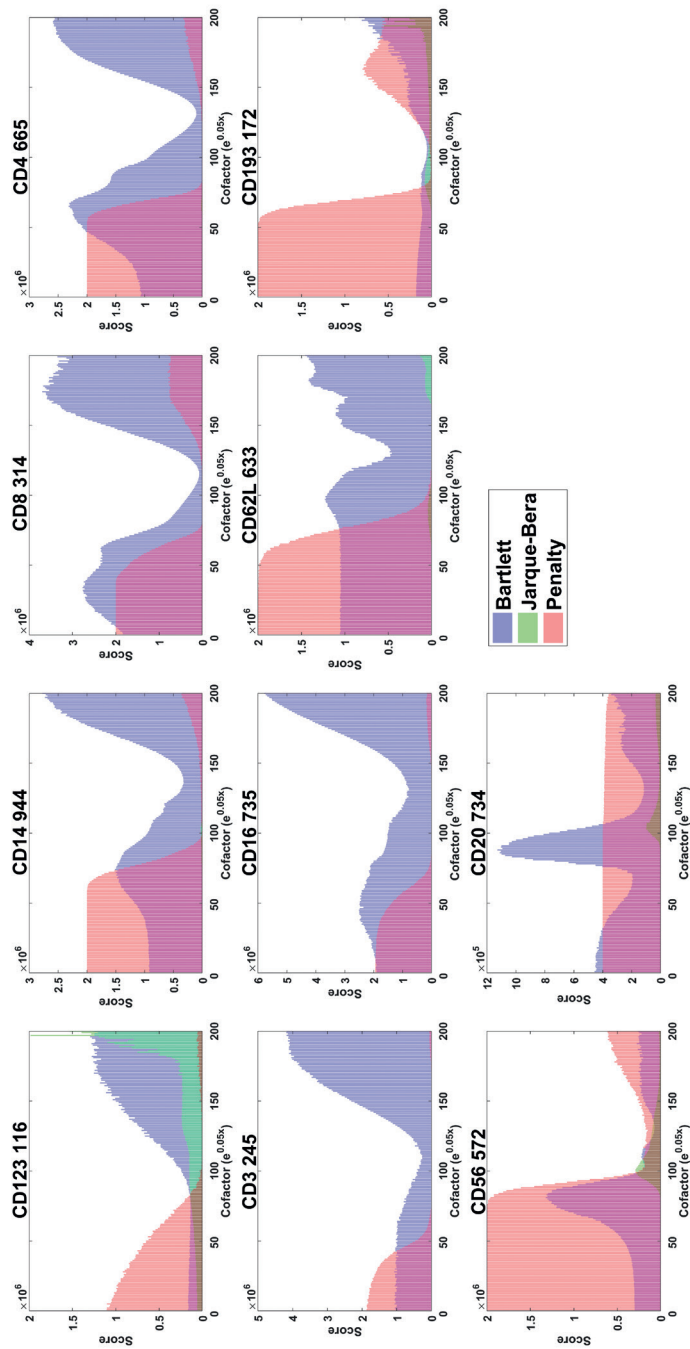
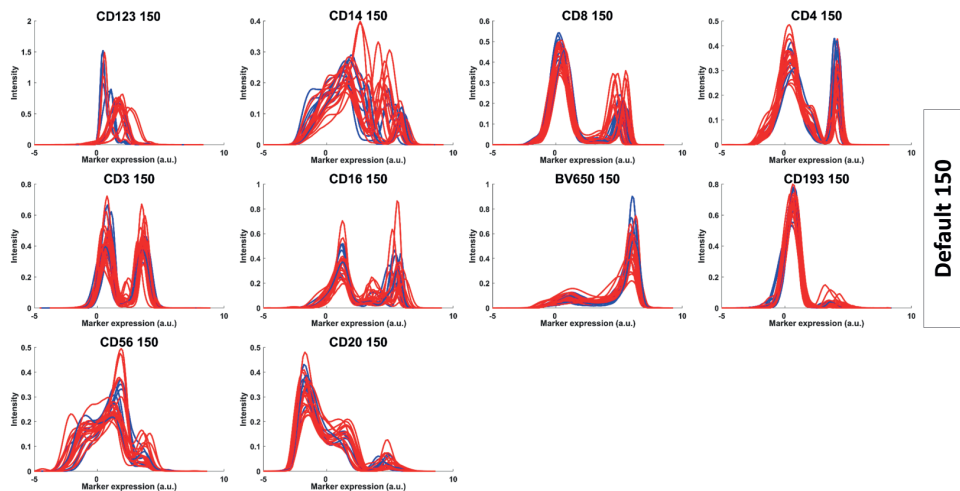


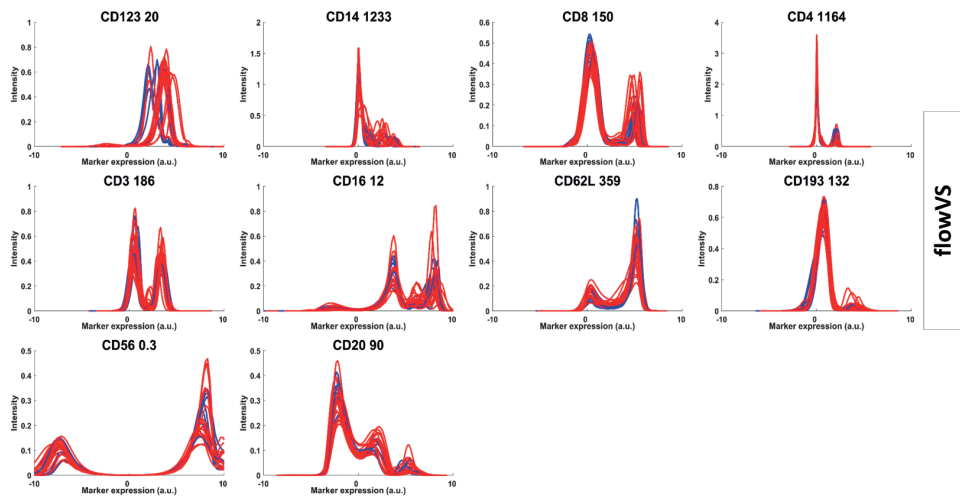
Figure S5. Otfow score calculated per marker on the HIV data.

The graphs show the contribution of Bartlett (blue), Jarque-Bera (green) and Penalty (red) to the average score.

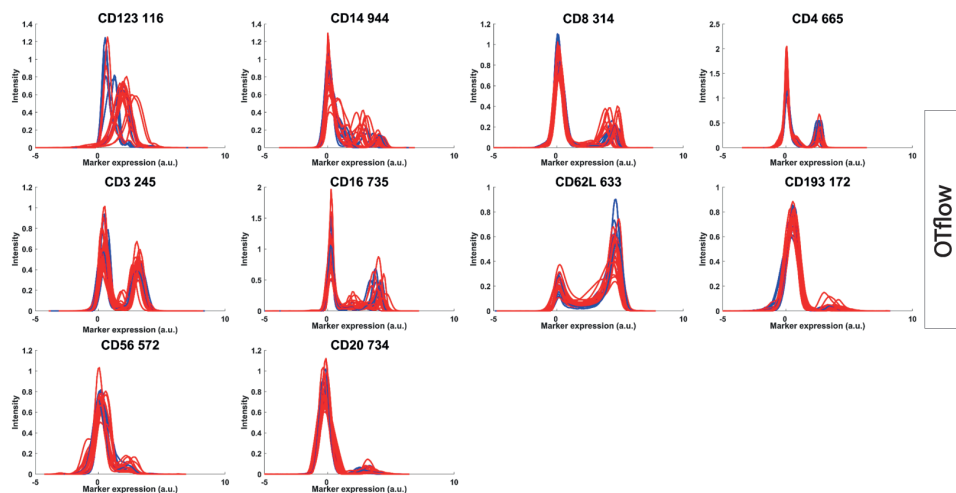
A



B



C



**Figure S6. Single marker expression histograms of arcsinh-transformed HIV data. Control individuals are depicted as blue, while HIV patients are shown in red.**

A) Data transformed by using default cofactor 150; B) Data transformed by using flowVS-defined cofactors; C) Data transformed with cofactors estimated by OTflow algorithm.

## OTflow scores plot explained

Penalty contribution is predominant when the dataset is transformed using low cofactors, as these provoke splitting of the leftmost peak around zero. For all markers Jarque-Bera statistics is quite insignificant throughout the entire range of tested cofactors ( $e0 - e10$ ), compared with the Bartlett statistics. From this we can deduce that per marker two or more peaks are identified by OTflow. This was expected, as the HIV dataset contains multiple cell populations which are defined by either high or low expression levels of various markers present in the dataset, leading to a dim and bright population (2 peaks) per maker. Only for very high cofactors, Jarque-Bera is estimated, indicating the presence of one single peak. For higher cofactors the linearly transformed region broadens and, as result, in some samples multiple cell populations can be squeezed into a single peak.

## Negative intensities per dataset

**Table S1. Overview of the three datasets used in this study.**

Dataset	Flow cytometer	Number of channels used	% of cells with $\geq 1$ negative intensities	Minimum negative value	Maximum negative value
HIV	BD LSRFortessa	10	96%	-104370	-0.00024
Obese	BD FACS-Canto II	7	39%	-4632	-0.00033
Tuberculosis	BD LSRFortessa	10	54%	-5745	-0.0000028

Percentage of cells having negative intensities in one or more channels is shown. For each dataset, the range of the negative intensities is indicated by the minimum and maximum negative value.

## Additional datasets

Two additional datasets were used to show the effect of suboptimal transformations by using selected MVA methods. Lean vs Obese <sup>6</sup> and Tuberculosis studies were chosen as complementary datasets consisting of multiple and single (gated) cell populations, respectively.

### Lean vs Obese data

The obesity study comprises 29 individuals divided in 16 obese and 13 lean individuals based on the value of Body Mass Index, as described in <sup>6</sup>. Seven surface markers were measured on peripheral blood cells of both classes of individuals: CXCR1, CD14, CD56, CD11c, CD16, HLA-DR, CD3/CD19/CD66b. Cells expressing CD3+, CD19+ or CD66b+ (T lymphocytes, B lymphocytes and neutrophils) were excluded by gating. Samples size ranged between 2,000 and 24,000, after gating. Previous research showed an increase in Natural Killer cells (CD16<sup>bright</sup>CD56<sup>bright</sup>), an increase in CD11b expression on monocytes (HLA-DR<sup>bright</sup>) and an increase in non-classical monocytes (CD14<sup>dim</sup>CD16<sup>bright</sup>) associated with obese individuals.

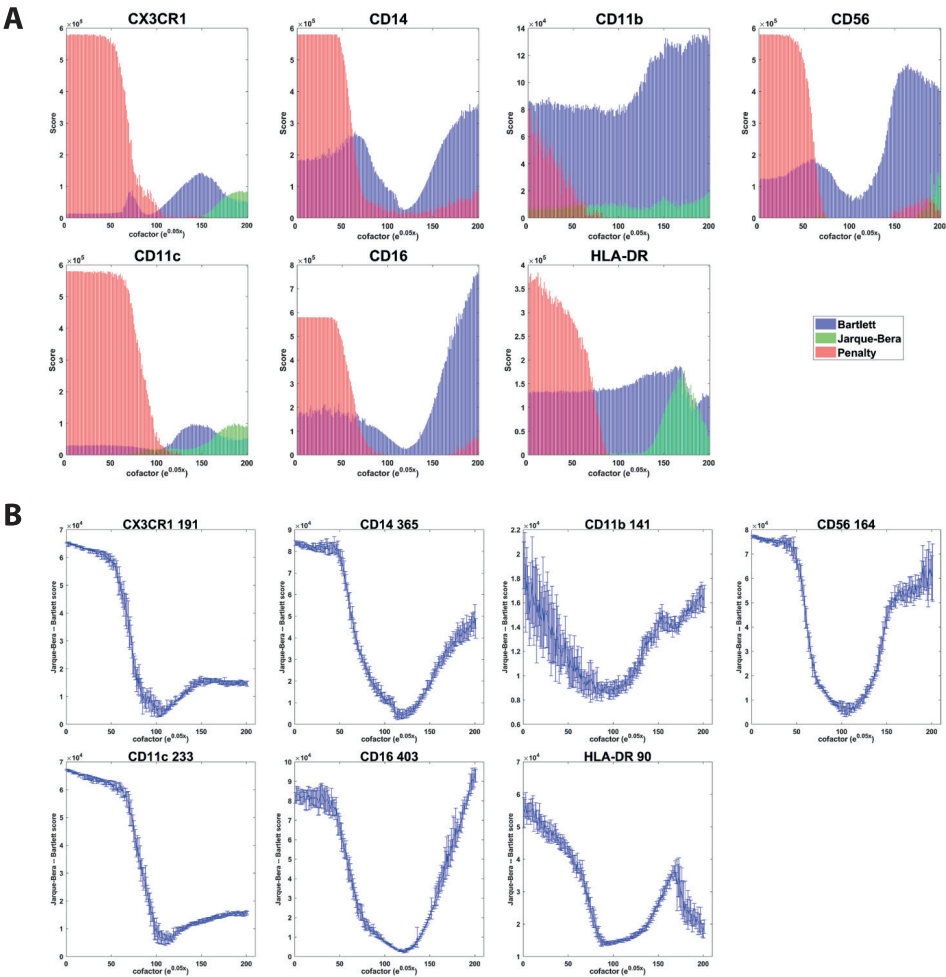
### Tuberculosis data

The tuberculosis dataset comprises 26 peripheral blood samples, of which 13 were obtained from healthy controls and 13 from Tuberculosis patients. Blood samples were taken from the antecubital vein and anticoagulated with sodium heparin. After erythrocyte lysis with ice-cold lysis buffer consisting of 155 mM NH<sub>4</sub>Cl and 10 mM Na<sub>2</sub>EDTA, the cells were incubated with a 10-color flow cytometry panel containing the markers CD16, CD62L, CD35, CD11b, CD66b, CBRM1/5, CD64, LAIR1, CD49d, CD14 for 30 min. Thereafter cells were fixated by incubation with 1% PFA for 30 min. The samples were measured on

a BD LSR-II Fortessa flow cytometer (Becton Dickinson, Mountain View, CA). In Flowjo analysis software (Tree Star Inc., Ashland, Oregon) the granulocytes were gated based on FSC/SSC and eosinophils were excluded based on their CD64<sup>bright</sup>CD49d<sup>bright</sup> expression, leaving only the neutrophils. Only cells within the neutrophil gate were exported in fcs files and used for subsequent transformation and data analysis.

**Lean vs Obese – arcsinh transformation**

An optimal arcsinh transformation cofactor was estimated for each of the 7 fluorescence channels in the lean vs obese dataset using the OTflow algorithm. Contributions of the different statistics and penalty to the average final OTflow score are shown in Figure S7A. Only for markers CD14 and CD16 no Jarque-Bera was calculated, meaning that even when high co-factors are applied 2 or more peaks are present for these markers. The final average score calculated by OTflow is displayed in Figure S7b. Figure S8 shows how the expression of a representative marker (CD11c) varies when the channel is transformed with increasing cofactor values. The two lower cofactor values cause peak splitting. Peak splitting does not occur when transforming the data from this marker with cofactor 150. However, using cofactor 150 results in a higher between-peak variance compared to the optimal cofactor value of 233, corresponding to the minimum of the score curve (Figure S7b). Optimally transformed data show 2 peaks with approximately equal variance. These CD11c<sup>high</sup> subpopulation may be identified as monocytes, while all other cell populations in the sample are CD11c<sup>low</sup>. High cofactors (2500 and 5000) provide a wrong representation of the CD11c populations: the peaks become highly compressed (cofactor 2500) and merge into a single peak with the highest cofactor (cofactor 5000).



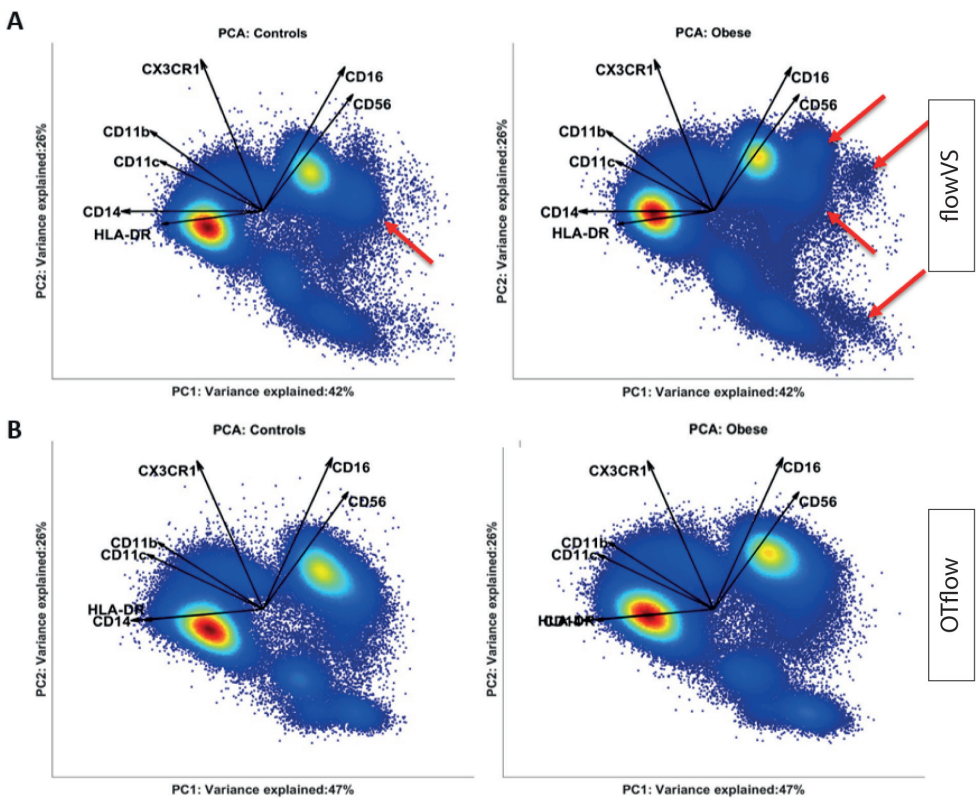
**Figure S7.**

A) Graphs of the contribution of Bartlett, Jarque-Bera and Penalty to the average score, calculated per marker. B) Graphs of the average total score and standard deviation per measured marker for cofactors ranging between  $e^0$  and  $e^{10}$ . Optimal cofactors correspond to the lowest scores and their values are shown in the title of each plot.

The optimal cofactors found by OTflow and by flowVS algorithm are shown in Table S2.

**Table S2. Cofactors for arcsinh transformation calculated per marker by the flowVS and OTflow algorithms on the lean vs obese dataset.**

	CX3CR1	CD14	CD11b	CD56	CD11c	CD16	HLA-DR
flowVS	73	343	54	142	3	481	0.4
OTflow	191	365	141	164	233	403	90



**Figure S9. Single marker expression histograms of arcsinh-transformed lean vs obese data.**

A) Data transformed by using default cofactor 150. No peak splitting is present; B) Data transformed by using flowVS-defined cofactors. Peak splitting caused by low value for the cofactor are marked by a green box; C) Data transformed with cofactors estimated by the OTflow algorithm. No peak splitting is observed around zero.



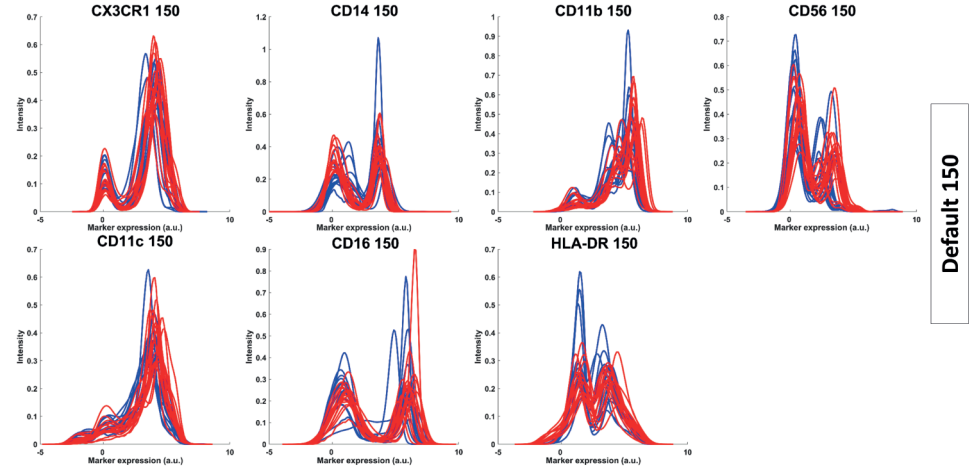
Significantly lower cofactors for marker CD11c and HLA-DR are calculated by flowVS compared to the results of OTflow. These, however, cause peak splitting with concomitant emergence of an extreme negative peak for these two markers, as shown in Figure S9. The optimized transformation by OTflow leads to similar results as the transformation with default cofactor 150 in terms of absence of spurious splits of the leftmost peaks. However the transformation using OTflow-defined cofactor might be preferred as it stabilizes better the variance between the peaks for all the fluorescence channels.

### **Principal Component Analysis applied to arcsinh transformed lean vs obese data**

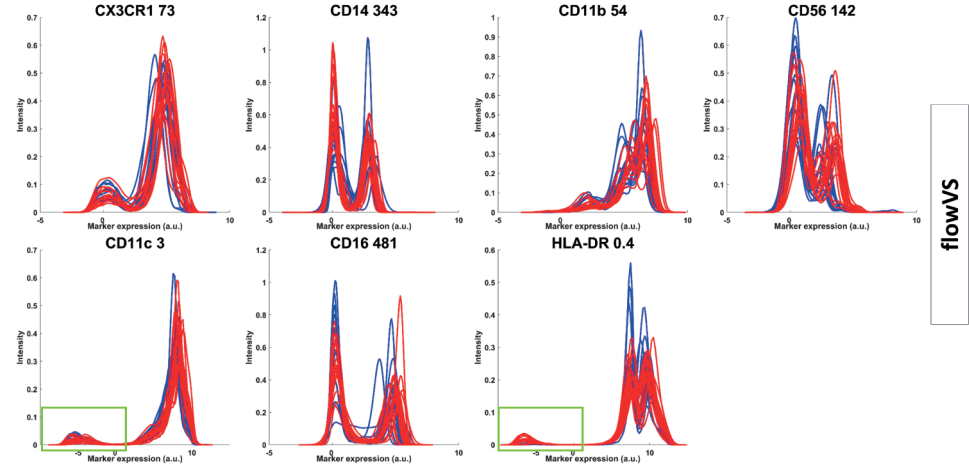
Score plots of the PCA analyses performed on the flowVS and OTflow transformed data appear substantially different (Figure S10). Various additional 'populations' are present when transforming the data with the flowVS-defined cofactors, versus the transformation based on OTflow (Figure S10, arrows). Strikingly, these additional subsets appear to be more predominant in the obese samples than in the samples of the lean (control) individuals (4 densely populated subsets vs 1 densely populated subset) and this would suggest greater biological heterogeneity in the peripheral blood of obese individuals compared to the peripheral blood of the control group.

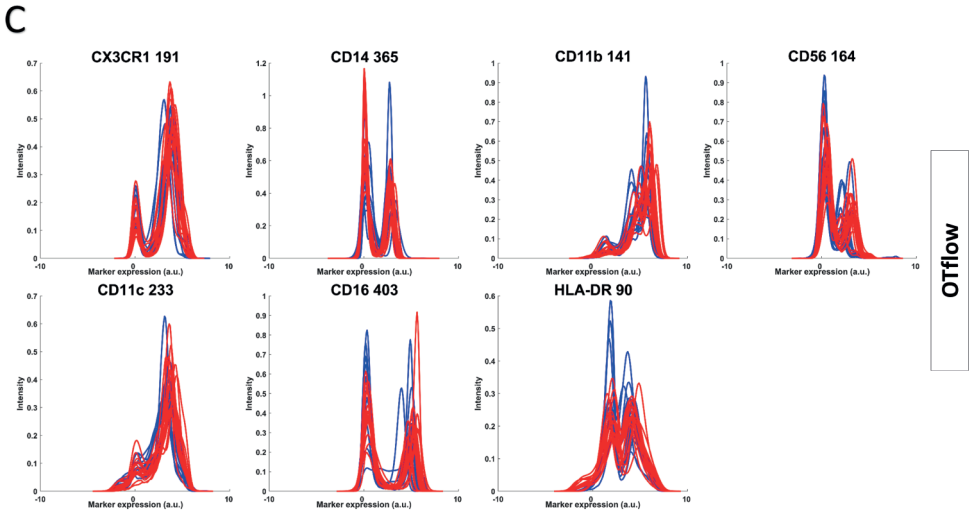
Interested regions in the PCA space were backgated to investigate the cellular marker expressions of the corresponding subpopulations (Figure S11). The light blue and dark blue populations (Figure S11A+B) are present in both PCA models (Figure S10). In the flowVS transformed dataset, the light green population emerged from the light blue cell population; the red, orange and dark green cell populations emerged from the dark blue cell subset. Due to peak splitting these cell populations could be interpreted as low CD11c and/or HLA-DR expressing cells, and the cell populations they emerged from will then be interpreted highly expressing those markers (Figure S11C+D), while actually the expressions of the additional cell populations are similar to the expressions of the original (blue gated) cell populations. The four additional cell populations are purely separated based on their negative intensities. These cell populations represent the negative region of the Gaussian curve at fluorescence intensity 0.

A



B

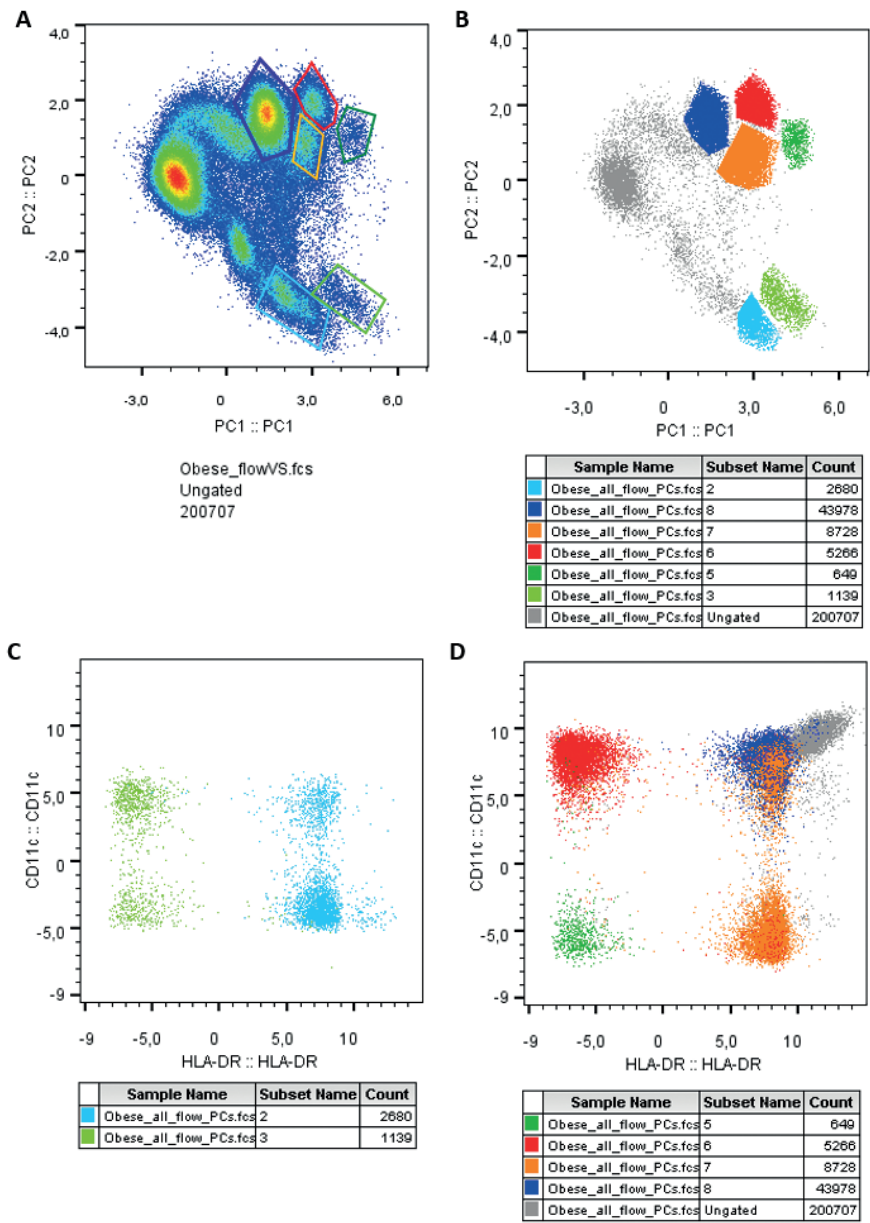




**Figure S10.**

Density representation of PCA analysis scores. The loadings are plotted as arrows in the PCA space: their length indicates the contribution of each marker to the cell-to-cell variability; their mutual direction offer indication of co-expression among the markers. A) Data are transformed with OTflow cofactors and pre-processed by mean-centering and scaling over all the samples prior the PCA analysis. The red arrows highlight additional cell populations which were created due to suboptimal transformation. B) Data are transformed with flowVS cofactors and pre-processed by mean-centering and scaling over all the samples prior the PCA analysis.

This shows that transformations with suboptimal cofactors, as defined by flowVS, causing peak splitting mainly in makers CD11c and HLA-DR, may have a great impact on multivariate analysis results of dimensionality reduction methods such as PCA and this could lead to immunological misinterpretations. The comparison of the loadings in the OTflow and flowVS plots suggests that the additional clusters (subpopulations, artifacts) in the flowVS plots are related to HLA-DR, as this loading changes most going from OTflow to flowVS.



**Figure S11.**

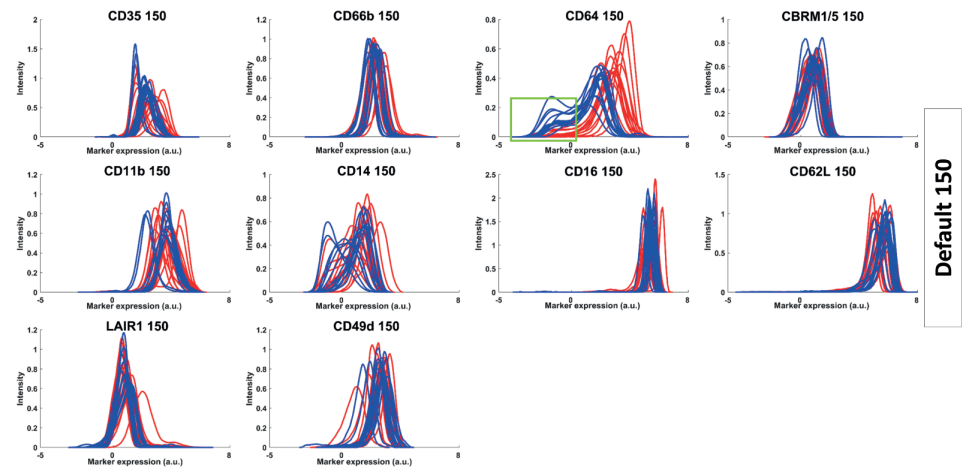
A) gating of populations in the PCA space; B) overlay of gated populations in the PCA space. C+D HLA-DR vs. CD11c biplots of gated populations. C) the green population which emerged from the blue population is separated based on its negative intensities for HLA-DR. Both populations contain negative and positive intensities for CD11c, which are also separated from each other due to peak splitting. D) The orange, red and green population which emerged from the dark blue population have negative intensities for HLA-DR and/or CD11c. The grey cells are the ungated cells, which are doubly positive for HLA-DR and CD11c.

## Tuberculosis data – arcsinh transformation

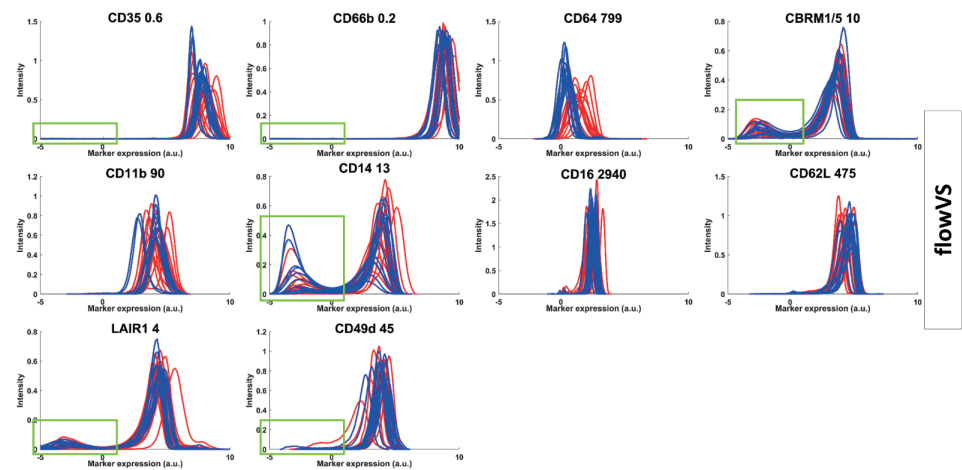
The Tuberculosis (TBC) dataset was used to further test the OTflow algorithm against the results of the flowVS-based cofactors and default cofactor 150 for the arcsinh transformation. The TBC dataset consists of peripheral blood samples from 13 healthy controls and 13 tuberculosis patients. Peripheral blood samples were stained with 10 markers, mostly neutrophil degranulation or activation markers. To focus on neutrophil phenotypes in TBC patients vs controls, only the peripheral blood neutrophils were gated and exported as fcs files.

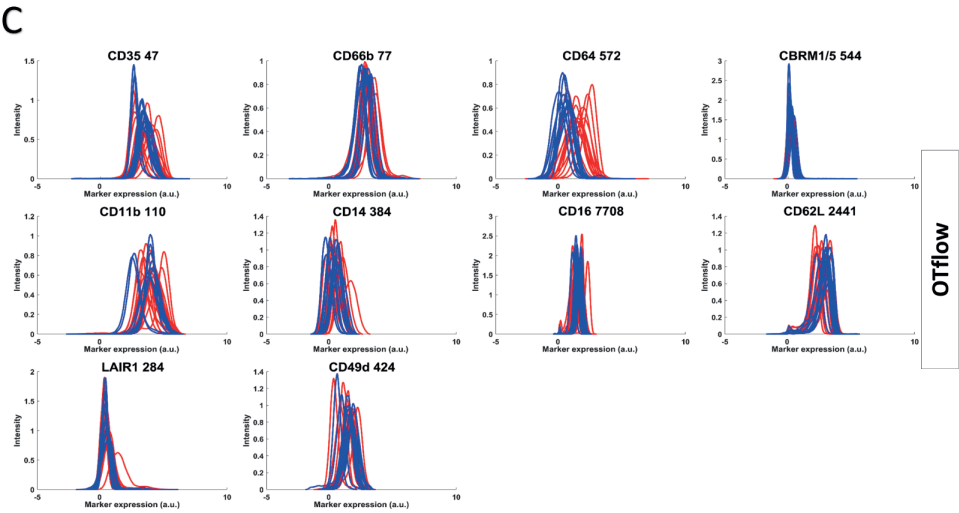
When using default cofactor 150, the single CD64 peak is actually split in two peaks which would be interpreted as differential expression of the marker (low and bright). This peak splitting seemed to be more predominant in the control individuals compared to the TBC patients as shown in Figure S12A (blue for control histograms, red for TBC patients histograms). Homogeneous low expression levels of CD64 on neutrophils are expected in control individuals, while the expression of CD64 is upregulated on neutrophils after TBC infection. Also for markers CD16 and CD62L the default cofactor 150 was too small. In both channels a bright population is visible, while the low populations are flattened and greatly extend into the negative region. A higher cofactor solved the problem of peak splitting for CD64 as shown by the single-marker histograms of the data transformed with flowVS and OTflow defined cofactors (Figure S12B+C). A single CD64<sup>low</sup> population is present for both transformations with the single peaks resembling a more normal distribution in the case of the OTflow-based transformation. The flowVS algorithm failed in optimally transforming the intensities of various markers. The estimated cofactors led to peak splitting of the markers with low intensities such as CBRM1/5, CD14 and LAIR-1, as shown in Figure S9B. Low cofactors were calculated also for markers CD35, CD66b and CD49d resulting in a flattened population with low intensities spread in the negative range (Figure S9B). No peak splitting artifacts were present when transforming the TBC data with OTflow (Figure S12C). The novel algorithm, in fact, provided an optimal transformation of the intensities from the 10 fluorescence channels with small low populations correctly represented around zero (as for CD16 and CD62L) and peaks having a normal distribution.

A



B





**Figure S12. Single-marker expression histograms of arcsinh-transformed Tuberculosis (TBC) data. Control individuals are depicted as blue, while TBC patients are shown in red.**

A) Data transformed by using default cofactor 150. Peak splitting caused by a too low cofactor are marked in a green box; B) Data transformed by using flowVS-defined cofactors. Peak splitting caused by a too low cofactor are marked in a green box; C) Data transformed with cofactors estimated by OTflow algorithm. No peak splitting is present; the peaks resemble normal distributions.

Table S3 reports the values for the cofactors found by flowVS and OTflow algorithm for a direct comparison.

**Table S3. Cofactors for arcsinh transformation calculated per marker by the flowVS and OTflow algorithms on the tuberculosis dataset.**

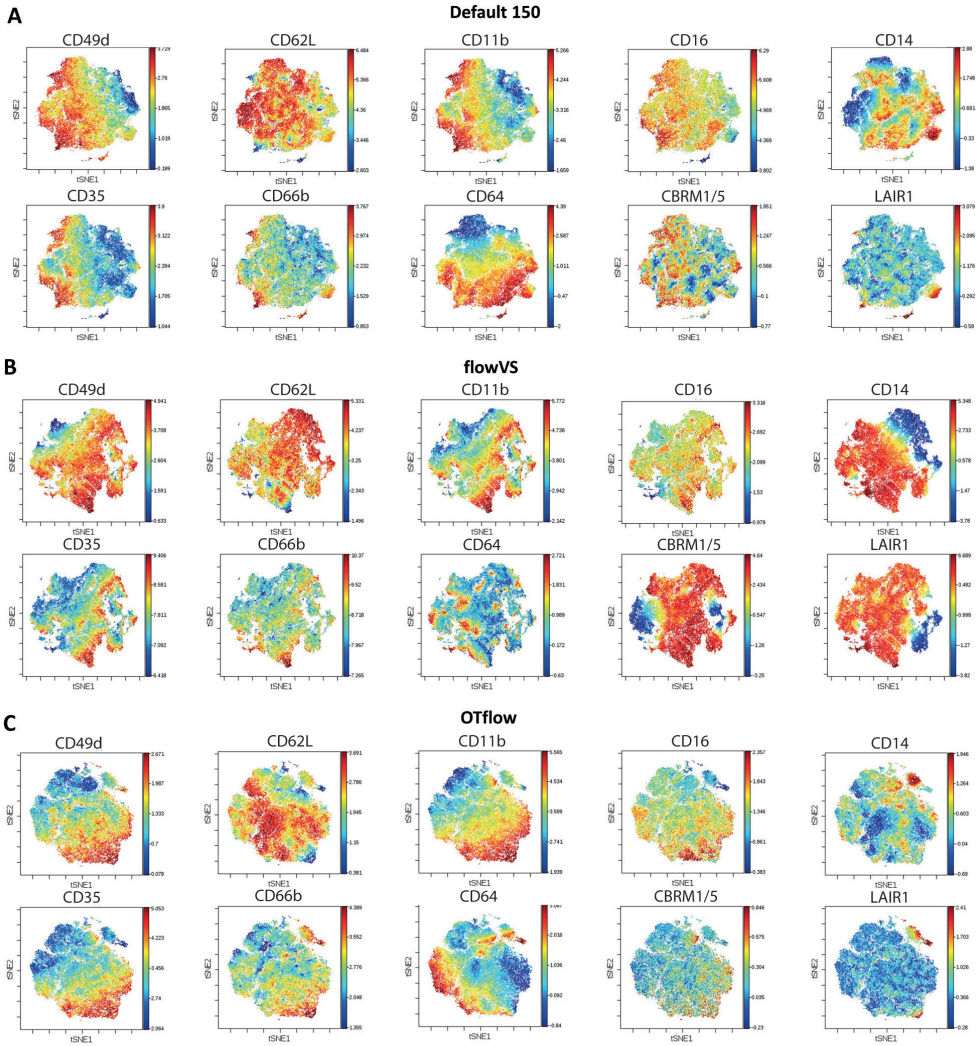
	CD35	CD66b	CD64	CBRM1/5	CD11b	CD14	CD16	CD62L	LAIR1	CD49d
flowVS	0.61	0.20	799	10	90	13	2940	475	4	45
OTflow	47	77	572	544	1010	384	7708	2441	284	424

**viSNE applied to arcsinh transformed TBC data**

We investigated the effect of suboptimally transformed data in MVA results by performing viSNE analyses on the differently arcsinh-transformed TBC data. The resulting viSNE maps plotting both patient and control neutrophils, colored based on marker expression levels, are shown in Figure S13. The most striking differences between the differently transformed viSNE maps are observed for the CD14, CBRM1/5 and LAIR-1 expression levels. In the default cofactor transformed results more than half of neutrophils seem to have high CBRM1/5 expression levels. These CBRM1/5<sup>bright</sup> cells are distributed over the

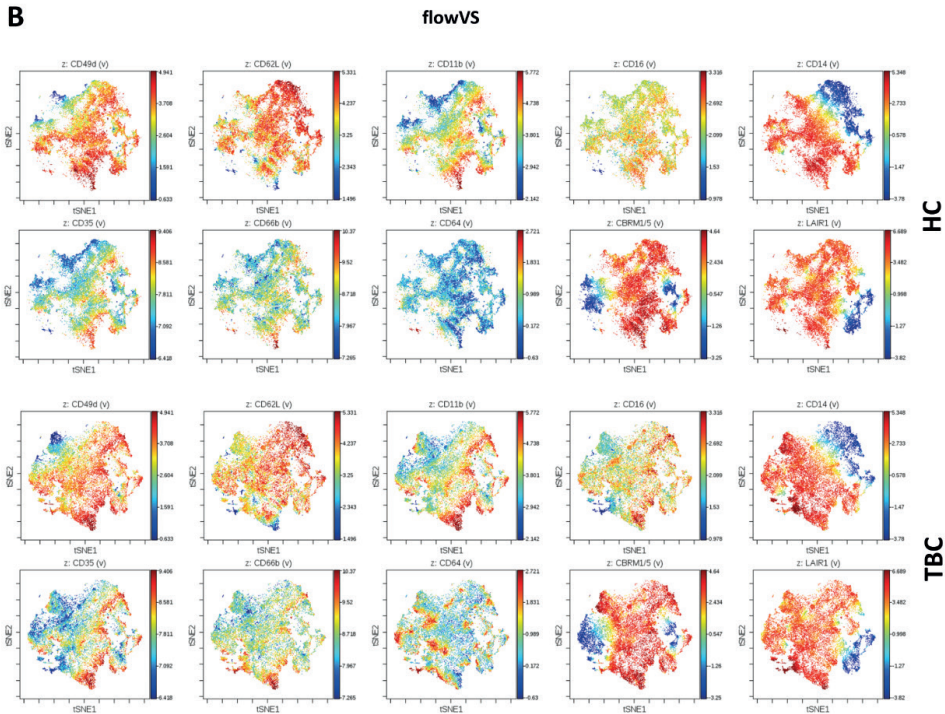
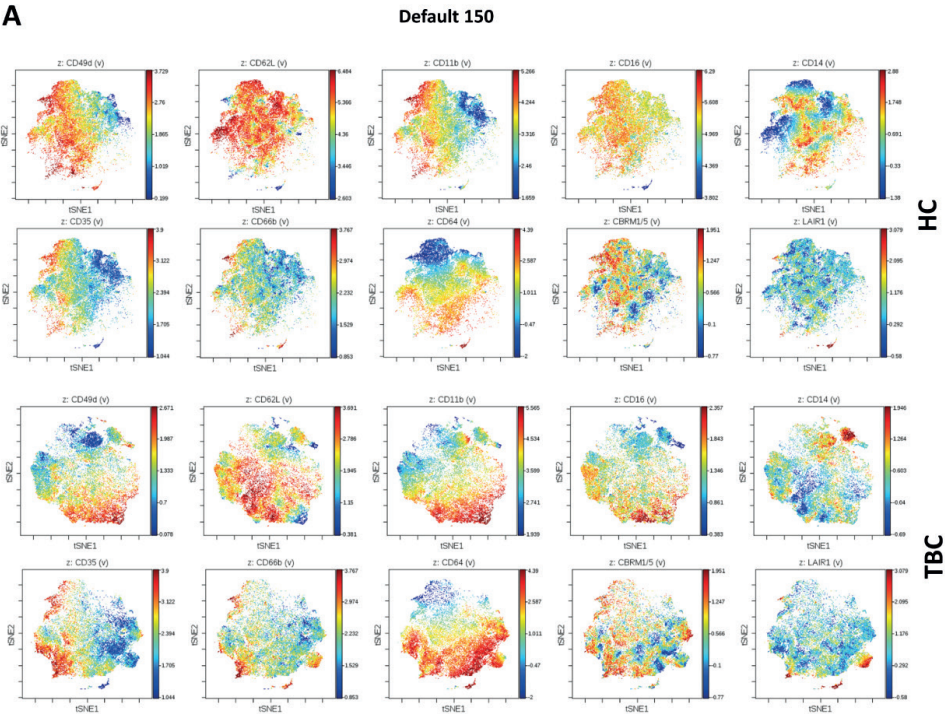
whole map in small clusters. In the flowVS transformed viSNE map, even more cells with bright CBRM1/5 intensities were found. Similarly low to predominantly high expression patterns were found for LAIR-1 and CD14 in the flowVS transformed data. This could very well be caused by the peak splitting of the homogeneous low expressing population into 2 separate populations as observed in Figure S12B for CD14, CBRM1/5 and LAIR1. In the OTflow transformed viSNE map the vast majority of the cells have low or dim CBRM1/5 expression levels, the same applies for LAIR-1 and CD14. The variance in LAIR-1, CBRM1/5 and CD14 expression in the flowVS transformed dataset influences the distribution of the cells greatly, since highly expressing cells are clustered together and low expressing cells are also clustered together on a different position in the map. Interestingly, the CD64<sup>bright</sup> cells are dispersed over the flowVS transformed viSNE map, while in the OTflow and default transformed viSNE maps these cells are clustered together. The viSNE maps of controls vs. patients are shown in Figure S13. Strikingly, in the viSNE maps of the flowVS transformed data both healthy controls and patients have cells with high expression levels of LAIR-1 and also of CBRM1/5. This is unexpected, as peripheral blood neutrophils of healthy controls normally do not express LAIR-1 and CBRM1/5.<sup>7,8</sup> These seemingly high-expressing cells are most likely the cells in the “positive” peak which is formed due to peak splitting. Comparing the viSNE map of the control neutrophils to the viSNE map of the TBC patients neutrophils of the OTflow transformed data (Figure S14C), a small CD16<sup>low</sup>CD62L<sup>dim</sup>CD66b<sup>bright</sup>LAIR-1<sup>bright</sup> neutrophil population is present in the patients, but not in the controls. These cells could be identified as metamyelocytes, and might be recruited from the bone marrow during Mycoplasma-induced inflammation. In the default and flowVS transformed data this cell subset is not clearly visible in the viSNE map with the patient cells, which is probably due to differences in the expression range of LAIR-1 and other markers. This is an illustrative example showing how peak splitting causes both the gain of false information in a dataset and the loss of immunologically relevant information.

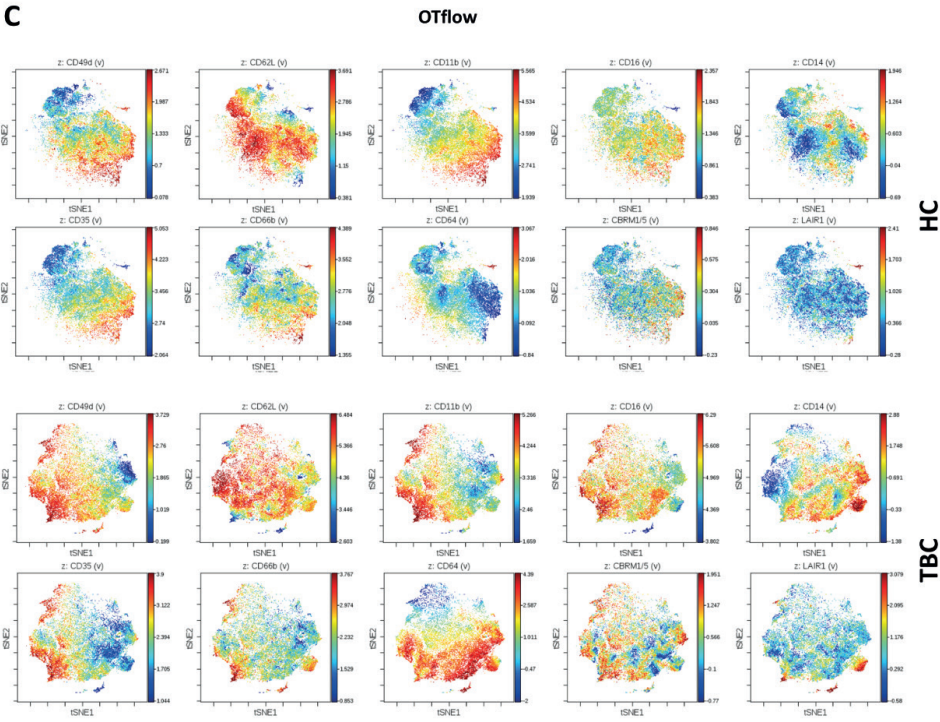




**Figure S13. The viSNE maps of the variously transformed TBC datasets differ greatly from each other.**

viSNE maps of default transformed cofactor 150 (A), flowVS transformed (B) and OTflow transformed (C) TBC datasets. The color coding of the cells represent the expression levels for the specific marker named in the title of the plots.





**Figure S14. Control vs. patients viSNE maps of differently transformed datasets.**

A) viSNE maps of default transformed (cofactor 150) of healthy controls (HC, upper panels) and TBC patients (lower panels), B) flowVS transformed and C) OTflow transformed TBC datasets. The color coding of the cells represent the expression levels for the specific marker named in the title of the plots.

Supplementary Material IV

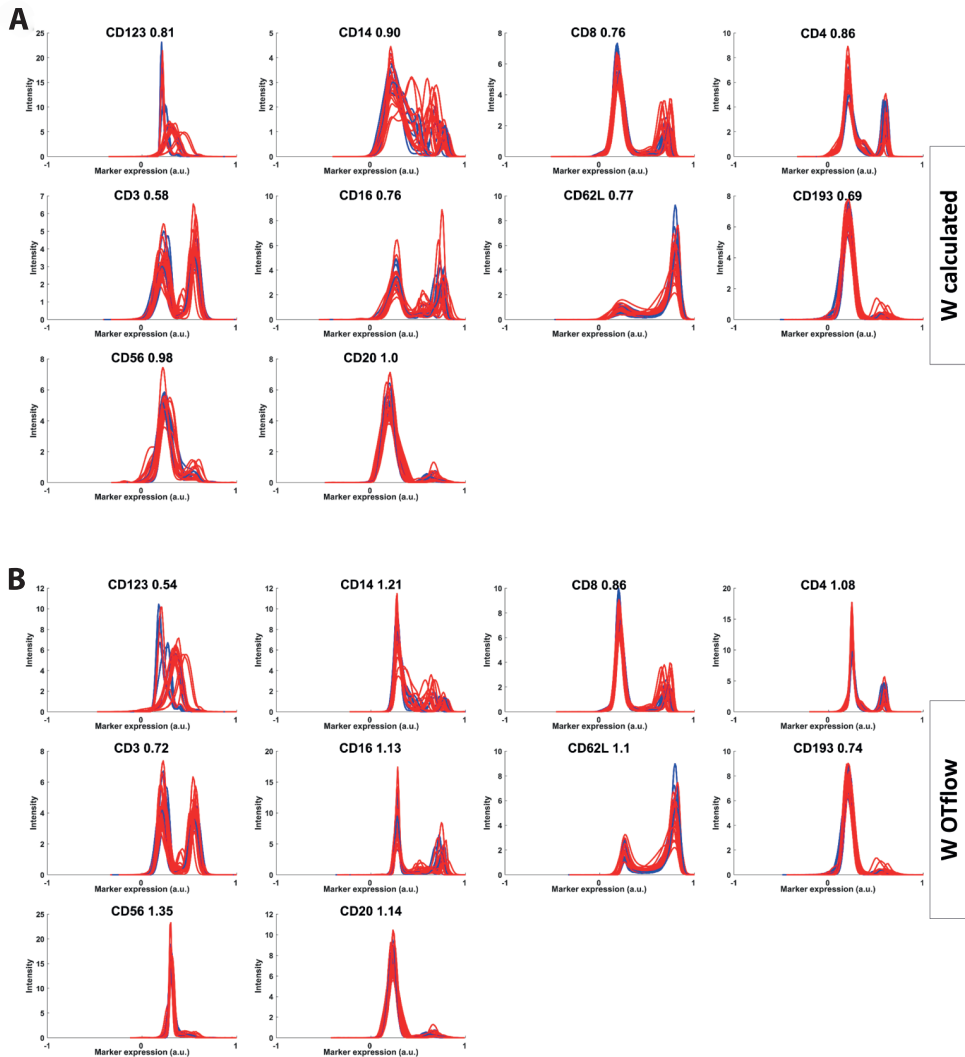
Logicle transformation

As an alternative to the arcsinh transformation, the logicle transformation is often used to transform MFC data and it is available in most of the conventional gating software packages for MFC data analysis. Mathematical details and formulation of this bi-exponential function are described in the Method section (Eq. 2). As described, the logicle function depends on various parameters, most of which are fixed to a default value or set by the experimental conditions, except the parameter  $W$ . This parameter is the essential part of the logicle function as it defines the width of the linear transformation. It can assume values between 0 and 2.25, corresponding to an increase of the linear region when going from  $W = 0$  to the maximum  $W = 2.25$ . This maximum value specifies a transformation that is essentially linear for the whole range considered. The authors Parks et al.<sup>9</sup> proposed to derive  $W$  by considering the 5<sup>th</sup> percentile of the negative values present in the data, as described in Eq. 2b.:

$W = \frac{M - \left(\log\left(\frac{T}{ r }\right)\right)}{2}$	Eq. 2b
---	--------

Where  $M$  generally assumes value of 4.5 and  $T$  is equal to 262,144; while  $r$  corresponds to the 5<sup>th</sup> percentile of the negative values as described before.

Here we use the OTflow algorithm to calculate optimal parameter  $W$  for the HIV data. The results of the OTflow-based transformation were compared to the results obtained by using the calculated  $W$  with Eq. 2b. Figure S12 shows the single marker expression of the logicle-transformed data using the differently estimated parameters  $W$ . When using Eq.2, the calculated parameters for each channel seem to be appropriate for the transformation of most of the markers as displayed in Figure S15A. Only marker CD56 is found to be an exception, as a small population can be found in the negative range which might be associated to peak splitting. Different values for  $W$  for each channel were estimated by the OTflow algorithm. Similarly to the estimation of the cofactor for the arcsinh function, OTflow optimizes the variance between the peaks as criteria to find optimal  $W$ , while preventing peak splitting and other artifacts. The results of the variance stabilization are presented in Figure S15B.



**Figure S15. Single marker expression of the logicle transformed HIV data.**

A) The parameter  $W$  used in the logicle function is calculated using the formula provided by Parks et al. 9, displayed in Eq.2. Peak splitting appearing in one channel is marked in a green box; B) The parameter  $W$  used to transform the data with logicle function is estimated by OTflow. When multiple peaks are present, OTflow uses Bartlett's statistics to optimize the variance between peaks and it provides the parameter associated to most homogeneous variance between the peaks. Peak splitting and other artifacts are prevented.

This is an example that also for logicle transformation peak splitting is prevented when using the OTflow calculation.



## References

1. Jolliffe IT. *Principal Component Analysis. Second Edition.*; 2002. doi:10.2307/1270093
2. Amir ED, Davis KL, Tadmor MD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol.* 2013;31(6):545-552. doi:10.1038/nbt.2594
3. Van Gassen S, Callebaut B, Van Helden MJ, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytom Part A.* 2015;87(7):636-645. doi:10.1002/cyto.a.22625
4. Bruggner R V., Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci.* 2014. doi:10.1073/pnas.1408792111
5. Tibshirani R. Regression shrinkage and selection via the lasso: A retrospective. *J R Stat Soc Ser B Stat Methodol.* 2011. doi:10.1111/j.1467-9868.2011.00771.x
6. Wouters K, Gaens K, Bijnen M, et al. Circulating classical monocytes are associated with CD11c+ macrophages in human visceral adipose tissue. *Sci Rep.* 2017;7:42665. doi:10.1038/srep42665
7. Verbrugge A, Ruiter T de, Geest C, Coffe PJ, Meyaard L. Differential expression of leukocyte-associated Ig-like receptor-1 during neutrophil differentiation and activation. *J Leukoc Biol.* 2006;79(4):828-836. doi:10.1189/jlb.0705370
8. Van Staveren S, Ten Haaf T, Klöpping M, et al. Multi-dimensional flow cytometry analysis reveals increasing changes in the systemic neutrophil compartment during seven consecutive days of endurance exercise. *PLoS One.* 2018;13(10). doi:10.1371/journal.pone.0206175
9. Parks DR, Roederer M, Moore WA. A new “logicle” display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. *Cytom Part A.* 2006. doi:10.1002/cyto.a.20258



## CHAPTER 5

5



# Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm

Rita Folcarelli<sup>1</sup>, Selma van Staveren<sup>2,3</sup>, Roel Bouman<sup>1</sup>, Bart Hilvering<sup>3</sup>, Gerjen H. Tinnevelt<sup>1,2</sup>, Geert Postma<sup>1</sup>, Oscar F. van den Brink<sup>2</sup>, Lutgarde M. C. Buydens<sup>1</sup>, Nienke Vrisekoop<sup>3</sup>, Leo Koenderman<sup>3</sup>, Jeroen J. Jansen<sup>1</sup>

<sup>1</sup> Radboud University, Institute for Molecules and Materials, Analytical Chemistry, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands

<sup>2</sup> TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands

<sup>3</sup> Department of Respiratory Medicine and laboratory of translational immunology, University Medical Center Utrecht, Heidelberglaan 100, 3584CX, Utrecht, The Netherlands

Multicolor Flow Cytometry (MFC)-based gating allows the selection of cellular (pheno)types based on their unique marker expression. Current manual gating practice is highly subjective and may remove relevant information to preclude discovery of cell populations with specific co-expression of multiple markers. Only multivariate approaches can extract such aspects of cell variability from multi-dimensional MFC data. We describe the novel method ECLIPSE (Elimination of Cells Lying in Patterns Similar to Endogeneity) to identify and characterize aberrant cells present in individuals out of homeostasis. ECLIPSE combines dimensionality reduction by Simultaneous Component Analysis with Kernel Density Estimates. A Difference between Densities (DbD) is used to eliminate cells in responder samples that overlap in marker expression with cells of controls. Thereby, subsequent data analyses focus on the immune response-specific cells, leading to more informative and focused models. To prove the power of ECLIPSE, we applied the method to study two distinct datasets: the *in vivo* neutrophil response induced by systemic endotoxin challenge and in studying the heterogeneous immune-response of asthmatics. ECLIPSE described the well-characterized common response in the LPS challenge insightfully, while identifying slight differences between responders. Also, ECLIPSE enabled characterization of the immune response associated to asthma, where the co-expressions between all markers were used to stratify patients according to disease-specific cell profiles.

## Introduction

Multicolour Flow Cytometry (MFC) is a powerful analytical technique, widely used in biomedicine as a diagnostic tool to evaluate and characterize health and disease.<sup>1</sup> It enables quantitative detection of marker expression, among other cell characteristics, at the single-cell level by specific antibodies conjugated to a multitude of fluorophores. The power of MFC lies in the simultaneous measurement of multiple surface or intra-cellular markers. This allows both a comprehensive biological and physical characterization of cells and cell populations of interest. Advances in technology and fluorophore chemistry have drastically increased the number of parameters that can be concurrently measured.<sup>2,3</sup> Fluorescence-based flow cytometry allows simultaneous measurement of more than 20 markers, while the most recent generation of mass cytometry platforms (Cytometry-Time of Flight) can routinely run experiments with more than 40 parameters.<sup>4</sup> In fact, massive amounts of data are generated in a single experiment, for which many different dedicated data analysis methods have been proposed.<sup>5</sup>

One of the major objectives of MFC data analysis is the identification of homogenous cell types of interest. In the conventional MFC data analysis software, cells of interest are selected through a selection process called 'gating', based on uni- or bivariate marker expressions. Manual 'multiple' gating on binary combinations of cell characteristics is by far the most widely used method. This is however highly subjective and resource-intensive, because expert technicians need to establish quantitative thresholds in several bi-dimensional plots that cannot be mutually compared on the single-cell level. Manual gating of a data set with seven measured cellular markers would already require inspection of 21 bivariate plots per individual sample. The number of possible combinations becomes difficult to manage with increasing numbers of measured markers, to the extent that the manual gating approach becomes unfeasible quite soon. Aside from the extensive time-consumption involved, it would place additional requirements in consistency of operation and expertise between operators. Moreover, this bi-dimensional approach is done hierarchically, by which cell populations may be overlooked like in sequential gating on single markers.<sup>6</sup>

Recently, several multivariate methods have been proposed to overcome these problems. The viSNE method<sup>7</sup> is commonly used as a visualization tool for high-dimensional MFC data. Clusters of single cells are visualized in a biaxial viSNE map, using the non-linear t-Stochastic Neighbour Embedding (t-SNE) algorithm for dimensionality reduction. Even though viSNE may be beneficial in the presence of strongly non-linear data, the use of such a non-convex objective algorithm brings about several drawbacks. Each run performed on the same dataset would result in a different map, making the maps difficult to validate. Consequentially and importantly, the arrangement of the cells cannot be

directly and easily associated with the marker expression and it is not possible to project a new individual in an existing map without a complete new run. This highly limits the comparison of new, incoming datasets to a model calibrated and validated as a diagnostic instrument for single-cell analysis.

Spanning-tree Progression Analysis of Density-Normalized Events (SPADE)<sup>8</sup> uses hierarchical clustering to connect different cell subpopulations in 'minimum spanning trees' which represents their mutual relations. The cell distribution is visualized as nodes of clustered cells in the SPADE tree that have specific phenotypes. Unlike viSNE, a new MFC sample may be represented into a spanning minimum tree previously built on a dataset, by matching all the cells to the nodes with the most similar phenotype. However, if an extra cell population is present in the new sample, these cells are forced to incorrectly belong to one or more of the available nodes. The (high) residuals of the projection of the cells are not directly detectable. Another recently developed method, Citrus<sup>9</sup>, also uses hierarchical clustering to identify phenotypically similar cell populations. The method is particularly used for intergroup analysis, for which a regularized classification model detects group-specific cell clusters for each sample. Both SPADE and Citrus adopt multivariate and linear approach but they fail, like viSNE, to show the multivariate co-expression that underlies the diversity of cell phenotypes. Co-expression relationships among all the measured markers either remain hidden in the models or are limited to a visual inspection of all the single marker expression levels.

Two alternative methods to analyse Flow Cytometry data are APS (Automatic Population Separator)<sup>10</sup>, and FLOOD (Flow cytometric Orthogonal Orientation for Diagnosis)<sup>11</sup>. Both methods employ Principal Component Analysis (PCA)<sup>12</sup> as dimension reduction technique to condense the relevant biological cell marker information in a few, mostly two, dimensions. Although they use the same underlying principle, both methods differ considerably in important aspects. APS builds a PCA model on different phenotypes of disease known as reference groups. After that, each case is compared against each reference group<sup>13</sup> and it is matched to the closest group in the PCA space. The reference groups need to be established a priori, which becomes challenging in absence of information about a disease and/or with a small number of individuals within each reference group. The solution to circumvent this disadvantage is to use healthy individuals as reference group and highlight the deviation of diseased individuals from the normal cell marker variability, as done in FLOOD. The FLOOD approach first describes the characteristic cell-to-cell marker variability among healthy (or control) individuals; then it models the cell variability present only in responding individuals within a second model, the response model. FLOOD provides multivariate 'biplots' that offer an intuitive view of the cell marker variability associated to a specific immune response, interpreted in terms of co-expression between all the markers, in a representation that is analogous to that used in bivariate

gating. However, the information within the FLOOD response model is reduced by the contribution of considerable numbers of 'healthy' (or not response-specific) cells in the analysis of responding individuals.

We present ECLIPSE (Elimination of Cells Lying in Pattern Similar to Endogeneity), which combines Simultaneous Component Analysis (SCA), a generalization of PCA suitable for MFC data, and Probability Density Functions of the cell distribution in the dimension reduced space, to identify and eliminate the 'healthy' or 'normal' cells from the patient (or responder) data. Since these normal cells in the responder individuals are not the discerning factor in the disease response (only the change in number is), they can be removed from the MFC dataset. When initially healthy cells change in relative abundance or when they upregulate or downregulate certain markers as a response to a stimulus or a disease, these cells are considered response-specific and are, therefore, not removed from the dataset.

## Methods

### Data

#### LPS study

MFC measurements of the 'Lipopolysaccharide study' were performed on cells obtained from healthy individuals challenged with systemic endotoxin (LPS). (NCT01374711; [www.clinicaltrials.gov](http://www.clinicaltrials.gov)) The LPS dataset comprises gated neutrophils from eight 'control' individuals who did not receive LPS intravenously, and eight different 'response' individuals who received LPS administration and from whom blood was collected 180 minutes post LPS administration. Seven surface markers were measured on the neutrophils in samples of both classes of individuals: CD62L, CD11b, CD11c, CD64, CD32, CD69, and CD16. Further details regarding the Flow Cytometry experiments that generated the data can be found in a previous publication.<sup>11</sup>

#### Asthma study

The asthma dataset consists of a 'response group' of 11 patients classified as suffering from severe asthma (aged 22–78,  $\bar{x} = 57$ ) and a 'control group' of 10 non-asthmatic individuals (aged 25–57,  $\bar{x} = 40$ ). Whole-blood cells were stained with the following eight fluorescent markers: CD3, CD4, CD8, CD14, CD16, CRTH2 (CD294), CD123, and CD193. After staining, red blood cells were lysed using a FACS Lysing solution (Becton Dickinson). Cells were measured on a LSR Fortessa flow cytometer (Becton Dickinson).<sup>14</sup>

In both studies, the data were compensated before the analysis to limit the spectral ‘spill over’ between the markers. This assures that conclusions drawn on marker co-expressions were mainly due to biological correlation rather than spectral overlap.

Details about the pre-processing procedure to transform the data can be found in the Online Supplementary Material I.

## **ECLIPSE algorithm**

The MFC data were arranged in a multiset structure illustrated in the Supplementary Material I. This arrangement is relevant throughout the quantitative analysis of Flow Cytometry data and each following step of ECLIPSE therefore takes this distinctive structure by design, while retaining the information of each single individual sample. In fact, the ECLIPSE method, through the whole algorithm, takes into account the fact that cells in MFC data may come from different individuals. Difference in terms of the number of cells measured per individual is preventively addressed in the pre-processing (extensively described in Step 0: Data pre-processing, in Online Supplementary Material I), so that each individual equally contributes to the models, disregarding the size of each set. The ECLIPSE algorithm is described step by step and the FlowChart of the methodology is shown in Figure 1.

### ***Step 1: Simultaneous Component Analysis on control individuals.***

Chemometric dimension reduction techniques such as Principal Component Analysis (PCA) reduce the multidimensional information into a few dimensions to facilitate the information-rich representation of, in this case, flow cytometry data<sup>6</sup>. PCA is an unsupervised technique for the multivariate explorative analysis that operates dimension reduction of multidimensional data. The idea underlying PCA is to contract the information present in the original variables in a lower-dimensional space that best fits the data, with the aim of retaining all the relevant information and removing minor, non-systematic sources of variability. PCA operates a projection of the original data onto a new, reduced space where each dimension is a linear combination of the original variables. The new dimensions, called *Principal Components* (PCs), capture as much variability in all immunological features among all measured MFC events as possible. They are mutually orthogonal and the first Principal Component explains most of the original variance and the successive principal components explain then the highest percentage of remaining variance.<sup>12</sup>

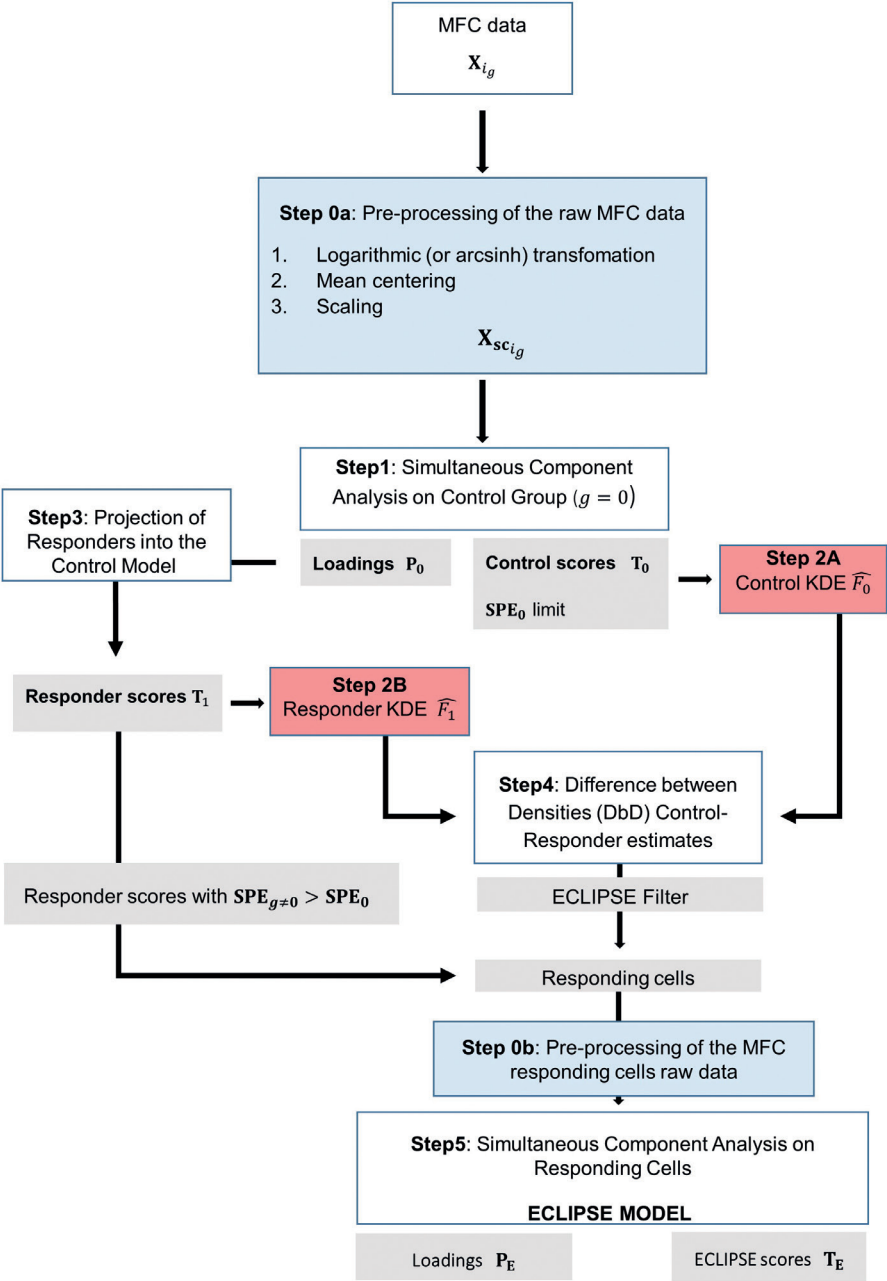


Figure 1.

Flowchart of the ECLIPSE algorithm. In light blue are highlighted the pre-processing steps (0a: pre-processing on the raw MFC data, 0b: pre-processing on the raw MFC data, after removal of normal cells); in red the probability densities distribution estimations (2A: estimate of control PDF, 2B: estimate of responder PDF, from the projection of responder-group cells); in grey the results of the corresponding previous steps.

PCA has been widely used to reduce the dimensionality of Multicolour flow cytometry data<sup>10,15,16</sup>. Because these earlier methods disregard the MFC multiset structure (described in Figure S1 of the Online Supplementary Material I), application of such methods may lead to individuals with large numbers of cells dominating the resulting model. Multiset extensions of PCA, such as described here, consider these differences, in a way that each individual will contribute equally to the constructed model. Secondly, PCA on both control and responder individuals simultaneously may mix normal and response-related cell variability in the resulting PCA model. Both aspects need to be addressed in a comprehensive modelling approach.

From the multiset structure, we consider the control individuals and we build a *Control Model*, which describes the cellular markers' variability within all the control class individuals. This variability is not of primary interest for describing the immune response; it may therefore serve as a benchmark, where the complement to the cell variability in immunologically perturbed samples or responder individuals is more likely response-related. The availability of identical loading vectors across all control individuals explicitly models the similarities between all control individuals and allows direct comparison between the distinctive scores for each control individual. The Simultaneous Component Analysis with equal patterns (SCA-P) imposes the constraint of identical loading vectors.<sup>17</sup>

The pre-processed MFC data are arranged in the matrix  $\mathbf{X}_{sc_0}$  from Equation S5 (Supplementary Material I, *Data pre-processing*) of size  $N_{i_g} \times J$ , where  $N_{i_g}$  is the number of cells of the  $i_g$ -th individual,  $i_g = 1_g, \dots, I_g$ , with  $g = 0$  for the control and  $g \geq 1$  for the responder groups;  $J$  corresponds to the markers measured,  $1 \dots j \dots J$ . Before applying the SCA decomposition,  $\mathbf{X}_{sc_0}$  is normalized such that each individual contributes with the same amount of information, regardless of the number of cells measured per set (Eq. 1). The resulting matrix  $\mathbf{X}_{sc_0^*}$  is then decomposed according Equations 2a-b-c:

$$\mathbf{X}_{sc_0^*} = \begin{bmatrix} \mathbf{X}_{sc_{1_0}} \mathbf{N}_{1_0}^{-1} \\ \vdots \\ \mathbf{X}_{sc_{I_0}} \mathbf{N}_{I_0}^{-1} \end{bmatrix} \quad \text{Eq.1}$$

$$(a) \mathbf{X}_{sc_0^*} = \mathbf{T}_0 \mathbf{P}_0^T + \mathbf{E}_0 \quad \text{Eq.2}$$

$$(b) \mathbf{T}_0 = \mathbf{X}_{sc_0} \mathbf{P}_0^T$$

$$(c) \mathbf{E}_0 = \mathbf{X}_{sc_0} - \mathbf{T}_0 \mathbf{P}_0^T$$

where the matrix of the scores  $\mathbf{T}_0$ , of size  $\sum_{i_0}^{I_0} N_{i_g} \times R_C$ , with  $R_C$  dimensionality of the new low-dimensional space, holds the coordinates of the  $i_g$ -th control individual in the new space;  $\mathbf{P}_0$  the  $J \times R_C$  matrix of loadings, common to all the control individuals;  $\mathbf{E}_0$  of size  $\sum_{i_0}^{I_0} N_{i_g} \times J$  contains the residuals of the  $i_g$  control individuals, with  $i_g = 1_0, \dots, I_0$ .



The scores  $\mathbf{T}_0$  represent the major variability across cells in all cellular markers simultaneously. This is however an inherently simplified representation. The residuals matrix  $\mathbf{E}_0$  contains the cell variability that remains unmodeled by loadings  $\mathbf{P}_0$ . These residuals may be diagnostic for outlying or abnormal cells, where cells with high residuals are poorly described by the Control loadings. This abnormality is quantified in the Sum Prediction Error (SPE) in Equation 3:

$$\mathbf{SPE}_{i_0} = \text{sum}(\mathbf{E}_{i_0} \mathbf{E}_{i_0}^T) \quad \text{Eq. 3}$$

Where  $\mathbf{E}_{i_0}$  contains the residuals for the  $i_0$ -th individual. The statistics for all the control individual might be collected in the cumulative  $\mathbf{SPE}_0 = \begin{bmatrix} \mathbf{SPE}_{1_0} \\ \vdots \\ \mathbf{SPE}_{I_0} \end{bmatrix}$  of size  $\sum_1^I N_{i_0} \times 1$ . A confidence limit corresponding to the  $(1 - \alpha)$  percentile may be evaluated for the statistics  $\mathbf{SPE}_{i_0}$  of each individual as the value exceeded by  $\alpha\%$  of the cells. These individual values with  $\alpha=0.05$  can be collected in the vector  $\mathbf{SPE}_{0,0.05} = \begin{bmatrix} \mathbf{SPE}_{1_{0,0.05}} \\ \vdots \\ \mathbf{SPE}_{I_{0,0.05}} \end{bmatrix}$  of size  $I_0 \times 1$ . The total confidence limit, is calculated as the mean of the individuals  $\mathbf{SPE}_{i_{0,0.05}}$  and it is used as reference to establish the degree of "normal condition" of the cells.

**Step 2: Probability density estimation of cell variability with KDE.**

The Control Model may be represented in a multivariate biplot. This two-dimensional representation, obtained by plotting two principal components and superimposing these onto the loadings, allows simultaneous visualisation of the  $\sum_{i_0}^I N_{i_g}$  observations and the  $J$  variables and therefore how these associate.

In the resulting space, the single cell coordinates of an individual are represented as single points. To enable interpretation of the distribution of the cells and the presence of distinctive cell subpopulations we approximate this distribution from the measured single cells by multidimensional density estimation. The probability density function (PDF) may be used to describe the probability that a cellular marker expression assumes a certain value in a D-dimensional space, which can be established by Kernel Density Estimation.<sup>18,19</sup>

This estimate does not require *a priori* assumptions about the shape of the underlying distribution of the expressions, such as Mixture Modeling<sup>20</sup>. The density estimation results from the sum of the individual contributions of the estimated function centered at each sample position. In our case the sample position is represented by the scores' coordinates in the  $R_C$ -dimensional PCA space for each cells belonging to the  $i_g$ -th individual.

Step 2A, in the Flowchart Figure 1, estimates the PDF of the control cell scores with KDE. Equation 4 contains the formula for the KDE for each cell score belonging to the  $i_0$ -th control individual:

$$\widehat{f}_{i_0}(t, h) = \frac{1}{N_{i_0} h} \sum_{1_0}^{N_0} K\left(\frac{\mathbf{t}^T - \mathbf{t}_{n_{i_0}}^T}{h}\right) \quad \text{Eq. 4}$$

where  $\mathbf{t}^T$  is a vector holding the location where the function is being evaluated, expressed in score coordinates in the PCA space  $\mathbf{t}_{n_{i_0}}^T$ ; contains the location of the  $i$ -th cell score belonging to the control individual  $i_0$ -th;  $N_{i_0}$  is the total number of cells;  $K$  the kernel function centered at the cell score location and  $h$  the width of the such kernel function, that works as smoothing parameter.

We used Gaussian function as kernel functions  $K$ . Several studies have shown, however, that this choice of  $K$  does not strongly determine the estimate<sup>21</sup>. The kernel bandwidth  $h$  in Equation 4 determines the model outcome to a much larger extent. Figure S2 in the Online Supplementary Material I, illustrates the effects of the bandwidth choice on the estimate of a normal distribution. Large bandwidths will mask finer structures of the data that may be biologically relevant. Too-small bandwidths may result in spiky density estimates that may be challenging to interpret. Finding the bandwidth where the estimate best approximates the true distribution is therefore imperative. Various methods for optimal bandwidth selection have been proposed in literature<sup>22,23</sup>. The method proposed by Botev et al. based on the Fast Gauss Transform performs best in estimating simulated distributions<sup>24,25</sup>. This relatively fast and accurate method was particularly better in representing multimodal distributions of the cellular marker expressions that are characteristic of multiple cell populations. A specific advantage of using the aforementioned method to fit the KDE lies in the fact that it automatically evaluates the optimal bandwidth of Gaussian kernel functions  $K$ . The KDE formulation (Eq. 4) already includes a normalization over the number of cells per individual  $N_{i_0}$ , facilitating quantitative comparison between estimates of different individuals within the control class. The KDEs for each individual may be pooled into a KDE<sub>0</sub> of all control individuals by summing up the individual KDEs and subsequent normalization to unit probability.

**Step 3: Projection of responder individuals onto the Control Model**

The Control Model created in Step 1 explains the cellular marker variability of the control individuals. Step 3 consists of orthogonal projection of the cells of the responder individuals into the control space. Assuming the value for the case group  $\neq 0 \wedge g = 1$ , the orthogonal projection is provided by the equation:

$$\mathbf{T}_1 = \mathbf{X}_{sc1} \mathbf{P}_0 \quad \text{Eq. 5}$$

$$\mathbf{E}_1 = \mathbf{X}_{sc1} - \mathbf{T}_1 \mathbf{P}_0^T \quad \text{Eq. 6}$$

Where Eq. 5 is further equivalent:

$$\begin{bmatrix} \mathbf{T}_{11} \\ \vdots \\ \mathbf{T}_{I_1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{sc11} \\ \vdots \\ \mathbf{X}_{scI_1} \end{bmatrix} \mathbf{P}_0 \quad \text{Eq. 7}$$

$\mathbf{T}_1$  ( $\sum_{i_g=1}^{I_1} N_{i_g} \times R_C$ ) is the matrix of the scores for all the  $i_1$ -th responder individuals, projected in the Control Space using the orthogonal loadings  $\mathbf{P}_0$  (Eq. 1) from controls and  $\mathbf{E}_1$  ( $\sum_{i_1=1}^{I_1} N_{i_g} \times J$ ) refers to the matrix containing the residuals of the projection. Cells with high residual are outliers in the Control Model and considered abnormal or response-specific. Such cells can be detected by evaluating the Sum Prediction Error (SPE) of the projections using Eq. 3, against the confidence limit for the control calculated as mean of  $SPE_{i_{0,0.05}}$  the of the  $i_0$ -th control individuals. Cells of which the SPE exceeds this limit are considered 'non-normal' in their marker expression, as it does not overlap with the variability within the Control model density estimate. These non-normal cells are specifically interesting for the immune response and therefore excluded from the Control Space and they will be reintroduced in the model just after the ECLIPSE filter provided by Step 4, Flowchart Figure 1).

We estimate the Probability Density Function for the scores of the remaining cells of the responder individual with KDE using Eq. 4 (Step 2B of the Flowchart, Figure 1). In addition to the amenable representation of the cell scores distributions in a 2D-space, the PDF estimations might allow a direct and easy comparison of cell distribution drawn from different populations.

**Step 4: Difference between Densities**

We may now match the KDE of each responding individual against the cumulative control kernel estimate, resulting in a Difference between Densities (DbD) for each responder individual<sup>26</sup>. This step is the core of the ECLIPSE algorithm, with the DbD as an automated filter to exclude those cells with a marker expression that is also found in control individuals. Cells are binned onto the DbD; bins with a positive value contain cells with a higher chance to be found in control individuals. This isolates the cells that are more

likely to be found in responders than in control individuals, and therefore likely response-specific together with those cells with a SPE that exceeds the limit observed in the control individuals isolated in Step 3.

#### **Step 5: ECLIPSE model**

As a result from previous steps the new matrix  $\mathbf{X}_{E_1}$  contains the raw signals of the response specific cells while retaining the information of the individuals, which means that the multiset structure is maintained. The resulting matrix needs to be pre-processed following the recipe chosen in Step 0. After log or arcsinh transformation,  $\mathbf{X}_{E_1}$  is autoscaled using the mean and the standard deviation of each responder individual (i.e. individual centering), according to Eq. S2b and S3c, in the Online Supplementary Material I. The autoscaled matrix  $\mathbf{X}_{E_1}$  is normalized over the number of cells of each individual, following Eq. 1, and then decomposed through Simultaneous Component Analysis, according to Eq. 8:

$$\begin{aligned} (a) \quad \mathbf{X}_{E_1^*} &= \mathbf{T}_{E_1^*} \mathbf{P}_{E_1^*}^T + \mathbf{E}_{E_1^*} \\ (b) \quad \mathbf{T}_{E_1} &= \mathbf{X}_{E_1^*} \mathbf{P}_{E_1}^T \\ (c) \quad \mathbf{E}_{E_1} &= \mathbf{X}_{E_1} - \mathbf{T}_{E_1} \mathbf{P}_{E_1}^T \end{aligned} \quad \text{Eq. 8}$$

where the matrix of scores  $\mathbf{T}_{E_1}$  has size  $\sum_{11}^{I_1} N_{E,ig} \times R_E$ , with  $\sum_{11}^{I_1} N_{E,ig}$  the total number of activated or responding in the responder individuals, and  $R_E$  the number of dimensions of the new ECLIPSE space in which the majority of the original variability is explained;  $\mathbf{P}_{E_1}$  the  $J \times R_E$  matrix holding the loadings, common to all the responder individuals;  $\mathbf{E}_{E_1}$  of size  $\sum_{11}^{I_1} N_{E,ig} \times R_E$  contains the residuals for the ECLIPSE space.

The ECLIPSE model, built on the activated cells within the responder individuals, allows the investigation of the cellular marker variability introduced by the immune response to the disease. This variability is indeed modelled through SCA, which makes it directly comparable across of the responder individuals with the common matrix of loadings  $\mathbf{P}_{E_1}$ . The loadings contained in  $\mathbf{P}_{E_1}$  exclusively describe relations and correlations among cellular markers exhibited by cells that produced or activated in the immune response. These relations are the ones important for the understanding of the pathogenesis and progress of the disease under study. No information about the normal cells is left in this model, because the elimination Step 4 excludes cells with marker expressions similar to those observed in the control individuals.

#### **Stratified ECLIPSE model**

The ECLIPSE model may be very insightful to reveal similar cell distribution patterns across different individuals. This information can then be used to subsequently stratify individuals based on their ECLIPSE scores into responder subgroups. This can be very useful to study immune responses involving subtypes that are either known or unknown,

with cell variability both unlike each other and unlike that in the control individuals. Such stratification could be done by visual inspection of the score distributions of  $\mathbf{T}_{E1}$ , perhaps including prior knowledge, but here we propose to cluster the scores of different individuals with an automated algorithm. Cells of each individual are binned into the ECLIPSE space; pairwise distances are calculated between bins across all the responders. The obtained distance matrix is then a measure of (dis)similarity of the score distributions between the individuals. Responders that present cells in the same bins (small distances), express similar cell populations and are therefore allocated to the same cluster. When multiple individuals cluster closely together and away from other individuals, they can be modelled by an ECLIPSE (partial) cluster-specific model dedicated to those individuals to better focus upon their specific cell variability in marker expression of likely response-specific cells.

### Quantitative Validation of the ECLIPSE approach

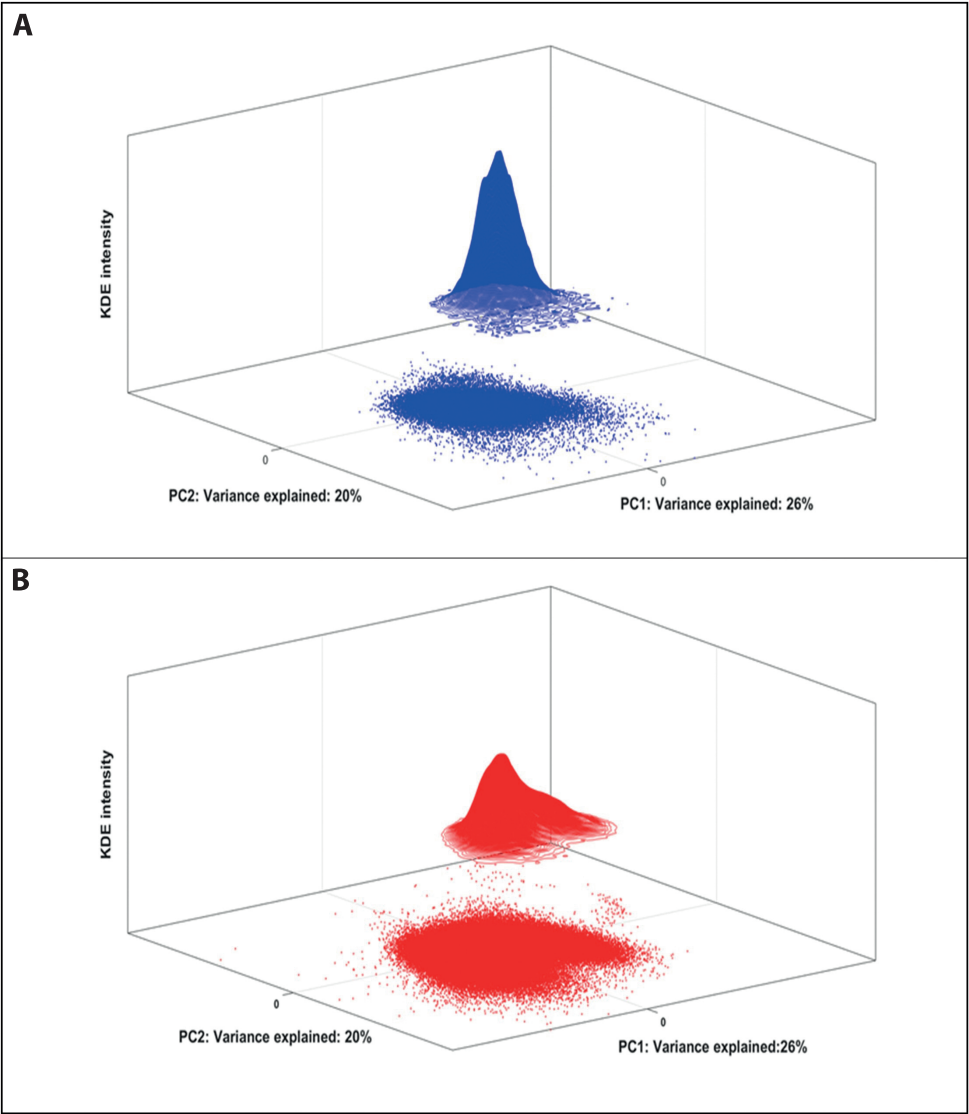
Steps 1 and 5 (Flowchart Figure 1) involve a dimension reduction to condense the cellular marker variability within the control individuals (Step 1) and the responder cells (Step 5) respectively into fewer representative dimensions. The number of significant principal components needed for this should be estimated with a bootstrapping procedure to result in a stable model solution<sup>27,28</sup>.

The cells within each individual  $\mathbf{X}_{i_g}$  from the multiset structure are bootstrapped and SCA is applied to each new resulting multiset matrix; the stability of the angle between the loading subspaces of the bootstrapped samples should be high to indicate that the SCA components describe relevant and systematic information rather than cells occurring in only one or a few individuals.

## Results

We illustrate ECLIPSE with two case studies. The first investigates the immune response in individuals that intravenously received endotoxin Lipopolysaccharide (LPS) as a model of systemic inflammatory response syndrome (SIRS).<sup>29</sup>

In the second study, the immune response in peripheral blood of severe asthma patients is examined. Asthma is a chronic and heterogeneous disease, consisting of several disease phenotypes with distinct etiologies and underlying immune responses. Distinguishing these is essential for efficient medical treatment of the relevant phenotypes in the asthma patients.<sup>30</sup>



**Figure 2.**  
A) Scores plot of control cells of all the control individuals in the Control Space, plus 3D-contour of the PDF of the control scores estimated with KDE; B) Scores plot of all the responder cells projected in the Control Space, plus 3D-contour of the PDF of the responder scores estimated with KDE.

The running time for both the LPS data (281505 events) and the asthma study data (2157326 events) did not exceed 2 minutes when computing the analyses on a laptop with 2.7 GHz, i7-7500U processor, 8GB of memory.

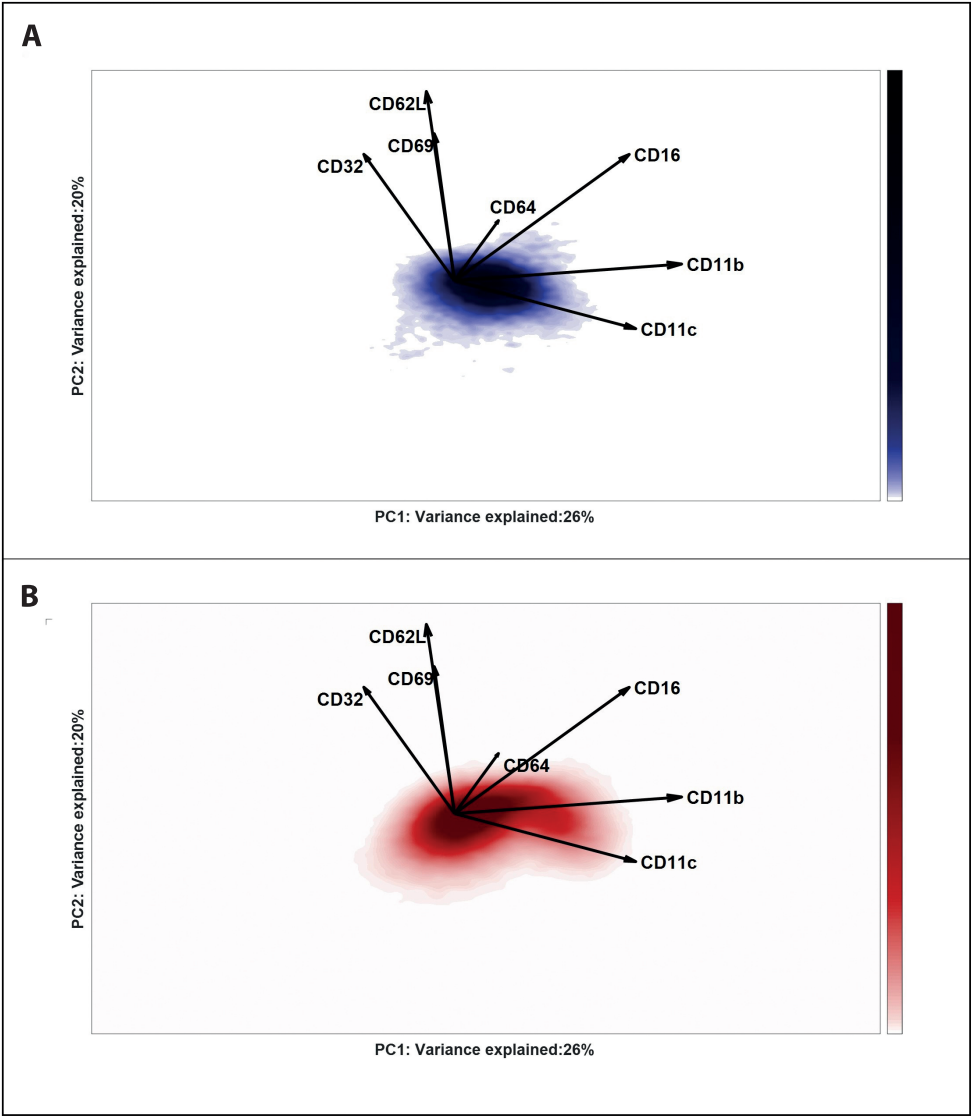
### Lipopolysaccharide challenge

Intravenous administration of LPS in humans induces the rise of neutrophil subpopulations in the peripheral blood that are normally absent. This is most prominent 180 minutes after LPS administration. The two neutrophil subpopulations that are induced after LPS challenge can be distinguished based on their expression of CD16 and CD62L. One of the populations is CD62L<sup>dim</sup>CD16<sup>bright</sup> and the second population is CD62L<sup>bright</sup>CD16<sup>dim</sup>. Next to these markers it has been shown that co-expression of multiple other surface markers characterize these neutrophil subpopulations.<sup>11,31</sup>

After log-transformation, the data were mean-centred per individual and scaled over the control group, according to Equations S2b and S5b, respectively. Supplementary Material I contains more detailed information on the data pre-processing. A control individual identified as outlier from previous analysis<sup>11</sup> was excluded from the ECLIPSE modelling.

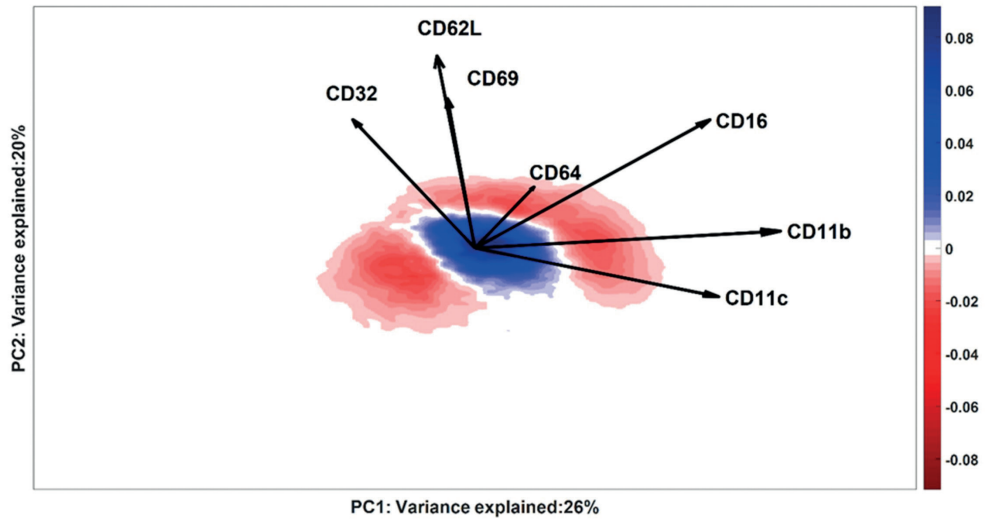
The bootstrapping procedure indicated that two components for the Control Model describe the variation of the marker expression in a stable model of the Control individuals. These components together explain 45% of variability in marker expression across cells from control individuals, ranging between 42-53% per individual. This indicates that the model provides a balanced representation of all seven control individuals. Although this percentage might appear nominally low because the majority of marker variability remains unexplained, statistical model validation shows that models with additional components are more unstable and therefore inappropriate as generic filters for 'normal' cells.

Figure 2 shows the Control score distributions of cells in control individuals (a, in blue) and of the responder individuals (b, in red) projected onto the control model. Dots represent single cells within the dimension-reduced space. The third dimension is the 'Kernel density estimation (KDE) intensity', i.e. the local density estimate at each cell score position. Locations with higher KDE intensities are more densely populated by cells. The inhomogeneous estimate in B indicates a distribution of cells with a continuum variation of fluorescent signal intensities, i.e. surface marker expressions, compared to the control individuals.



**Figure 3.**  
A) Biplot representation of PDF of the control scores, estimated with KDE; B) Biplot representation of the PDF of the responder scores, estimated with KDE. The loadings of the SCA Control Model are plotted as vectors: their length indicates the contribution of each marker to the cell-to-cell variability; the mutual directions suggest a positive (same direction) or negative (opposite direction) co-expression. An angle of 90 between the direction of the vectors indicates no co-expression.

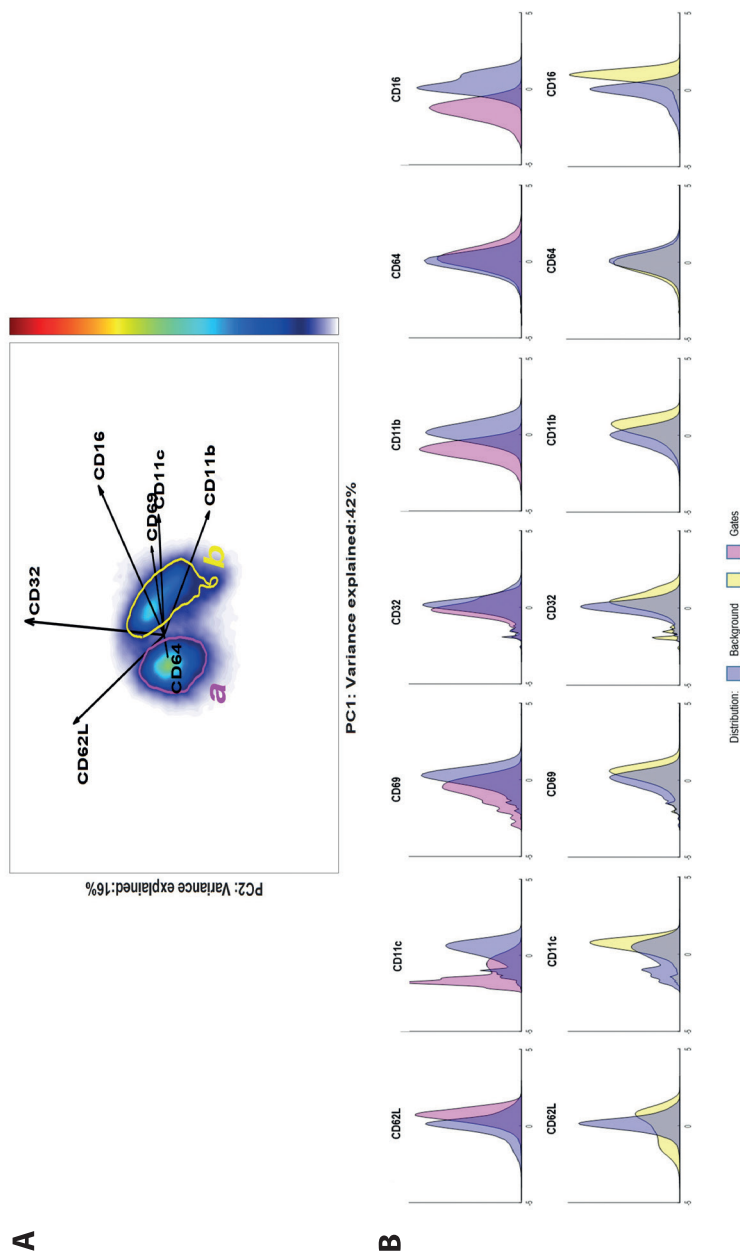




**Figure 4.**

KDE of the scores distribution of a typical LPS-responder (ID #9) is subtracted from the cumulative KDE of the control scores distribution. The negative intensity (red) indicates the location where cells over-produced in the responder are more likely to be present; the positive intensity (blue) indicates the location where healthy cells are more likely to be present. The white area between the red and blue areas corresponds to a value of KDE=0, which can indicate bins with no cells or equal intensity of control and responder estimates. The axes report the variance of ID#9 explained by the respective PCs (Control Model).

The biplot (Figure 3a-b) shows the scores distribution of the cells, as a 2D-density representation, together with the surface marker loadings. This plot shows how the continuum of the cell scores associates with the co-expression of the different surface markers. The homogeneous neutrophil population in the control individuals (Figure 3a) varies mostly in expression of CD11b, CD11c and partially CD16, where cells on the right side of the distribution co-express these markers above-average and more than cells on the left side. The large loadings of CD32, CD62L and CD69, in the vertical direction, indicate that the expression of these markers varies considerably among the neutrophils, along the second component.



**Figure 5.**

A) ECLIPSE model built on all the responding cells from the responder individuals. The cells generally attributed to the control group have been eliminated from the analysis, and the SCA space has been recomputed. The loadings show the LPS-response specific surface marker expression for all cells. Gates a (magenta) and b (yellow) encircle the two most densely populated areas; Panel b, Histograms of the surface marker expression of the cells lying in the two different gates. The distributions are displayed with the colour of the corresponding gate (magenta for gate a, yellow for gate b); the blue distribution corresponds to the cells not included in the specified gate.

The responder cells projected into the Control Model (Figure 3b) vary in a characteristic 'bean-like' shape, showing neutrophil subpopulations towards the markers CD16/CD11b/CD11c from one side and CD32/CD62L/CD69 from the other, that emerge upon LPS response. However, responder individuals contain a considerable fraction of cells with cell surface marker expressions identical to the cells within samples from non-responding Control individuals. These cells are less-interesting for the immune response characterization and would considerably affect any quantitative models of cells variability in responder individuals, at the expense of the rarer cells associated with the LPS induced immune response

We therefore selected those cells that deviate from the normal cell surface marker variability in the space defined by the first two PCs of the Control model. These abnormal cells are selected in two independent steps.

First, cells with high Sum Prediction Error (SPE), analogous to Equation 3, are worth further investigation. The Control Model cannot explain these cells well: they exceed the limit established for the Control Model, and are therefore designated as 'abnormal'. The percentage of these abnormal cells ranges between 12-21% for the different responder individuals—considerably and consistently higher than the 5% of cells in control individuals.

Secondly, cells with model residuals similar to those in the control individuals may be further filtered by quantitative comparison of the kernel density estimates of each Responder against that of the control group. The resulting responder-specific Difference between Densities (DbD) for a typical responder is shown in Figure 4. Regions with negative intensity contain cells with a higher chance to be found in a responder than in the control individuals on the basis of their Control Model score. These cells are therefore potentially response-related. Those cells lying in regions with positive intensity may be found with higher probability in control than in responder individuals and are therefore excluded from subsequent analyses.

A dedicated new SCA model was modeled on those cells that were either identified based on SPE or DbD as response-related, to specifically fit the variability in marker expression of these LPS-induced cells.

The ECLIPSE Model computed on the response-related cells that were identified before (Figure 5) shows how two groups of neutrophils arise in the LPS-induced immune response.

The loadings, superimposed on the space, indicate how the surface markers in both groups of cells differ in expression. CD62L is more strongly expressed in gate **a**, and CD69, CD16, CD11c, and CD11b are more intensively co-expressed on the cells in gate **b**. Although CD32 varies much in expression across response-related cells, its intensity does not significantly differ between both cell groups.

The population with higher expression of CD62L and lower expression of CD16, in gate **a**, corresponds to the premature neutrophils (cells with a banded nucleus) and the cells with higher expression of CD16 are mature neutrophils, which were both identified in earlier studies.<sup>31</sup> Single-marker expression profiles of the cells present in both gates are shown as histograms in Figure 5b, against a background containing the cells that are not included in the specified gate. The background contains both cells of the responders as well as cells of controls outside of the depicted gate. This is analogous to the histogram representation used in Citrus. A differential expression of the marker CD62L, indicated also by the direction of the loading, is present in gate **b**. This suggests the presence of two sub-clusters within the gate: a cluster which is CD62L and a CD62L cluster. These sub-clusters might occur due to variability between the different responder individuals. The density estimates of the ECLIPSE cell scores of two different responders (Figure S3, Online Supplementary Material II) confirm that there is heterogeneity in the surface marker expressions of the cells specifically induced by LPS. The variability in the continuum of surface marker expressions for all the responders are shown in the Table S2 of the Supplementary Material II. All responder individuals react with increases in premature neutrophils three hours after LPS was administered, while the presence and expression profiles of the mature neutrophils vary between individuals. In some cases, these cells consist of two subpopulations. The diversity in responses to the administered LPS is enhanced after the elimination of normal cells. Especially when the number of response-specific cells is limited or when the response is heterogeneous, these normal cells will considerably affect quantitative models of the responder individuals, at the expense of the rarer cells associated to the immune response. As proof of concept, we performed a simultaneous component analysis on the original cells of the responder individuals and compared the model against the SCA Control Model and the SCA ECLIPSE Model, defined previously. The angle between the subspaces spanned by the loadings of each model, built with the first two Principal Components, is used as metric of dissimilarity. A procedure similar to the procedure as described by Liu and al<sup>27</sup> gives the distribution of the angles between the subspaces. The smaller angle distribution in Figure S4 (Online Supplementary Material II) confirms that a model built on all cells of the responders grasps a variability in surface marker expression similar to that of the control individuals. The larger angle difference between the ECLIPSE SCA model and Control or Response SCA model shows that the elimination step is essential to capture the surface marker expressions specifically related to the response.

To compare the ECLIPSE results with other multivariate methods, we performed viSNE and Citrus analyses on the same dataset. The cell phenotypes found by the three methods can be compared. From this, we can conclude that the cell phenotypes, highlighted in the ECLIPSE space, are similar to the ones found by the Citrus and viSNE analyses in Supplementary Material III. We performed Citrus and viSNE analyses on three differently pre-processed LPS datasets: the arcsinh transformed (cofactor) LPS data, the LPS dataset pre-processed as used in ECLIPSE (log transformation, mean-centering per individuals, scaling over controls) and the LPS dataset after removal of normal cells by ECLIPSE. When trained on the arcsinh transformed LPS data, the two most relevant clusters for the classification detected by Citrus may be identified as premature and mature neutrophils (Figure S6). In contrast to the ECLIPSE results, the Citrus cluster of mature neutrophils only consists of CD62L low cells. The cross-validation error rate of the model around 7% (Figure S5) however suggests that the identified clusters might not be specific for all the individuals within the control or responder groups.

Contrary to the arcsinh transformed data, the multi-set data pre-processed according to Equations S2b and S5b led to a perfect classification with Citrus (Figure S7). The similar pre-processing also allows a more direct comparison between both methods. In this case, Citrus recognized two clusters in the LPS data that are more abundant in the control group, with phenotypes that may be assigned to normal mature neutrophils, with CD16+CD62L+CD64- and CD16+CD62L+CD64dim phenotypes, respectively (Figure S8). Only a single cluster, identified as a premature neutrophil population (based on CD16<sup>dim</sup> marker expression), was found to be distinctive for the response group.

In the third analysis, which was performed on the LPS data after removal of normal cells, only a cluster of premature neutrophils was detected (Figure S10). The premature neutrophil population, which was more abundant in the response group, suffices for a correct classification (Figure S9). In this analysis, there were no clusters of normal mature neutrophils identified as most discriminant ones.

When performing the viSNE analysis on the arcsinh transformed (cofactor 5) LPS data, no clear subpopulations were found. The viSNE map shows single cells with distinctive expression of CD11b, CD11c, CD16, CD32 and CD62L, and the other markers in lesser extend (Figure S11).

Additionally, it seems that cells of the same individual are grouped together, suggesting that the cell distribution found by the algorithm are individual-specific. This grouping is removed when viSNE is performed on the pre-processed data. In the resulting map the cells are indeed evenly distributed across the space and no donor-specific clusters are present (Figure S13).

When the analysis was performed on the data after removal of normal cells, roughly three distinctive cell phenotypes were found based on CD16/CD62L expression: CD16-CD62L+, CD16+CD62L- and CD16+CD62L+ cells (Figure S14). The expressions of the other markers within these three cell populations are more easily interpreted after ECLIPSE-based filtering. Removing the normal cells results in a less crowded viSNE map, which helps identifying the cells of interest and focusing on more subtle differences between and within these cell populations.

### **Asthma-specific immune responses**

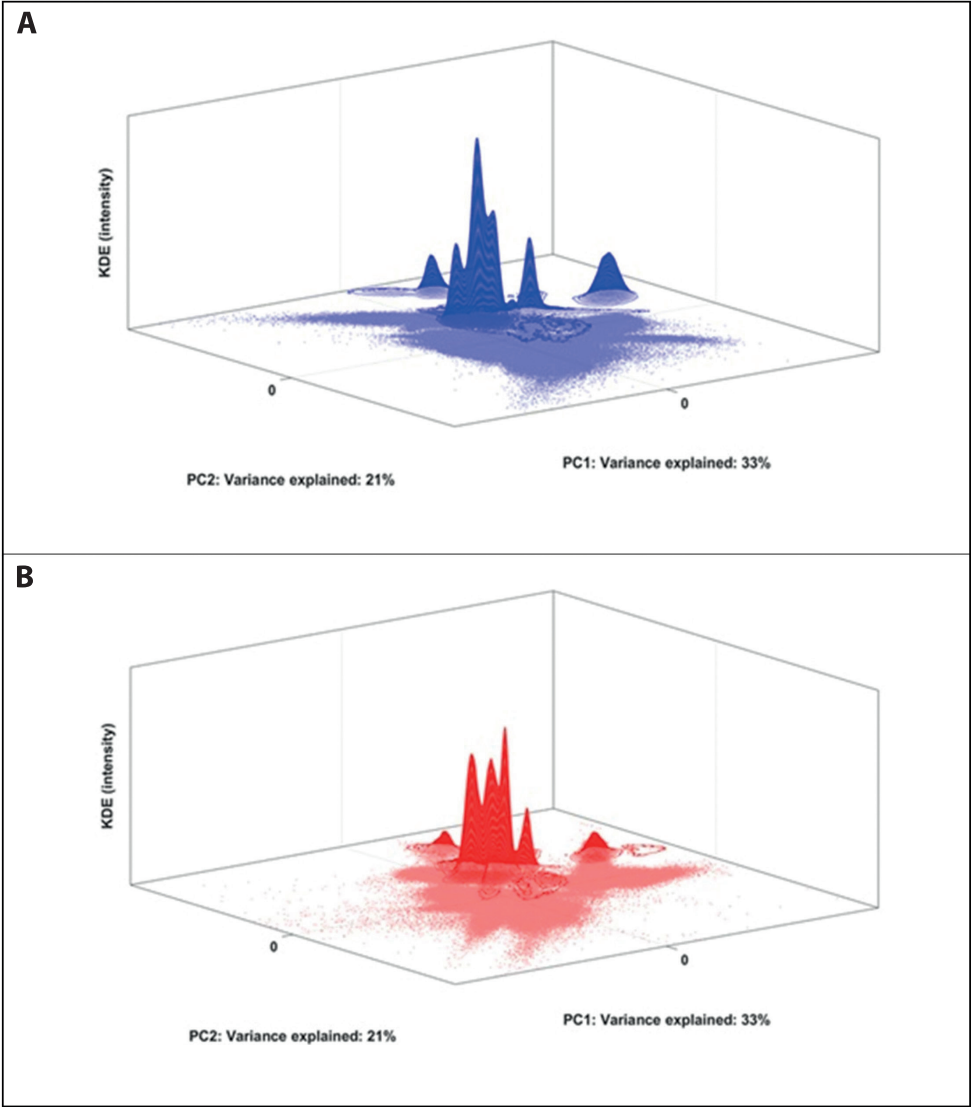
Asthma is a chronic inflammatory respiratory disease that affects the airways. Symptoms range from wheezing and coughing to bronchospasms and dyspnoea. Symptom severity differs among patients, with some barely experiencing problems in daily life, and others having frequent exacerbations, requiring hospitalization<sup>30</sup>. We show here how ECLIPSE may be used to model the asthma-specific immune response associated to differential expression of immune markers on white blood cells (WBCs).

After arcsinh transformation (cofactor 5), the data were mean-centred and scaled over the controls according to Equations S2b and S5b (Supplementary Material I contains more detailed information on the data pre-processing), respectively.

As a result of the validation step, we used a two-component Control Model for this, the PCs of which explain 33% and 21%, respectively, of the variability in surface marker expression of the 10 control individuals. Figure 6a-b shows the scores and the resulting kernel estimates of control individuals and patients.

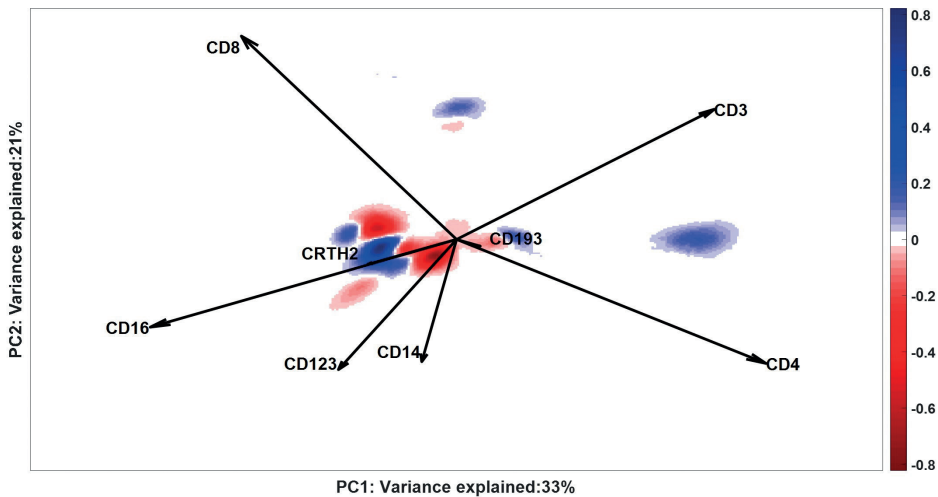
The large number of cells and the considerable differences in expression between the cells make the score distributions of patients and controls complex and difficult to interpret, although they are visually distinguishable. The complexity is due to the considerably larger heterogeneity in surface marker expression that can be expected in the full WBC compartment, compared to the gated neutrophils of LPS. In addition, the effects on the immune system due to asthma may be more subtle than the systemic response to LPS. Transcending the single-cell level into the kernel probability density estimates, shows how the multiple cell populations result in multimodal kernel estimates, both in the

control and patient group. The high maxima in the center of both control space plots correspond to the neutrophils, the largest constituent of all white blood cells. In addition to the changes in the neutrophils upon asthma immune response, comparing both KDEs already shows that the response involves several other WBC sub-populations as well.



**Figure 6.** A) Scores plot of control cells of all the control individuals in the Control Space, plus 3D-contour of the PDF of the control scores estimated with KDE; B) Scores plot of all the responder cells projected in the Control Space, plus 3D-contour of the PDF of the responder scores estimated with KDE. The kernel estimates display PDFs with multiple maxima; this suggests that multiple cell-subpopulations are present in the data.

The DbD plot (Figure 7) shows which cell populations are over- or underrepresent in patients with severe asthma, when compared to non-asthmatic controls. The DbD plot displays several negative regions (red) with a higher chance of cells with a surface marker co-expression typical of patients; cells in positive areas (blue) are more probable to be found in control samples.

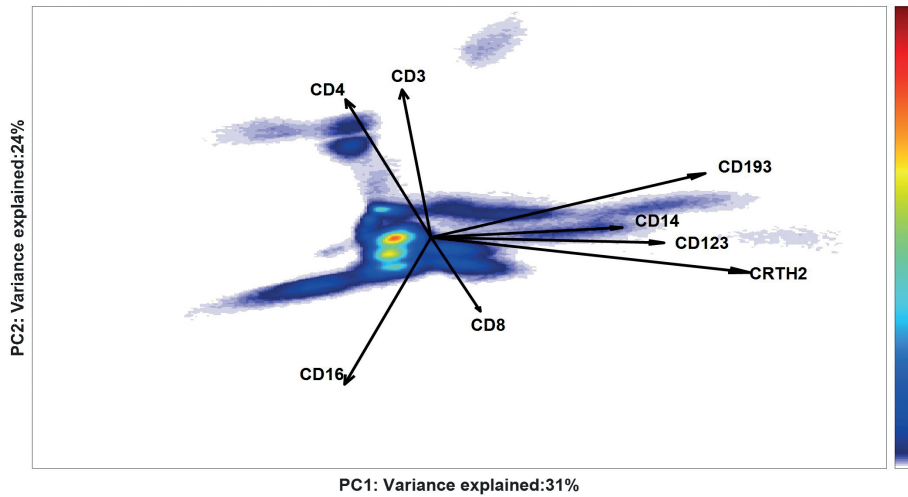


**Figure 7.**

Density plot of Difference between Densities: the negative intensity (red) shows the locations where a larger fraction of cells scores for patients than for control individuals; positive intensities (blue) contain cells with larger fractions of normal cells. White areas either contain no cells or have similar fractions in both groups of individuals. The axes report the variance explained by the SCA Control Model for the control individuals, on which the DbD is calculated. Loadings of the SCA Control Model are represented as vectors: loadings with similar directions indicate positive co-expression (like for CD16, CRTH2 and CD193, CD4) and with opposite directions indicate negative (like for CD3 and CD16) co-expression of both surface markers. Orthogonal vectors indicate surface markers that express independently, e.g. CD8 and CD123.

Cells within negative regions, together with the cells with high residuals are then selected for further analysis within a response-specific ECLIPSE model. The original compensated data of these cells, after arcsinh transformation (cofactor 5), are mean-centered and scaled across all the individuals (Eq. S2-S3, Online Supplementary Material I). The ECLIPSE algorithm models the variability of all the cells particularly found in the patients, as shown in Figure 8.



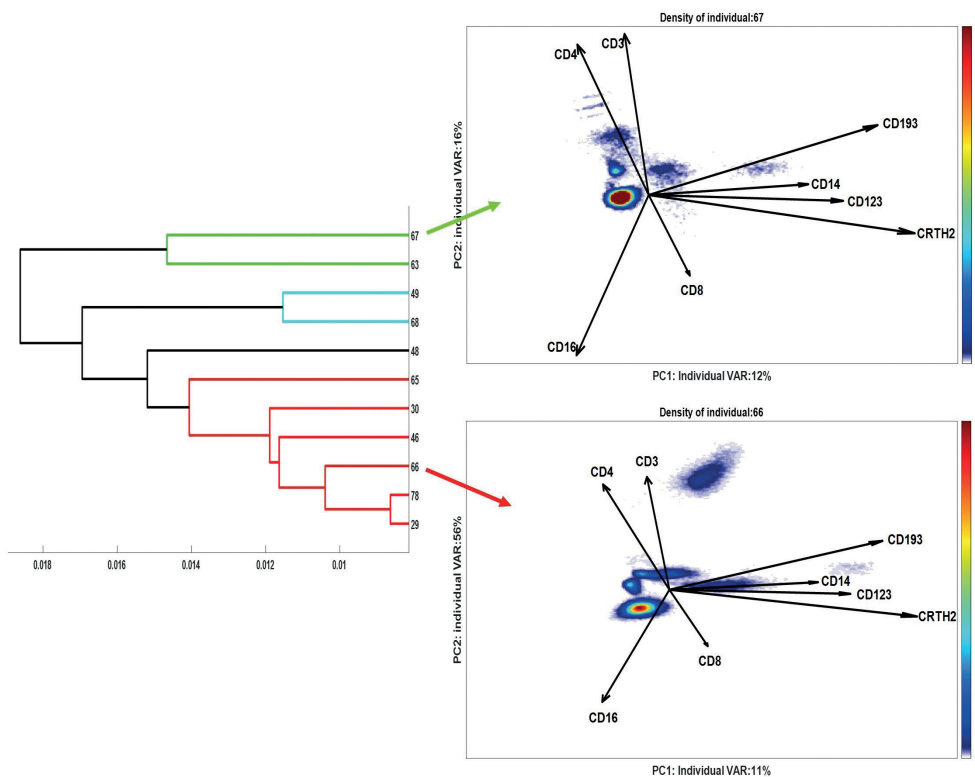


**Figure 8.**

General ECLIPSE model built on all non-normal cells of the 11 asthma patients. A density estimation of the cell scores distribution is shown, where density is quantified from low (blue) to high intensity (red). Multiple cell sub-populations with different intensities can be observed, mainly differing in the co-expressions of CD193, CD14, CD123, CRTH2 (along the horizontal axis), and by simultaneous positive co-expression of CD4, CD3 with a negative co-expression of CD16 (on the vertical axis). Both axes report the variance explained by each component of the two PCs in the SCA model for all the patients. The loading vectors greatly differ from the loadings of the Control Model because they explain the variability of only the not-normal cells.

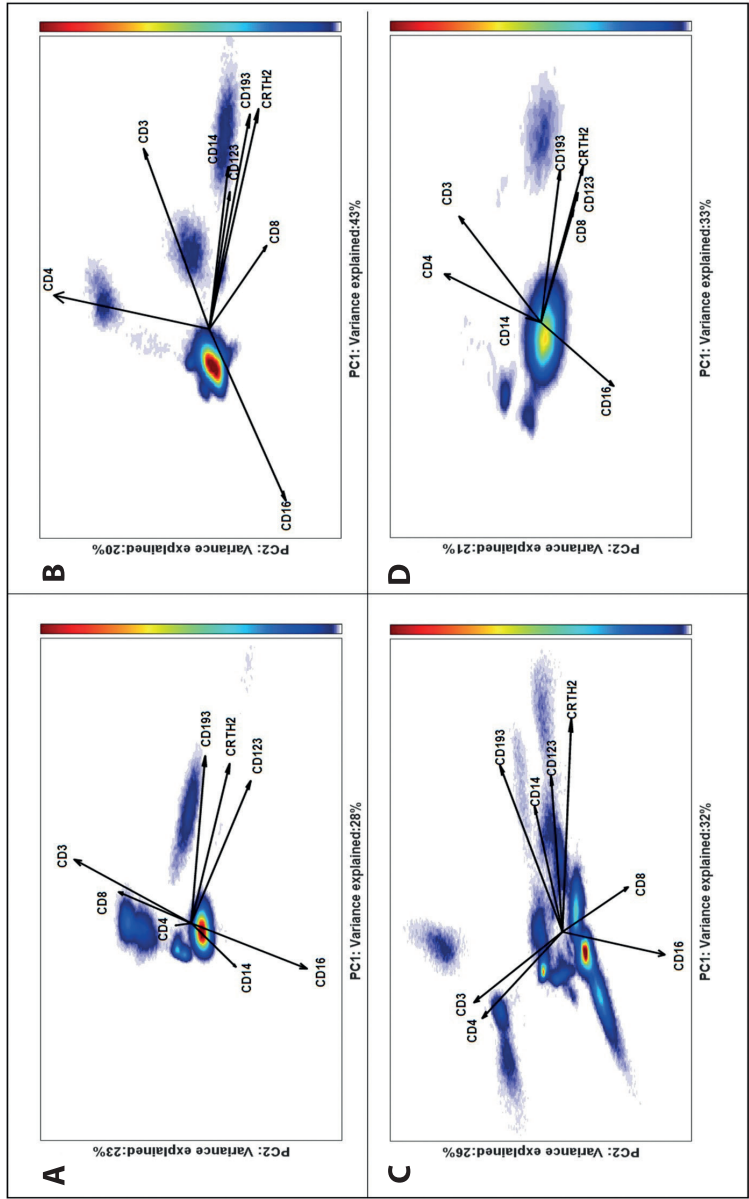
From this ECLIPSE model, scores of each individual asthma patient in the ECLIPSE space can be individually extracted and inspected (Table S3, Supplementary Material II, which shows the scores in the general models for all patients). The output of this general ECLIPSE model may be used as input for a clustering analysis, to stratify individuals in an automated and reproducible fashion. Hierarchical clustering by average linkage to the distance of the kernel estimates provides a dendrogram of the patients, visualized in Figure 9 (left panel), which shows clear subgrouping of this small cohort of asthma patients. Asthma patients may be grouped into three or four clusters, specifically three containing multiple patients (blue, green and red) and one single responder in a separate cluster (black).

Figure 9, right panel, shows the scores in the ECLIPSE general model of two patients grouped in different clusters (green and red) and thus reveals differences in the cell populations present. However, the variance in marker expression of the individual patients was not always well described by the general ECLIPSE model: only 28% of the total variance of responder ID #67 was explained in this way. To investigate the cell variability in surface marker expression that is relevant for a specific cluster, we built four partial cluster-specific ECLIPSE models, on each separate cluster of individuals (Figure 10). As expected for a heterogeneous immune response, the length and the orientation of the loadings are different between the 4 models.



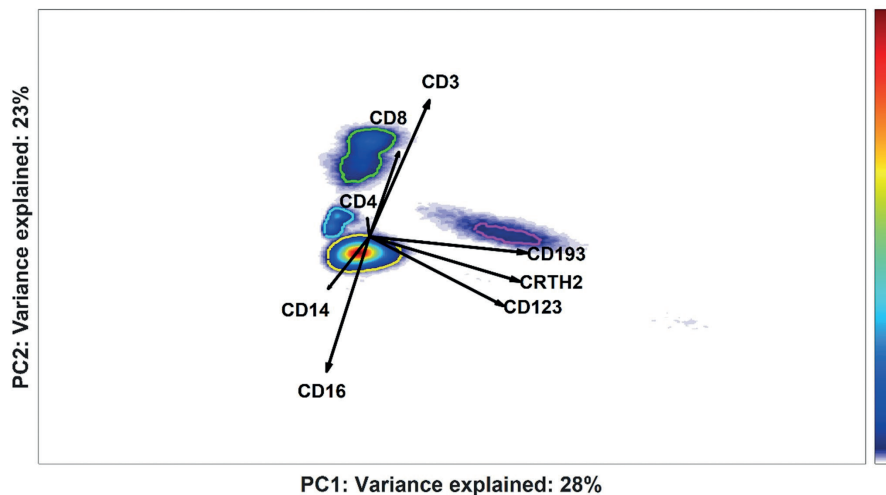
**Figure 9.**

Dendrogram (left panel) of the asthma patients, obtained by average linkage to the distances between the Kernel Density estimates of their ECLIPSE scores. Four different patient clusters are coloured as different dendrogram leaves. The Figures on the right panel show the density cells scores distribution of two representative individuals (#67 and #66) from different patient clusters.



**Figure 10.**

Partial cluster-specific ECLIPSE response models. A) Density estimation of cell scores distribution of patients #63 and 67 grouped in the green cluster; B) Density estimation of cell scores distribution of patients #49 and 68 grouped in the blue cluster; C) Density estimation of cell scores distribution of patients #65, 30, 46, 66, 78 and 29 grouped in the red cluster; D) Density estimation of cell scores distribution of patient #48 (black individual). Both axes report the variance explained in total for the patients of each group by the respective PCs of the partial cluster-specific ECLIPSE Models. The 4 models are qualitatively comparable in terms of percentage of variance explained. However, great difference in the loadings indicates that the type of variation accounted by each model is non-identical: each patient cluster is defined by a different and characteristic combination of cell populations.



**Figure 11.**

Partial cluster-specific ECLIPSE model built on the variability of the non-normal cells of individual #63 and #67 (green cluster in Figure 8). Density estimates of cell score distributions of both patients are shown. The intensity of the estimates is indicated with the colour bar on the right, from low (blue) to high intensity (red). The most densely populated regions are highlighted with coloured gates.

The most densely populated regions of the partial ECLISPE models can be automatically gated; the results for the partial ECLISPE model of individuals #63 and #67 from the green cluster are shown in Figure 11. One of the cell populations (gated in green) is identified as CD8+ T cells, due to its co-expression of markers CD3 and CD8. The partial ECLISPE models built on the other clusters do not show a predominant CD3+CD8+ T-cell population, in the first two PCs (Figure 10). On the contrary, in the partial models of the blue and the red clusters CD3+CD4+ T-cells had a more important contribution to the variance, as indicated by the long loadings in these models and shown in the Online Supplementary Material IV for the cells of the red clusters. Another characteristic population that we observed in multiple patient clusters most likely were basophils, as indicated by a relatively high expression of CD123, CD193 and CRTH2 (Figure 11, magenta gate). These latter cells were discriminatory only in some phenotypes of asthma. The mechanism behind the variation in recruitment of cellular populations to the peripheral blood in severe asthma patients will be examined in future studies and goes beyond the scope of this article.

For comparison, we performed Citrus and viSNE on the asthma dataset (Online Supplementary Material III)

The considerable heterogeneity associated with the asthma study results in a reduced classification accuracy in the Citrus analysis. In fact, the highest achieved accuracy corresponded to 25% of misclassified samples (Figure S15). This is obtained with a model which identifies the four most discriminant cell clusters. Two of these clusters were distinctive for the control group, with CD16+CD123dimCD8dim and CD3–CD14– expression levels and CD4+ T cell phenotype (CD3+CD4+), respectively. The over-representation of these cells in the control individuals, against the asthma group, was highlighted also by the Difference between Densities plot (Figure 7). The blue regions here show, among other cell populations, the presence of CD4+ T cells and CD16+CD3– cell, which are more likely to be found in the control individuals. After the removal of normal cells, however, we could observe the presence of CD4+ T cells also in the partial ECLIPSE model of red and blue clusters (Figure 10b, c). The patients belonging to these clusters, identified by ECLIPSE, might have CD4+ T cells with a differential markers expression compared to the controls, which is emphasized by the difference in the angles of the loadings in the more focused models. Both other clusters identified by Citrus were more represented in the asthmatic patients (Figure S17). In this case, it was harder to identify clear phenotypes, mainly because of very heterogeneous expression patterns. In fact, multiple peaks per marker suggested the presence of multiple cell subtypes within a same cluster. Unfortunately, we were unable to identify which cell types are present in this cluster, since the information about the co-expression of markers is lacking in the Citrus results.

The viSNE map obtained by the analysis of the asthma data showed substantial overlap of cells between the control and the asthmatic patients (Figure S18). Most overlapping regions are associated to CD4+ T cells; CD8+ T cells; CD16+ cells, which might be identified as neutrophils; and CD14+ cells, which might be identified as monocytes. We compared the map showing only the individuals #63 and #67 (Figure S19) to the ECLIPSE partial model of Figure 11. The phenotypical marker pattern observed is similar to the cell phenotypes shown in the ECLIPSE partial model. The CD16+ cells might be identified as neutrophils; CD123+CD193+CRTH2 cells as basophils; while CD3+CD8+ cells and CD3+CD4+ cells might be assigned to CD8+ T cells and CD4+ T cells, respectively.

## Discussion

The ECLIPSE methodology allows comprehensive exploration of multi-dimensional MFC data specifically for cells that respond upon disease or other perturbations of cellular systems that may be analyzed by flow cytometry. The vast number of cells in responders that are indistinguishable, 'normal', from cells present in healthy individuals hampers such an analysis. The methodology demonstrated removes such cells, accurately taking

into account the multiset and multivariate features of MFC data. Rather than making the selection of cell subsets based on the expression of marker pairwise combinations (as for manual gating), ECLIPSE performs an automated gating of response-specific cells based on continuous and multivariate variability in surface marker expression compared to the multivariate variability of healthy cells in a control group. This then results in a comprehensive overview and a quantifiable indication of the immune response in terms of specific expression and/or co-expression of cellular markers. Using ECLIPSE for flow cytometry data analysis will result in more precise and elaborate knowledge about the cells specific for the studied host response, in a fast and un-biased manner.

ECLIPSE will be useful in MFC analyses of cells activated upon response, such as in the LPS study. It will also facilitate the analysis of whole blood in individuals with responses of unknown etiology. The latter is illustrated by the analysis of the asthma patients, revealing heterogeneity in their response to the disease.

In the LPS analysis, we showed how ECLIPSE was able to increase resolution and interpretability of the endotoxin-induced immune response based on the relationships among all the markers, due to removal of 'normal' cells. In particular, the general ECLIPSE model shows expression (above the average) of the markers CD62L and CD16 for the LPS-induced immune response associated cells in the responder individuals. The same biplot representation indicated a co-expression of markers CD11c, CD11b, CD69 for the cells that highly express CD16. In this way, the algorithm offered a closer look at such a minority of cells, defining their characteristics in terms of phenotypic markers.

Additionally, visualizing individual responder cell distributions provided the exploration of the personalized response to the LPS and focus on diversity among responders. Small differences in the continuum of surface marker expressions may suggest variability in the time of reaction to the endotoxin among the different individuals.

The asthma study shows how ECLIPSE may be used as a flexible method to characterize heterogeneity of poorly understood diseases among individual patients. In such analysis, a general ECLIPSE model was shown to be insufficient in reproducing the heterogeneous response across all individuals. However, quantitative comparison of cell distributions between individuals in the general ECLIPSE model allows stratification into different groups, which can be further analyzed by partial ECLIPSE models. This allows dedicated interpretation of the cell variability associated to the response within each cluster of individuals, allowing a detailed group- or individual-based interpretation of differences in diseases profiles. The applicability and added value of such profiles in clinical decision-making should be examined in a prospective study. An individual personal response

model that is exclusively based on the activated cells, may provide an optimal view of a unique profile in the individual patient.

The advantage of increased resolution in individually expressed cell (sub-)populations however comes with the drawback that personal models cannot be compared between individuals as they lack common ground in the model scores and loadings, such that individual and general ECLIPSE models should always be interpreted simultaneously. Full interpretation of the different immune responses across the entire patient cohort therefore requires systematic exploration firstly in the common ECLIPSE model; the findings from this can then be explored in the clustered patient profiles to allow detailed stratification of patients in future studies.

### **ECLIPSE compared to classification by Citrus**

The quantitative objectives of ECLIPSE, Citrus and viSNE are distinctive for each method, which makes the results derived from their models inherently slightly different and the comparison between the methods not always as straight-forward. In the case of a binary classification problem, Citrus aims to find the differences between an assumed homogeneous control group and a similarly homogeneous group of diseased samples. ECLIPSE equally assumes a homogenous control group, but allows for heterogeneity in the cell variability between diseased samples; it can even explicitly model this heterogeneity in the partial or individual model. This also implies that ECLIPSE does not provide classification accuracies and other metrics that can be directly compared to those of Citrus. Both methods can, however, be compared by the cell subpopulation marker profiles that are distinctive between phenotypes (Online Supplement Material III). In the case of the LPS dataset, the phenotypes identified by Citrus were highly comparable to the marker profiles of the cell populations found in the ECLIPSE space (Figure 4). Citrus additionally identified clusters distinctive for the control group, corresponding to mature neutrophils. In fact, Citrus focuses on finding features that are most predictive for the classification endpoint (e.g. controls vs responders). Therefore, also cell populations that are more abundant or specific for a control group may be found to be best predictors. These cells are not considered of primary interest when studying an immune response by the current implementation of ECLIPSE.

We show here that filtering of the 'normal' cells with the ECLIPSE-based filter method, positively affects discrimination between control and responder samples, as indicated by the lower error rate also in Citrus. On top of that, removal of normal cells resulted in a considerably lower number of features necessary for the optimal Citrus classification. However, this also provides a fundamental distinction between Citrus and ECLIPSE: the former aims to find an as low as possible number of features to classify samples, and shows the cell clusters distinctive for the control or response group with the marker expressions

in histograms. ECLIPSE provides a comprehensive set of single cells that contributes to the distinction of control and responder samples. Features that are considered redundant by Citrus, because they do not improve the prediction, may be needed to describe patient-specific aspects of the response. The only distinctive phenotype found for the LPS responders by Citrus consisted of pre-mature neutrophils (as shown in Figure S10), while our ECLIPSE analysis, additionally to the presence of pre-mature neutrophils, showed a diversity in mature neutrophils associated to a personalized response across the responders.

These results indicate that Citrus is less-well suited for identifying differences between individuals belonging to the same response group and it will fail in identifying putative personalized cell subsets which are not representative of the whole cohort but only of few samples. This explains why Citrus did not provide satisfying classification accuracies for a heterogeneous disease as asthma, as it failed to classify 25-30% of the samples correctly. Although some specific marker profiles were found for the asthma group, such high cross-validation error rate indicates that the identified clusters are not consistent across the entire patient cohort. Therefore, they do not offer an exact overview of the variability present in the asthmatic patients.

Moreover, the clusters which are either predictive for the asthma group or the control group, were not easy to interpret. This is due to the way the Citrus results are visualized. When multiple cell populations are present within one cluster, the single marker histograms do not give information about the co-expression per cell population.

The limits of Citrus in handling heterogeneous dataset was also shown, when it was applied to the synthetic flow cytometry-like data of cell populations characterized by differences in terms of abundance and expression levels of different markers (included in Online Supplementary Material V). Here, Citrus could only identify cell populations with different relative abundance in the controls compared to the response group; however, it disregarded the relevant individual response-specific cell populations. ECLIPSE is better suited for exploration of the within-group differences, as it reviews individual patient variability. Additionally, we showed how ECLIPSE was able to detect a small cell population, which was only present in one individual. Moreover, removal of normal cells was necessary to correctly model the variability of all the responder cells, as was shown by the orientation of the loadings in the ECLIPSE model (Figure S38).

### **ECLIPSE compared to cell visualization by viSNE**

Both viSNE and ECLIPSE methodologies reduce the multi-dimensional flow cytometry data into fewer, commonly two, dimensions. In both cases, this provides a 2D-map



that retains only part of the variability in the measured data. However, viSNE provides no information about the variability that remains unexplained by the analysis, while in ECLIPSE, the percentage of unexplained variance is displayed and, if needed, this variance may be described by additional PCs without altering the already fitted part of the model. Due to the different algorithm used for the dimensionality reduction, in principle, no heuristic parameter can be used for the quantitative comparison of the two methods. However, the Differences between the cell variability expressed in the ECLIPSE space and in the viSNE maps may be qualitatively compared and interpreted.

The viSNE analyses performed on the two studied datasets showed a big overlap of the responder cells with the control regions, as expected for both studies. Such overlap, however, can hamper the identification of cell populations specific for the response. Secondly, in viSNE, we cannot affirm whether these normal-like cells are overrepresented compared to the controls and thus interesting to describe the immune response. The Difference between Densities (DbD, step 4), performed by ECLIPSE, takes into account both situations that might occur because of immune response: deviation from normal cell marker variability and higher relative abundance of normal immune cells.

An inherent drawback of viSNE is the need to downsample the original data, due to limit in computational power. Populations that are more abundant will be retained in the subsampled data at the expenses of rarer cells. In such situations, the ECLIPSE-based filtering of normal cells (which is a more specialized way of subsampling), focusing on the cells associated to the perturbation of the immune system, is more beneficial. We have shown how a less crowded viSNE map with more intuitive interpretation of responder-specific information can be constructed when combined with a responder-specific filter provided by ECLIPSE. An example of such a viSNE map for the LPS study is compared with the analysis run on the complete dataset (Online Supplementary Material III). Phenotypical patterns distinctive for the cells found in the challenged individuals are better appreciated. In fact, viSNE performed on the ECLIPSE subsampled data reveals three distinctive populations, which are mostly differentiated by CD16 and CD62L marker expressions. Another advantage of ECLIPSE is that the co-expression among the markers is displayed in one single biplot, while in viSNE the single-cell representation cannot be directly linked to the co-expression of the markers measured.

### Multiset pre-processing

Another advantageous hallmark of the ECLIPSE algorithm, next to the removal of normal cells, is the distinctive multiset pre-processing strategy. As described in the Online Supplementary Material I, the novel pre-processing strategy adopted by the algorithm is able to account for differences in terms of number of cells measured per individual,

such that each individual equally contributes to the model. Unwanted variation present between the individuals is also taken care of by the pre-processing steps, which helps to systematically remove variability unrelated to the problem while retaining relevant biological information. We have also shown in Figure S14 (Online Supplementary Material III) how the elimination of individual variability done by ECLIPSE was beneficial for analysis with the viSNE algorithm, which might have been affected by sample-to-sample variability. Such multiset pre-processing was also beneficial to outperform the predictive ability of the discriminant model obtained with Citrus (Figure S7).

### **Multivariate advantage**

The results obtained with the two different studies, show how the multivariate advantage may be used to unravel the correlations among all the measured surface markers and to distinguish cell subsets from one another. Other methods like SPADE and viSNE lack this view of mutual correlation among surface markers; any information on surface marker co-expression can only be obtained from a post hoc model interpretation, analyzing multiple single marker heatmap plots. The SCA loadings of ECLIPSE provide a direct compass-like representation of the surface marker co-expression on specific subsets of cells that may be much easier interpreted in terms of surface marker co-expression than, for example, viSNE maps. A direct interpretation of marker co-expression, compared to the qualitative evaluation necessary for non-linear models such as viSNE, justifies the use of linear SCA. Non-linear relationships are, anyway, still visible in the ECLIPSE model. In fact, the LPS-specific variability appears in the ECLIPSE space as a 'bean-like' continuum (Figure 2) across co-expression of both CD16 and CD62L.

Although FLOOD offers an explicit view of the correlations among surface markers through a similar biplot representation, this method lacks the filter aspect (removal of 'normal' cells) of ECLIPSE. This makes the FLOOD loadings considerably less informative than those from ECLIPSE, because the former also describe 'normal' cells to a large extent, which dilutes their response-specific information. This demonstrates the direct benefit of ECLIPSE for the comprehensive and specific interpretation of the immune response across a range of responders/patients. In addition, the use of Difference between Densities instead of the convex hull used in FLOOD, for determining the boundaries of the filter to gate out 'normal' cell brings several advantages. While the control data cannot be described using only a single convex shape, the Difference between Densities gating step can easily discern different populations. In addition, the DbD is able to take into account not only deviation from the normal cell surface variability but also changes in the fraction of normal cells, that might increase due to the immune response.

## Conclusion

ECLIPSE provides a filter for 'normal' cells present in individuals out of homeostasis, based on the multivariate co-expression among all the cellular markers measured. As a result, the algorithm allows the identification of detailed cellular information particularly present in response individuals (e.g. patients) completely independent of prior expertise knowledge, but through automatic and data-driven cell gating. ECLIPSE is specifically valuable for the identification of unbiased cellular information present in flow cytometry data of individuals with a condition of unknown etiology. The ECLIPSE method provides a higher resolution on response-specific cells that might be hidden among an overwhelming preponderance of normal cells. In future application of the ECLIPSE algorithm, we may expect that ECLIPSE could benefit researchers and clinicians handling intricate data, such as data containing rare cell subsets that only occur in diseased samples.

In addition, the versatility of the algorithm may allow the integration of the multidimensional ECLIPSE filter into other multivariate MFC data analysis methods, in order to enable a better visualization and an outperforming prediction accuracy.

## Software and Data availability

ECLIPSE is implemented in MATLAB®R2016b and the script is available for download at the <https://github.com/RoelBouman/ECLIPSE>.

The data sets used in the ECLIPSE analyses are available at <https://www.ru.nl/science/analyticalchemistry/research/data-analytical-chemistry/>

## References

- 1 Biancotto, A., Fuchs, J. C., Williams, A., Dagur, P. K. & McCoy, J. P., Jr. High dimensional flow cytometry for comprehensive leukocyte immunophenotyping (CLIP) in translational research. *Journal of immunological methods* 363, 245-261, doi:10.1016/j.jim.2010.06.010 (2011).
- 2 Perfetto, S. P., Chattopadhyay, P. K. & Roederer, M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol* 4, 648-655 (2004).
- 3 De Rosa, S. C., Brenchley, J. M. & Roederer, M. Beyond six colors: A new era in flow cytometry. *Nat Med* 9, 112-117, doi:http://www.nature.com/nm/journal/v9/n1/supinfo/nm0103-112\_S1.html (2003).
- 4 Bendall, S. C. *et al.* Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science (New York, N.y.)* 332, 687-696, doi:10.1126/science.1198704 (2011).
- 5 Bashashati, A. & Brinkman, R. R. A Survey of Flow Cytometry Data Analysis Methods. *Advances in Bioinformatics* 2009, 19, doi:10.1155/2009/584603 (2009).
- 6 Lugli, E., Roederer, M. & Cossarizza, A. Data analysis in flow cytometry: the future just started. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* 77, 705-713, doi:10.1002/cyto.a.20901 (2010).
- 7 Amir el, A. D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology* 31, 545-552, doi:10.1038/nbt.2594 (2013).
- 8 Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nature biotechnology* 29, 886-891, doi:10.1038/nbt.1991 (2011).
- 9 Bruggner, R. V., Bodenmiller, B., Dill, D. L., Tibshirani, R. J. & Nolan, G. P. Automated identification of stratifying signatures in cellular subpopulations. *Proceedings of the National Academy of Sciences of the United States of America* 111, E2770-E2777, doi:10.1073/pnas.1408792111 (2014).
- 10 Costa, E. S. *et al.* Automated pattern-guided principal component analysis vs expert-based immunophenotypic classification of B-cell chronic lymphoproliferative disorders: a step forward in the standardization of clinical immunophenotyping. *Leukemia* 24, 1927-1933, doi:10.1038/leu.2010.160 (2010).
- 11 Jansen, J. J. *et al.* FLOOD: FLOW cytometric Orthogonal Orientation for Diagnosis. *Chemometrics and Intelligent Laboratory Systems* 151, 126-135, doi:https://doi.org/10.1016/j.chemolab.2015.12.001 (2016).
- 12 Jolliffe, I. T. *Principal Component Analysis, Second Edition*. Vol. 98 (Springer, 1986).
- 13 Engel, J., Blanchet, L., Engelke, U. F. H., Wevers, R. A. & Buydens, L. M. C. Towards the Disease Biomarker in an Individual Patient Using Statistical Health Monitoring. *PLOS ONE* 9, e92452, doi:10.1371/journal.pone.0092452 (2014).
- 14 Hilvering, B. *et al.* Diagnosing eosinophilic asthma using a multivariate prediction model based on blood granulocyte responsiveness. *Allergy* 72, 1202-1211, doi:10.1111/all.13117 (2017).

- 15 Kosugi, Y., Sato, R., Genka, S., Shitara, N. & Takakura, K. An interactive multivariate analysis of FCM data. *Cytometry* 9, 405-408, doi:10.1002/cyto.990090419 (1988).
- 16 Lugli, E. et al. Subject classification obtained by cluster analysis and principal component analysis applied to flow cytometric data. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* 71, 334-344, doi:10.1002/cyto.a.20387 (2007).
- 17 Timmerman, M. E. & Kiers, H. A. L. Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika* 68, 105-121, doi:10.1007/BF02296656 (2003).
- 18 Green, P. J. & Seheult, A. H. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 37, 120-121, doi:10.2307/2347507 (1988).
- 19 Duong, T. & Hazelton, M. Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics* 15, 17-30, doi:10.1080/10485250306039 (2003).
- 20 Pyne, S. et al. Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences* 106, 8519-8524, doi:10.1073/pnas.0903028106 (2009).
- 21 Hall, P. & Marron, J. S. Lower bounds for bandwidth selection in density estimation. *Probability Theory and Related Fields* 90, 149-173, doi:10.1007/bf01192160 (1991).
- 22 Park, B. U. & Marron, J. S. Comparison of Data-Driven Bandwidth Selectors. *Journal of the American Statistical Association* 85, 66-72, doi:10.1080/01621459.1990.10475307 (1990).
- 23 Sheather, S. J. J., M. C. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 53, pp. 683-690 (1991).
- 24 Botev, Z. I., Grotowski, J. F. & Kroese, D. P. Kernel density estimation via diffusion. *Ann. Statist.* 38, 2916-2957, doi:10.1214/10-AOS799 (2010).
- 25 Yang, C., Duraiswami, R., Gumerov, N. A. & Davis, L. in *Proceedings Ninth IEEE International Conference on Computer Vision*. 664-671 vol.661.
- 26 Hall, P. & Wand, M. P. On nonparametric discrimination using density differences. *Biometrika* 75, 541-547, doi:10.1093/biomet/75.3.541 (1988).
- 27 Liu, Y. J. et al. Angle distribution of loading subspaces (ADLS) for chemical rank estimation in three-way analysis. *Chemometrics and Intelligent Laboratory Systems* 152, 146-156, doi:10.1016/j.chemolab.2015.11.006 (2016).
- 28 Besse, P. PCA stability and choice of dimensionality. *Statistics & Probability Letters* 13, 405-410, doi:https://doi.org/10.1016/0167-7152(92)90115-L (1992).
- 29 Pillay, J. et al. Functional heterogeneity and differential priming of circulating neutrophils in human experimental endotoxemia. *Journal of leukocyte biology* 88, 211-220, doi:10.1189/jlb.1209793 (2010).
- 30 Vijverberg, S. J. et al. Clinical utility of asthma biomarkers: from bench to bedside. *Biologics : targets & therapy* 7, 199-210, doi:10.2147/btt.s29976 (2013).
- 31 Pillay, J. et al. A subset of neutrophils in human systemic inflammation inhibits T cell responses through Mac-1. *The Journal of clinical investigation* 122, 327-336, doi:10.1172/jci57990 (2012).

# Supplementary Material

## Contents

Supplementary Material I	187
Multi-set control/response structure	187
Step0: Data pre-processing	188
Supplementary Material II	192
Additional results of ECLIPSE on the LPS dataset	192
Additional results of ECLIPSE on the asthma dataset	196
Supplementary Material III	200
Citrus and viSNE analyses on the LPS and asthma study dataset	200
Results of Citrus on LPS data	200
Results of viSNE on LPS data	205
Results of Citrus on the asthma data	210
viSNE analysis of asthma data	212
Supplementary Material IV	216
Manual sequential gating of the asthma red cluster and projection into ECLIPSE space	216
References	231

## Supplementary Material I

### Multi-set control/response structure and pre-processing of Multicolour Flow Cytometry Data

#### Multi-set control/response structure

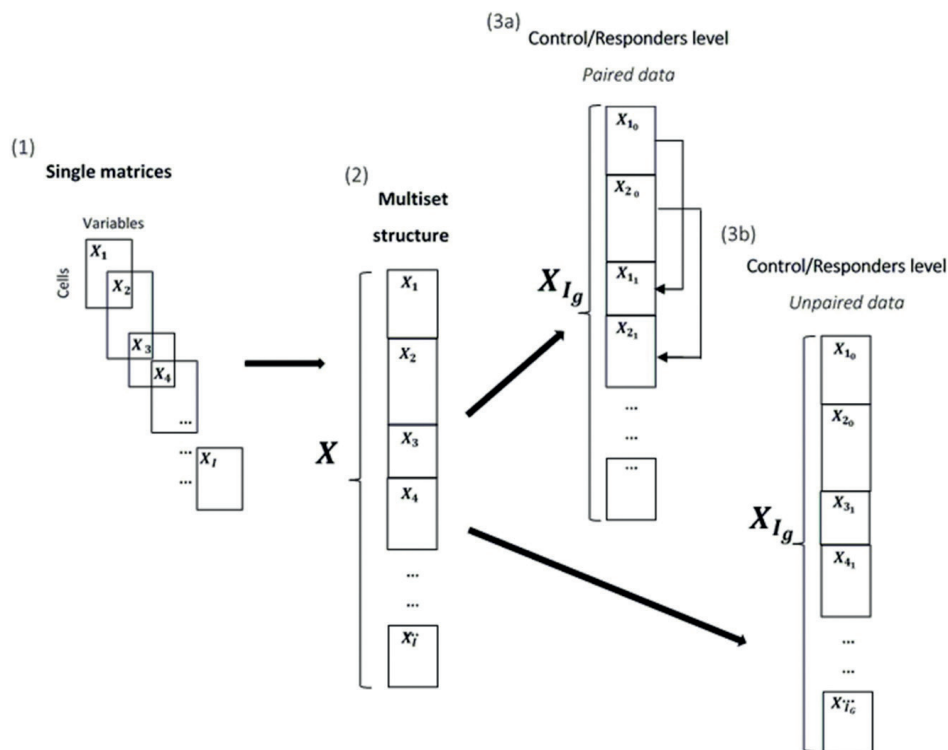
MFC data quantify the presence of fluorescently labelled cell properties, such as surface marker expressions, of a large number of single cells. These cells are collected per individual sample and the number of collected cells may vary considerably between cell 'sets'. Experiments often compare a 'Case' (or responder) group to samples obtained from 'Control' individuals. The Control individuals present surface marker expression on their cells typical for individuals that do not display the response of interest, while the individuals in the 'responder' group present cell populations with surface marker expressions characteristic of the immune response studied.

Figure S1 shows the possible arrangements of the MFC data by considering three different levels. (1) Single matrices, which hold the cell set measured per individual; comprehensive and comparative analysis of different samples require that the same surface markers are measured across all individuals. (2) Single matrices are concatenated column-wise leading

to  $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_I \end{bmatrix}$  of size  $\sum_1^I N_i \times J$ , where  $N_i$  is the number of cells of the  $i^{\text{th}}$  individual and

$J$  corresponds to the markers measured,  $1 \dots j \dots J$ . This resulting matrix  $\mathbf{X}$  therefore consists of a 'multiset', where each set contains the cells of one individual; each row within each set represents measurement of one single cell. Each individual might be either a control or a responder, the information of the respective group is displayed in the level (3) of the multiset structure with the matrix  $\mathbf{X}_{i,g}$  size  $N_{i,g} \times J$ .

The index  $g = 0$  for the control and  $g \geq 1$  for the responder groups that might correspond to different diseases or subtypes of the same disease. If all responder individuals are drawn from a population with the same disease, then  $g$  will assume values 0 and 1, for the control group and responder group, respectively. In some cases, experiments are paired, which means that the same person is followed and analysed before and after an immune response.

**Figure S1.**

(1) Single data matrices representing measurement per individual, (2) Multiset data arrangement obtained by linking the matrices column-wise, (3a) Control/Responder differentiation of the multiset structure, with paired data, (3b) Control/Responder differentiation of the multiset structure, with unpaired data

In this case the index  $i$ , representing the individuals, is not unique (3a). However, the two Flow Cytometry case studies we consider in this paper both consist of unpaired individuals, indicated by different  $i$  (as in 3b).

### **Step 0: Data pre-processing**

Data pre-processing is a very important aspect of chemometric data analysis.<sup>1</sup> It aims to remove variability in the data that is unrelated to the problem under study, while retaining the experimentally relevant information. In Flow Cytometry, such irrelevant variability might derive from instrumental artefacts due to misalignment of the laser source, baseline drift, laser power variability, interference of fluorescent labels, or uninformative noise coming from low intensity signals. Removing such irrelevant variability in a quantitative and reproducible way facilitates the quest for relevant biomedical information and guarantees that the fluorescents intensities are measures of level of protein expression on the cells.



The first step of the pre-processing consists of transforming matrix  $\mathbf{X}$  using log (or arcsinh) function. In MFC, this type of transformation is essential to cope with the broad dynamic range of emissions between fluorophores. The data matrix may contain negative values, due to background subtraction or to the compensation of overlap between emission spectra of different fluorophores<sup>2</sup>. These require a more dedicated transformation, such as the arcsinh scaling described by Finak et al.<sup>3</sup>. Multivariate analysis of the cell variability then requires centering and scaling of the log or otherwise-transformed matrix  $\mathbf{X}_{\log} = \log_{10}(\mathbf{X})$  Eq. S1. Mean (or median) centering subtracts the column mean (or median) from every element in the column, which removes surface marker expression consistently present across all cells, resulting in the variability in surface marker expressions across the cells. Scaling equalizes the variability of each surface marker across the cells, to allow them to contribute equally to a multivariate model of the data, regardless of the intensity of the used fluorophore or the absolute variability in abundance of every surface marker.

Both centering and scaling need to take into account the multi-set structure of the MFC data (Figure. 1). Several strategies for centering and scaling may accommodate this structure and the quantitative comparison between case and responder individuals<sup>4</sup>.

One strategy is centering and scaling the variables per set, i.e. per individual.

$$(a) \mathbf{m}_{i_g} = \frac{1}{N_{i_g}} \mathbf{1}_{N_{i_g}}^T \mathbf{X}_{\log, i_g} \quad \text{Eq. S2}$$

$$(b) \mathbf{X}_{m_i} = \mathbf{X}_{\log, i_g} - \mathbf{1}_{N_{i_g}} \mathbf{m}_{i_g}^T$$

Where  $\mathbf{m}_{i_g}$  is the mean calculated for the surface markers measurements of the  $i_g$ -th individual,  $i_g = 1_g, \dots, I_g$ , hold in the matrix  $\mathbf{X}_{\log, i_g}$  of size  $N_{i_g} \times J$ ;  $N_{i_g}$  corresponds to the number of cells of the  $i_g$ -th individual;  $\mathbf{1}$  is a column vector of ones with length  $N_{i_g}$ . The resulting mean centered matrix  $\mathbf{X}_{m, i_g}$  is thus scaled using as weight the inverse of  $\mathbf{S}_{i_g}$ , diagonal matrix of size  $J \times J$  holding the standard deviation of the mean-centered surface markers of the  $i_g$ -th individual, calculated as follows:

$$(a) \mathbf{s}_{i_g}^T = \sqrt{\text{var}(\mathbf{X}_{m_i})}$$

$$(b) \mathbf{S}_{i_g} = \text{diag}(\mathbf{s}_{i_g}^T)$$

$$(c) \mathbf{X}_{sc} = \mathbf{X}_{m, i_g} \mathbf{S}_{i_g}^{-1}$$

Eq. S3

Centering (Eq. S2b) and scaling (Eq. S3c) per individual may correct technical individual-specific offsets due to e.g. changes and/or misalignment of laser intensity, sample handling etc. that do not contribute to the biomedical information within the MFC dataset<sup>5</sup>.

Next to pre-processing per individual, there are other pre-processing options, based on the class-level (control or response) or the whole dataset. Considering a single class, specifically the 'control group'  $g = 0$ , as a reference for 'normal' cell variability may highlight those cells that are specific to a response in the other class(es). Centering and scaling based on the control individuals selectively emphasizes deviations from the 'normal' cell variability observed in the case individuals, when means and the standard deviations used in Equations. S2 and S3 are calculated across the control individuals. This requires correcting for the considerably different numbers of cells per analysed individual, in order to avoid the individual with most cells dominating the calculated means and standard deviations. This correction is accomplished by pooling the means of each set into a weighted class mean, according to Eq. S4.

$$\begin{aligned} \text{(a) } \mathbf{m}_0 &= \frac{\sum_{i_0}^{I_0} \mathbf{m}_{i_0}}{I_0} \\ \text{(b) } \mathbf{X}_{\mathbf{m}_0} &= \mathbf{X}_{\log} - \mathbf{1}_{N_{i_g}} \mathbf{m}_0^T \end{aligned} \quad \text{Eq. S4}$$

where  $\mathbf{m}_{i_0}$  is the mean of the log-transformed surface marker intensities of the  $i_0$ -th control individual, calculated according to Eq. S2a, with  $g = 0$ ;  $I_0$  is the total number of control individuals.  $\mathbf{X}_{\mathbf{m}_0}$ , of size of size  $N_{i_g} \times J$ , represents the multiset matrix centered using the class mean of the log-transformed surface marker intensities of the control class.

Also the cumulative control standard deviation may be calculated by Eq. S5:

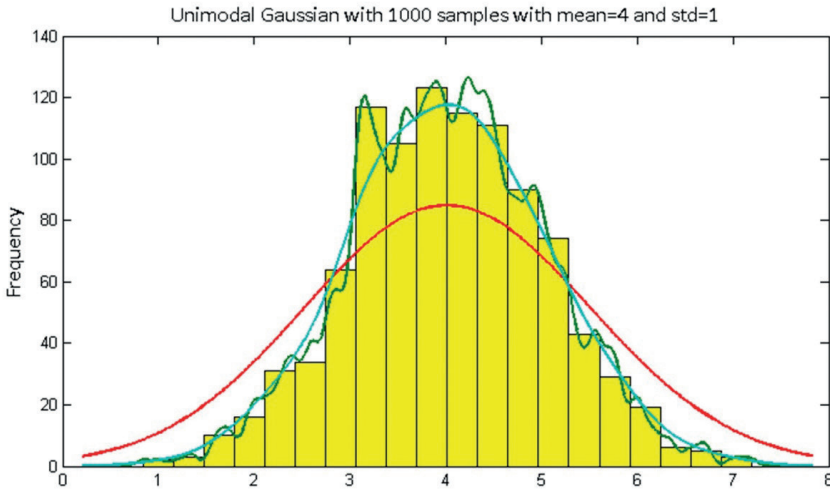
$$\begin{aligned} \text{(a) } \mathbf{s}_0^T &= \sqrt{\frac{\sum_{i_0}^{I_0} \text{var}(\mathbf{X}_{\mathbf{m},i_0})}{I_0}} \\ \text{(b) } \mathbf{X}_{\mathbf{s}_{c_0}} &= \mathbf{S}_0^{-1}(\mathbf{X}_{\mathbf{m}_0}) \end{aligned} \quad \text{Eq. S5}$$

with  $\mathbf{X}_{\mathbf{m},i_0}$  the mean-centered individual control matrix, estimated according Eq. S2b, with  $g = 0$ ;  $\mathbf{S}_0^{-1}$  the diagonal matrix holding the standard deviation  $\mathbf{s}_0^T$  weighted for the number of control individuals  $I_0$ ;  $\mathbf{X}_{\mathbf{s}_{c_0}}$ ; the multiset matrix resulting from the auto-scaling performed with the weighted mean and standard deviation across the control individuals. From  $\mathbf{X}_{\mathbf{s}_{c_0}}$  we can extrapolate the pre-processed matrix of the control and responder sets, expressed in Eq. S6 and Eq. S7, respectively:

$$\mathbf{X}_{sc_0} = \begin{bmatrix} \mathbf{X}_{sc_{I_0}} \\ \vdots \\ \mathbf{X}_{sc_{I_0}} \end{bmatrix} \quad \text{Eq. S6}$$

$$\mathbf{X}_{sc_1} = \begin{bmatrix} \mathbf{X}_{sc_{I_1}} \\ \vdots \\ \mathbf{X}_{sc_{I_1}} \end{bmatrix} \quad \text{Eq. S7}$$

Among the different options, centering and scaling based on the control class (Eqs. S4b, S5b), will enhance the deviation of the responder individuals from the cell variability observed in the control individuals. Alternatively, individual centering and scaling (Eqs. S2b, S3b) might be a preferable option when there are considerable shifts in the observed surface marker expressions of cells between different samples. These operations now rid the resulting data of uninformative offsets and allow each surface marker to *a priori* contribute equally to the model fitted subsequently, taking into account both the multiset and control/responder structure of the data. However it should be noted that the last option has a disadvantage. When all cells of one response individual show up or downregulation of one or multiple markers compared to the cells of the control individuals, this information will be lost due to individual centering.



**Figure S2.**

The effect of Kernel bandwidth ( $h$  in Equation 4, main article) on the estimate of a (normal) distribution. In yellow: a histogram of 1000 samples from a Unimodal Gaussian distribution with mean 4 and standard deviation 1. In green: an under smoothed estimate using a small bandwidth. In red: an over smoothed estimate using a large bandwidth. In light-blue an estimate using a near optimal bandwidth.

Supplementary Material II

Additional results of ECLIPSE on the LPS dataset

As an example of the heterogeneity between responders, the ECLIPSE results obtained from Individuals #14 and #16 are shown in Figure S3. Two groups of neutrophils, mainly separated along the first component, can be observed in both individuals. However, in Individual #14 the two populations are not as well distinguished from each other as in Individual #16. Cells with a continuum in surface marker expressions connects the two populations at the upper part. In addition, Individual #14 presents a quite heterogeneous distribution within the mature neutrophils that can be encircled by two different gates. The percentages of the cells included in the gates for Individual #14 and #16 are reported in Table S1. Individual #16 displays two already well distinguishable homogenous populations. This indicates a difference in response between the two individuals, which for instance could suggest that the immune system of Individual #16 responded faster to the LPS stimulus when compared to the other individual.

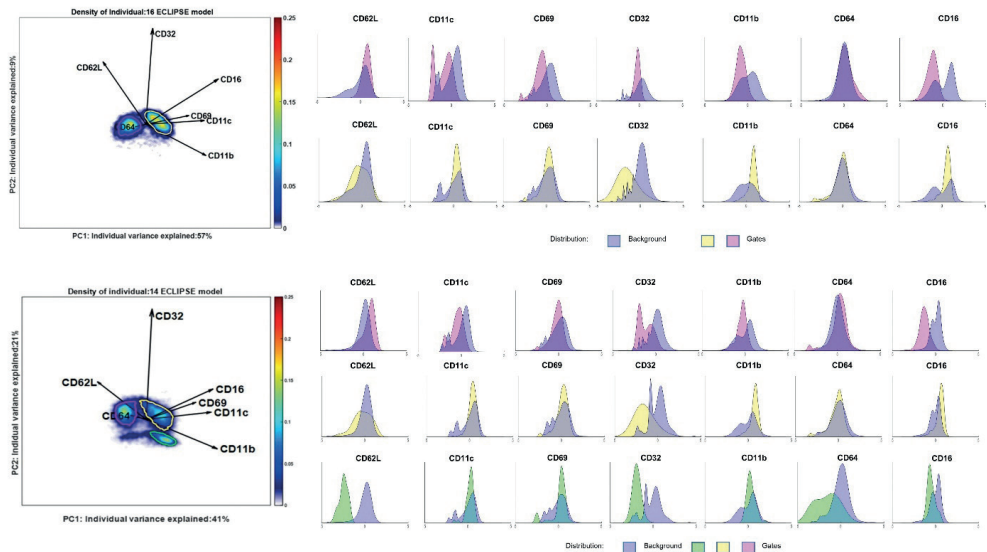


Figure S3.

Visualization of two responder individuals in the ECLIPSE model, built after eliminating the normal cells. The variance explained by the general model for each individual is displayed on the axis. The individuals present different distribution of the responding cells that can be selected by different gates, based on the higher density estimation. The histograms show the marker expression of the cells within the different gate; the distribution are displayed with the colour of the corresponding gate. Percentage of the cells, normalized on the original total amount of cells, is visualized in Table 1.

Table S1.

	Gate magenta	Gate yellow	Gate green
ID #14	13%	18%	7%
ID #16	16%	27%	0%

Percentages of the total cells of responder individuals #14 and #16 in the gates displayed in Figure. S3.

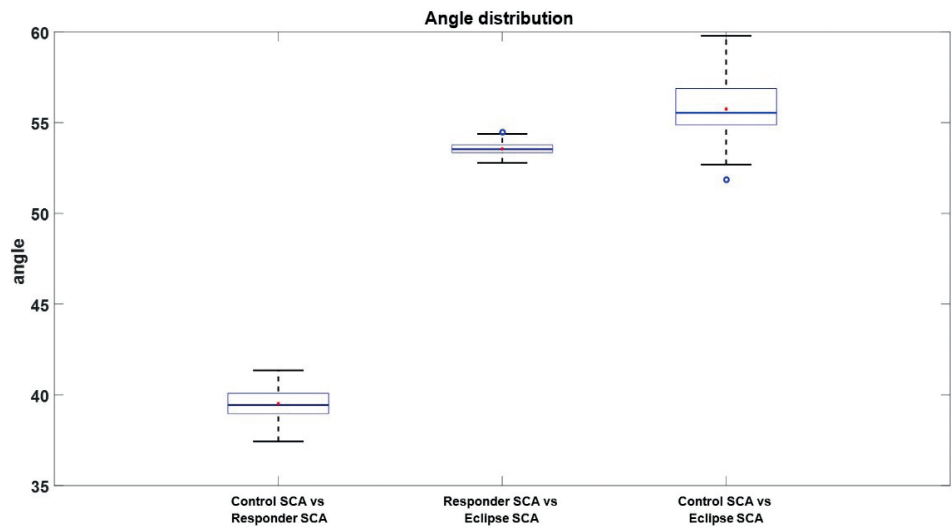
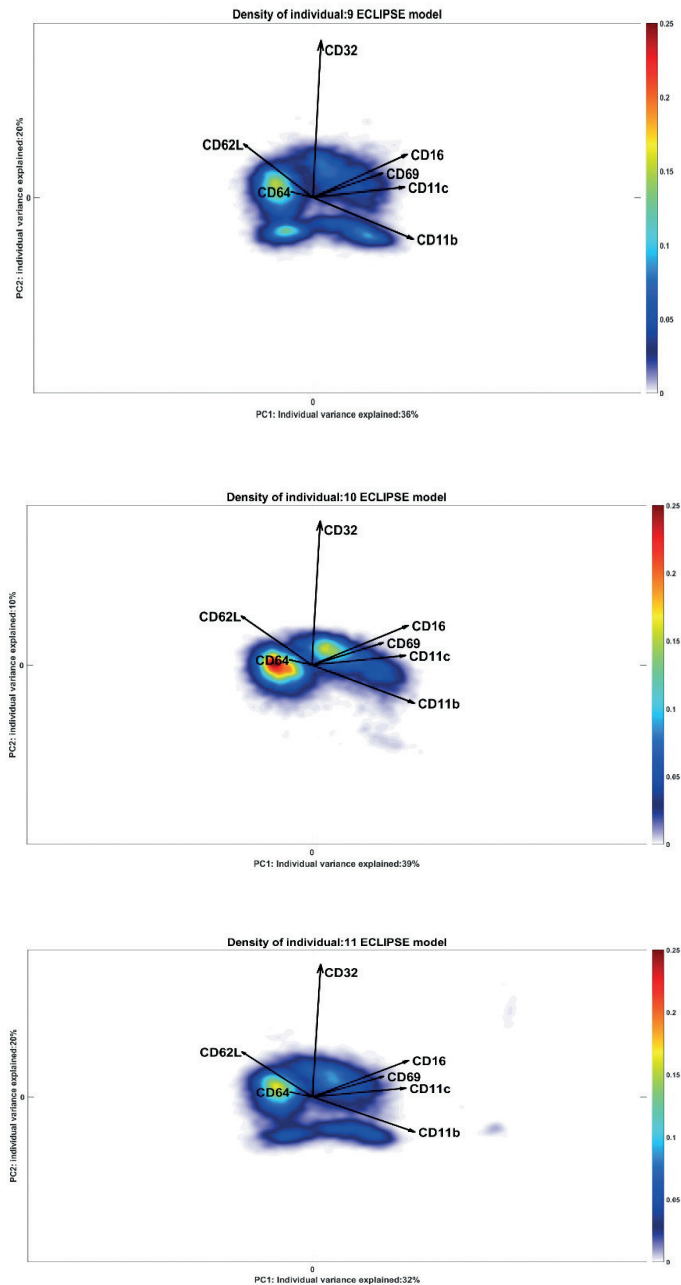


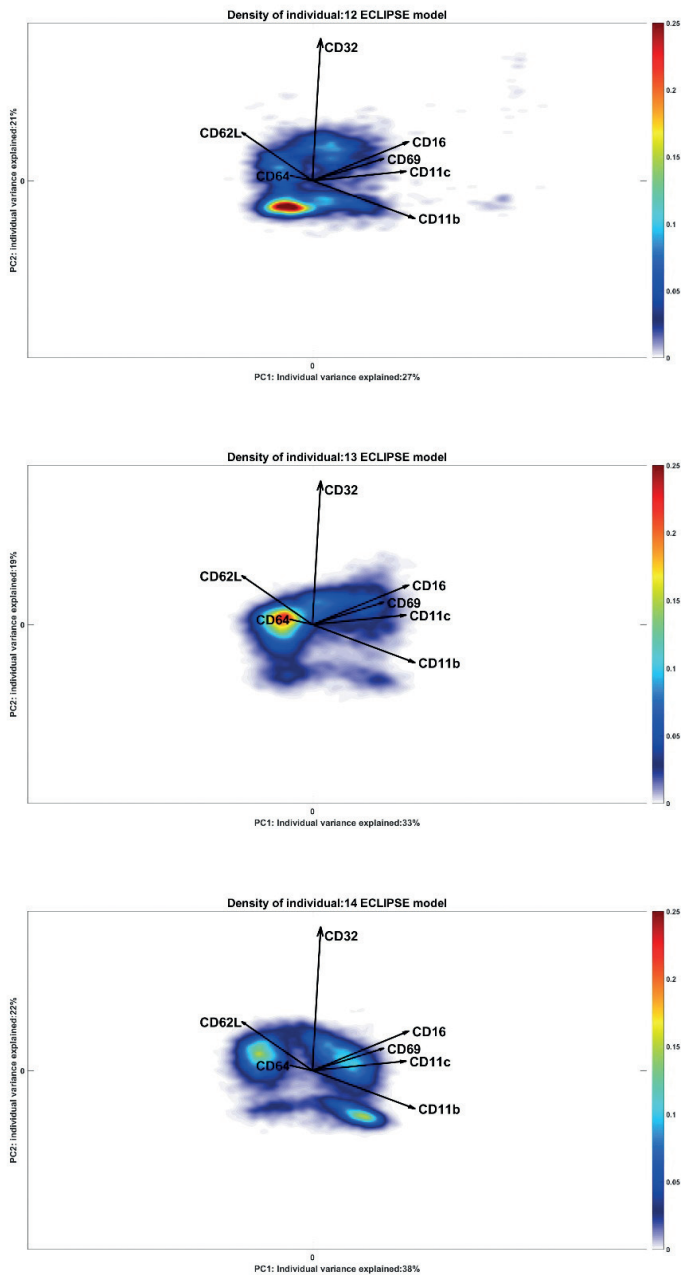
Figure S4.

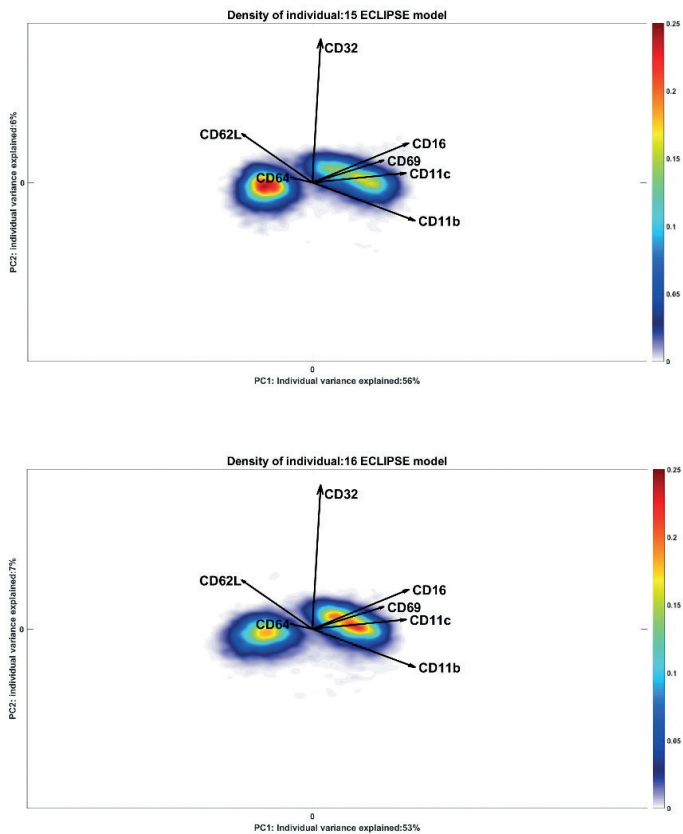
Box plot of the angle distributions between the sub-spaces defined by loadings of different SCA models. The first comparison shows that the Control SCA and Responder SCA models are the most similar as shown by the smaller angle than those obtained by the comparison of each of the models to the ECLISPE SCA model.

**Table S2.**

ECLIPSE plots of the LPS response group per individual.





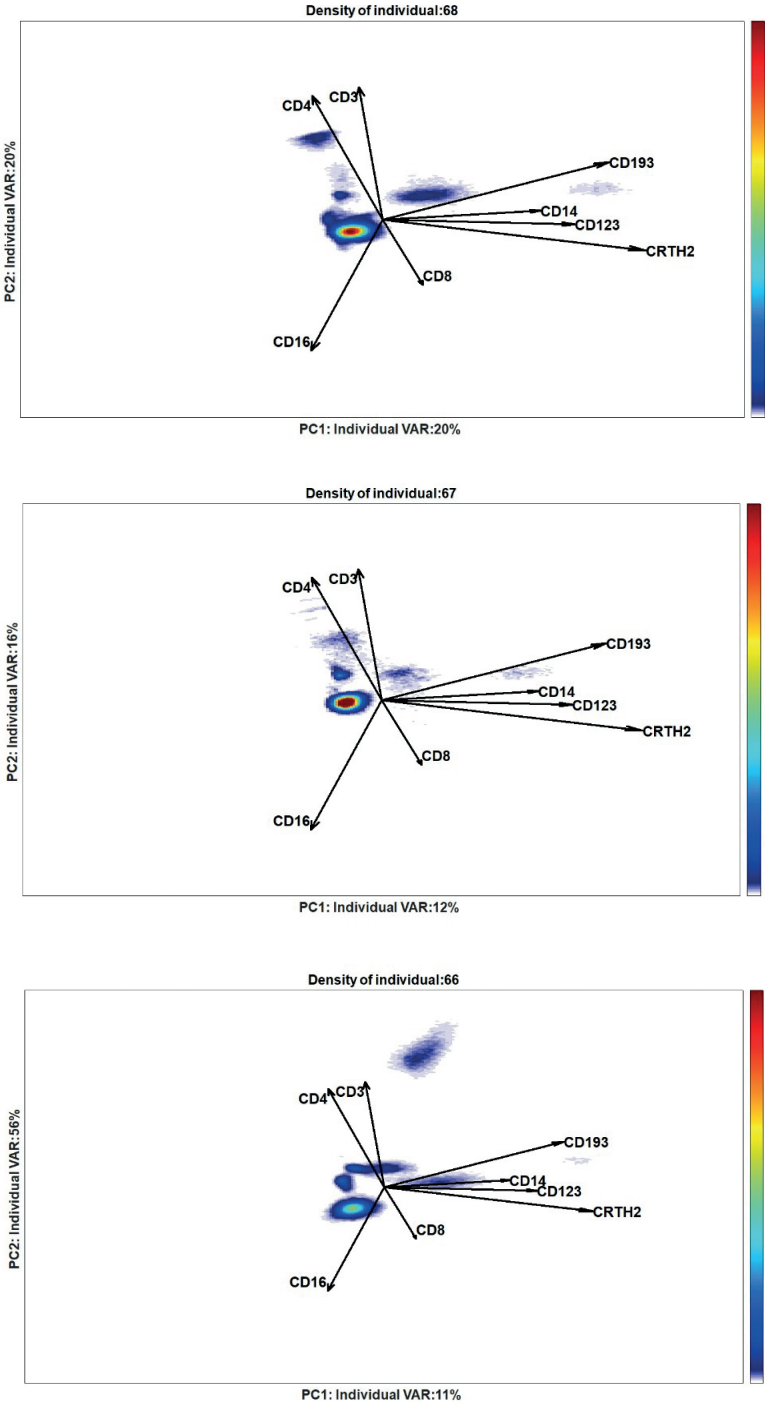


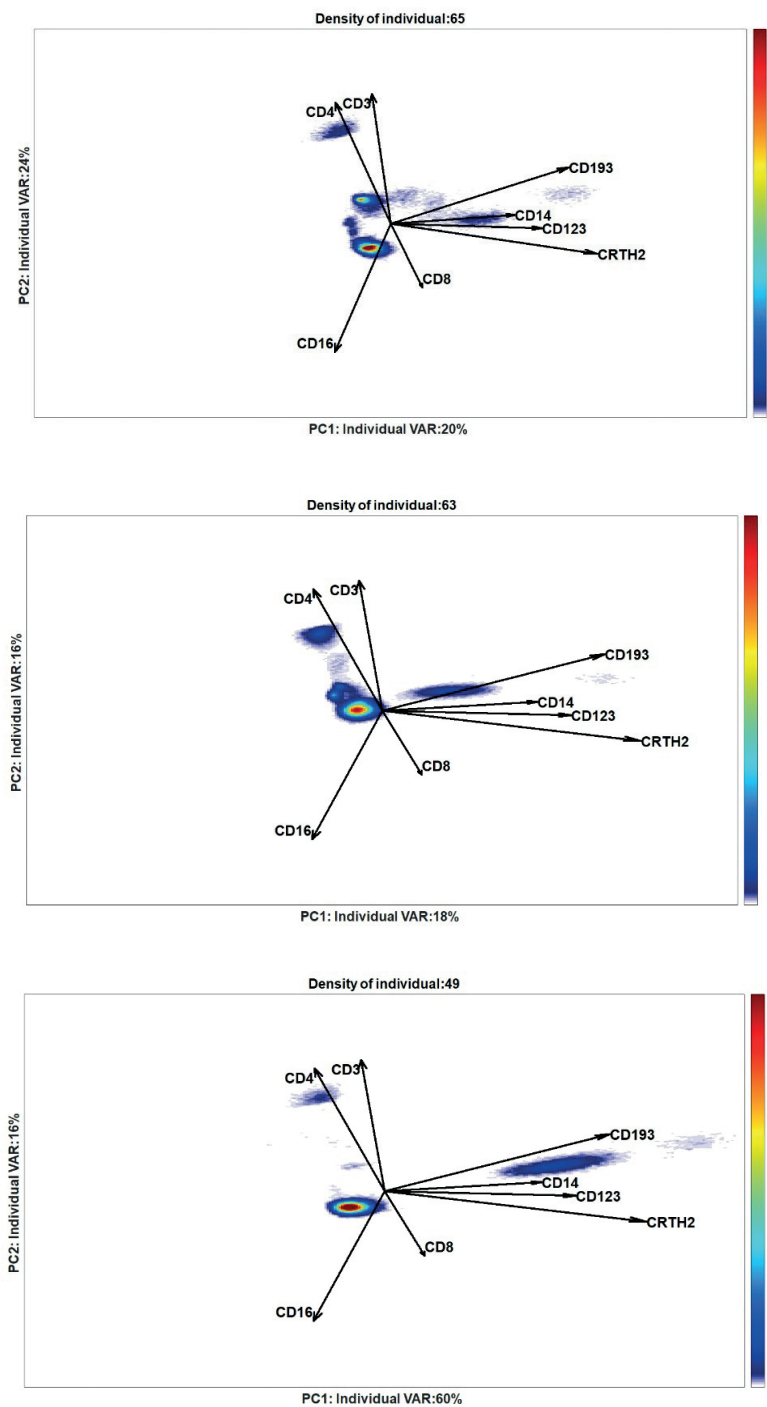
**Additional results of ECLIPSE on the asthma dataset**

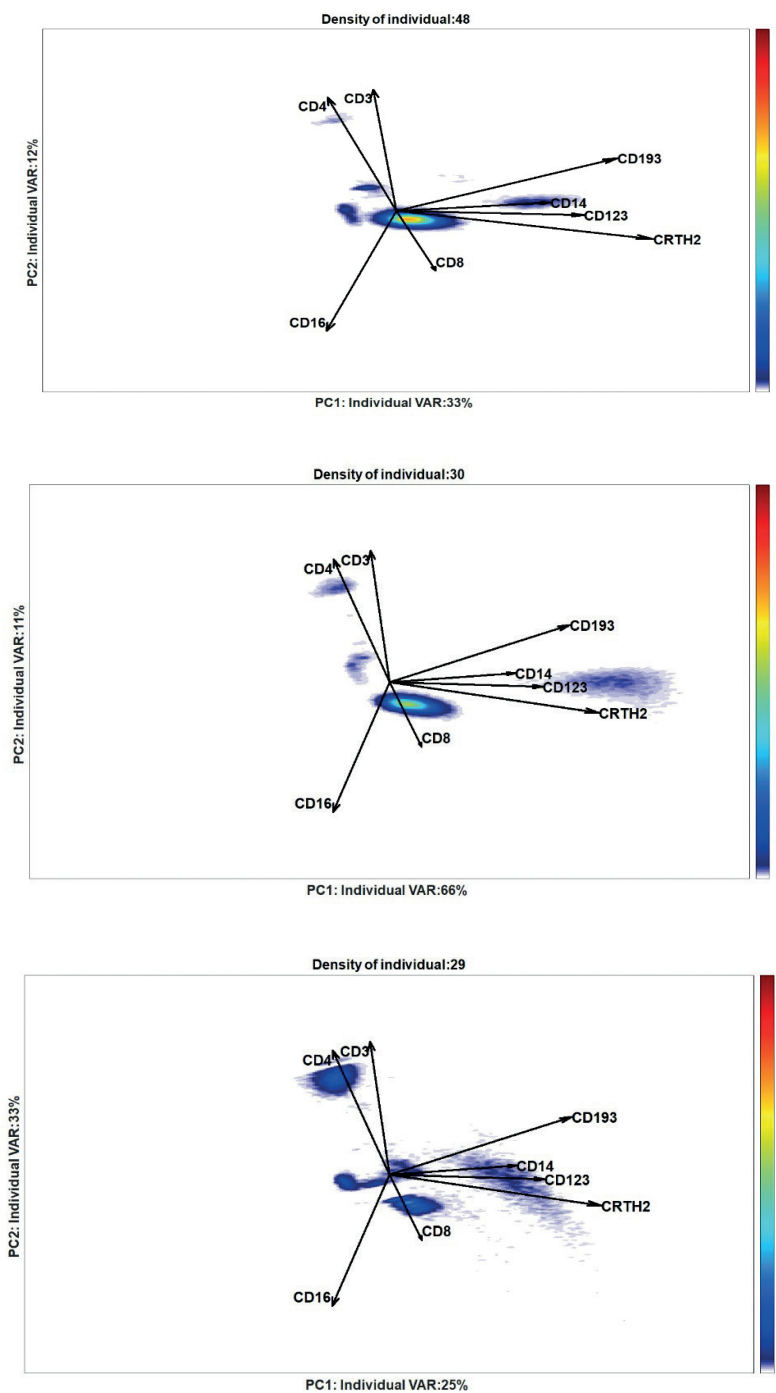
**Table S3.**

Density plots of all the asthma patients in the ECLIPSE model Space









## Supplementary Material III

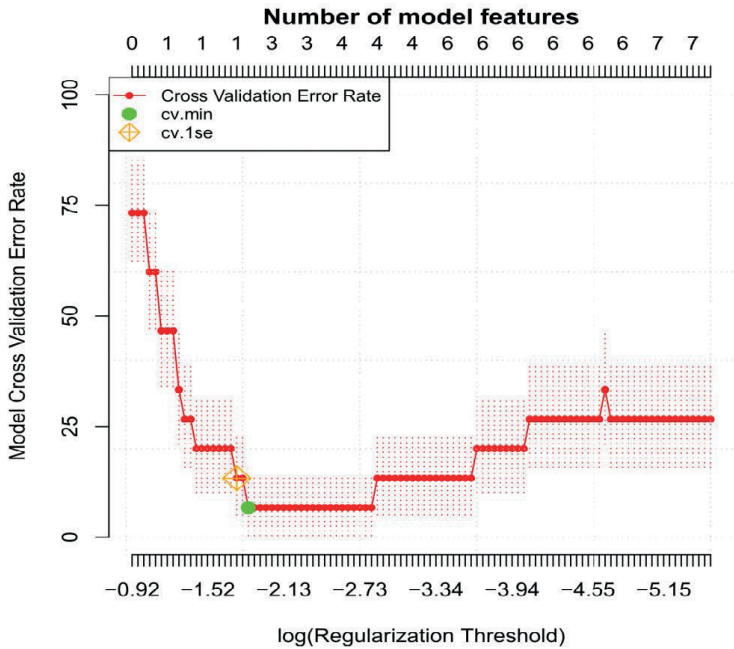
### Citrus and viSNE analyses on the LPS and asthma study dataset

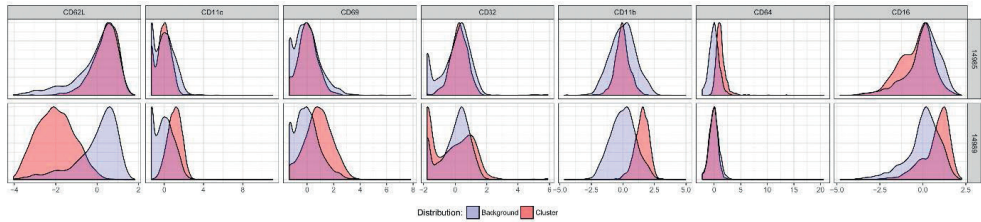
#### Results of Citrus on LPS data

Citrus<sup>6</sup> uses hierarchical clustering to identify phenotypically similar cell populations. Descriptive cell cluster-specific features are calculated on per-individual basis. These features can include, for example, the proportion of cells in the identified cluster for each individual. For intergroup analysis, the method employs a regularized classification model to detect the group-specific cell clusters for each sample.

By definition, a regularized model selects subsets of features of the data to achieve the best prediction, avoiding overfitting. The constant *regularization* threshold regulates the number of features used for the classification. A series of models with increasing complexity, i.e. with increased number of selected features, is built by varying this threshold. The fit of each model is estimated through cross-validation. A plot of the Model cross-validation Error Rate versus Regularization Threshold enables investigators to assess the quality of the results of the classification and the fit of the model chosen by Citrus analysis. An optimal model has low cross-validation error rate, which corresponds to a low percentage of misclassified samples. When the error remains constant within a range of increasing number of features, the regularized threshold associated to the fewest number of features is chosen in order to select the most informative features that differ between the two groups.

Citrus was applied to the LPS dataset using the R GUI. The regression classification model was trained on the data using the default pre-processing of the GUI, which implies arcsinh transform with cofactor 5. The accuracy of the classification models constructed is shown in Figure S5.



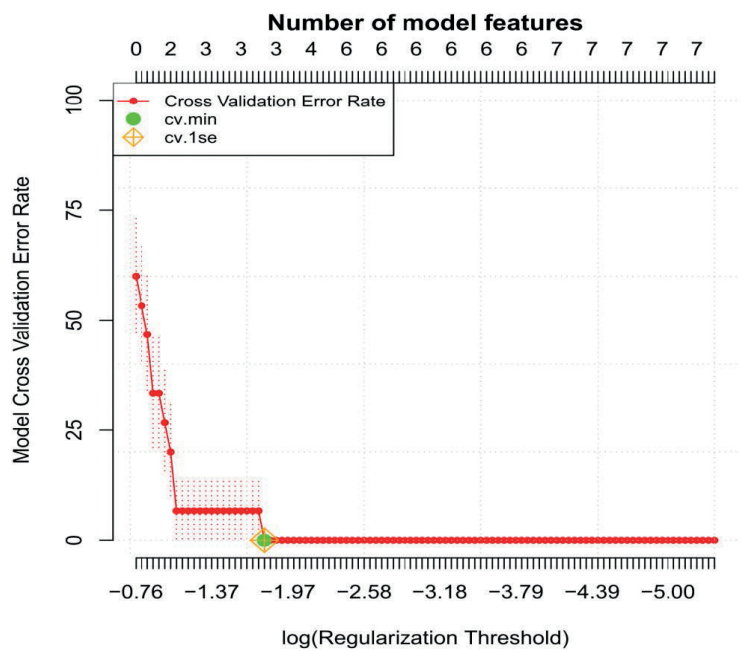


**Figure S6.**

The histograms show the phenotype of the cells belonging to the cluster (red) selected by the cross-validated model. The background histograms (blue) show the rest of the data, not included in the cluster. Cells from cluster 14985 are characterized by the distinctive phenotype CD16-and+CD62L+CD64+; cells from cluster 14989 are CD16+CD11b+CD69+CD11c+CD62L-. Both clusters are more abundant in the responder group.

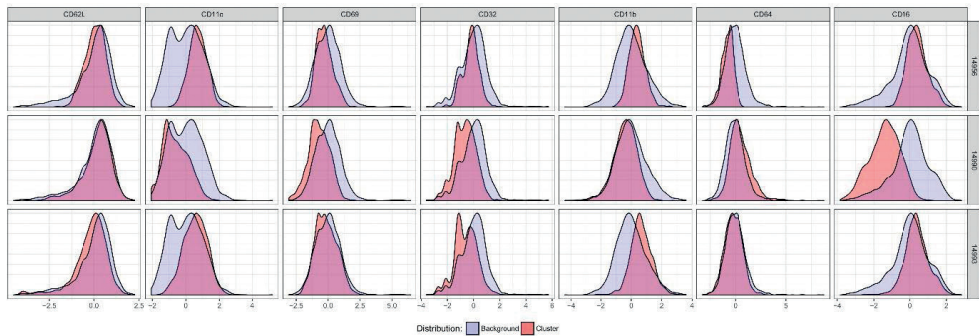
Cluster14989hasadistinctivephenotype,existingofCD16+CD11b+CD69+CD11c+CD62L– cells. Two CD32 populations are present in this cluster: CD32– and CD32+. In contrast, cells from cluster 14985 express high levels of CD64 and CD62L, while they have a dim expression of CD16. The cells present in the two cluster might be identified as activated mature and partly immature (CD16<sup>dim</sup>) neutrophils, respectively. Using the ECLIPSE method we also identified two LPS-specific cell clusters in the ECLIPSE space generated when plotting PC1 and PC2, Figure 4. To be able to compare these results, we applied Citrus on the pre-processed data as used in the ECLIPSE analysis. Subsequently, we applied Citrus on the LPS response-specific cells after ECLIPSE cell elimination.

First, Citrus was applied on the data pre-processed with the ECLIPSE algorithm, as described in *Pre-processing section*. In this analysis, all the cells were included. The cross-validated error rates of the obtained models are displayed in Figure S7.



**Figure S7.** Cross-validation error rate plot for LPS dataset transformed with the ECLIPSE pre-processing procedure: arcsinh transformation, individual mean center and control scaling. Model Cross Validation Error Rate equals to 0 is obtained with 3 stratifying features.

A perfect classification is achieved with a model which uses three distinctive features. This model is chosen as optimal by the Citrus Analysis. The most discriminant three clusters identified by the model are shown in Figure S8. Cluster 14956 and 14993 represent phenotypes that might be assigned to mature neutrophils, with CD16+CD62L+ and CD16+CD62L+CD32–, respectively. These clusters are found more abundant in the control group, compared to the responder group for which cluster 14990 is more present. This last cluster, with CD16–CD69– and CD62L+ can be assigned as immature neutrophils.



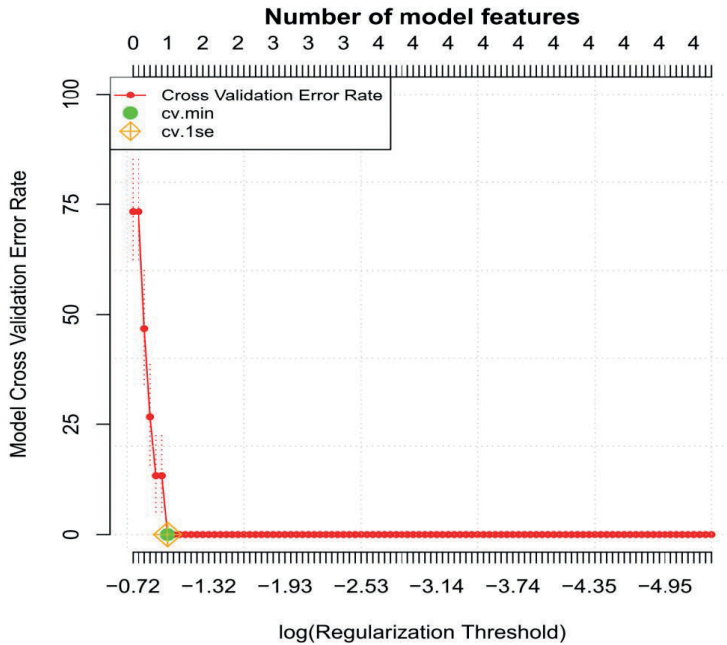
**Figure S8.**

Histograms show the phenotypes of the three cluster identified by the cross-validated classification model. Cluster 14956 and 14993 present phenotypes that might be assigned to mature neutrophils, with CD16+CD62L+and–. The cluster 14990, more abundant in the responder group, is identified as premature neutrophils with CD16–CD69– and CD62L+.

Finally, we applied Citrus to the LPS dataset, after the elimination of normal cells, provided by ECLIPSE. The model chosen by the Citrus analysis offers a perfect classification of the individuals in the two groups, as shown in the cross-validation error rate plot (Figure S9). In this case, however, the discriminating model requires a lower number of informative features. In fact, only one feature is necessary to correctly classify the samples. This feature corresponds to the abundance of cluster 14994 in the responder group. The phenotype of this cluster (Figure S10) indicates that the most discriminant population is represented by cells CD16–CD11b–CD62L+CD69–CD11c–/dim, which might correspond to gate **a** in Figure 4 of ECLIPSE, defined as immature neutrophils. In contrast to the previous analyses, there are no clusters of mature neutrophils identified as most discriminant ones between the groups.

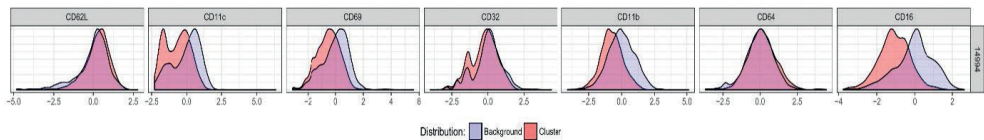
The results obtained by the three analysis show how individual mean-centering and control scaling has a positive effect in the predictive ability of the Citrus algorithm. The ECLIPSE algorithm, at the current stage, is not developed to be used as a classification model. It is more suited for explorative research, finding cell populations that arise upon an immune response. These results can give further insight in the mechanism behind the immune response studied. Due to the different purposes of the models, a direct comparison of the prediction accuracy of the two methods is not possible. However, the last analysis, performed on the data after removal of normal cells, brings a further improvement of the classification model in terms of complexity. This shows the power of the subsampling provided by ECLIPSE, since in principle the ECLIPSE subsampling may be beneficial for discriminant classification methods as Citrus.





**Figure S9.**

Cross-validation error rate vs Regularization threshold plot for LPS dataset transformed with the ECLIPSE pre-processing procedure (arcsinh transformation, individual mean center and control scaling), after elimination of normal cells. Model Cross Validation Error Rate null is obtained with 1 stratifying feature.



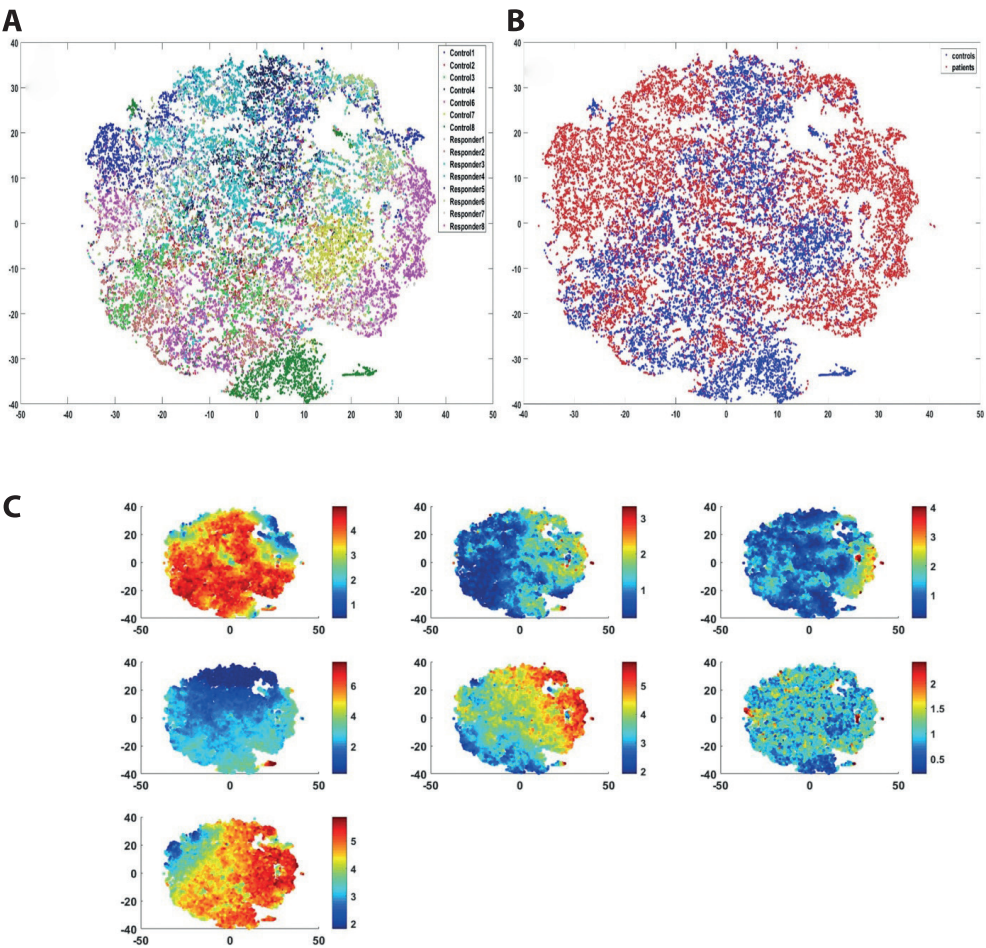
**Figure S10.**

Histograms of the phenotypes of the most discriminant cluster, more abundant in the responder group. Cells in the cluster are CD16–CD11b–CD62L+CD69–CD11c–/dim.

### Results of viSNE on LPS data

viSNE<sup>7</sup> analysis was applied to the LPS dataset using the Matlab GUI cyt (downloadable from the website <https://www.c2b2.columbia.edu/danapeerlab/html/cyt.html>). In order to obtain a better visualization of single-cell resolution, we run all the analyses on a subset of the LPS dataset. Each individual was subsampled to 2000 cells, so that the total amount of cells was 30,000. The analysis was first performed using the default transformation present in the GUI, consisting of arcsinh transformation with cofactor 5. The results are shown in Figure S11. The upper left panel (Figure S11a) show the cells in the viSNE map

coloured per individual. It seems that cells of the same individual are grouped together, suggesting that the cluster found by the algorithm are individual specific. The viSNE map in the upper right panel (Figure S11b) shows cells coloured per control/responder group: a considerable overlap of cells between control and responder individuals is observed. Distinctive region for the responders seems to be the upper left region and the right area. However, not all the response individuals show cells in those regions.

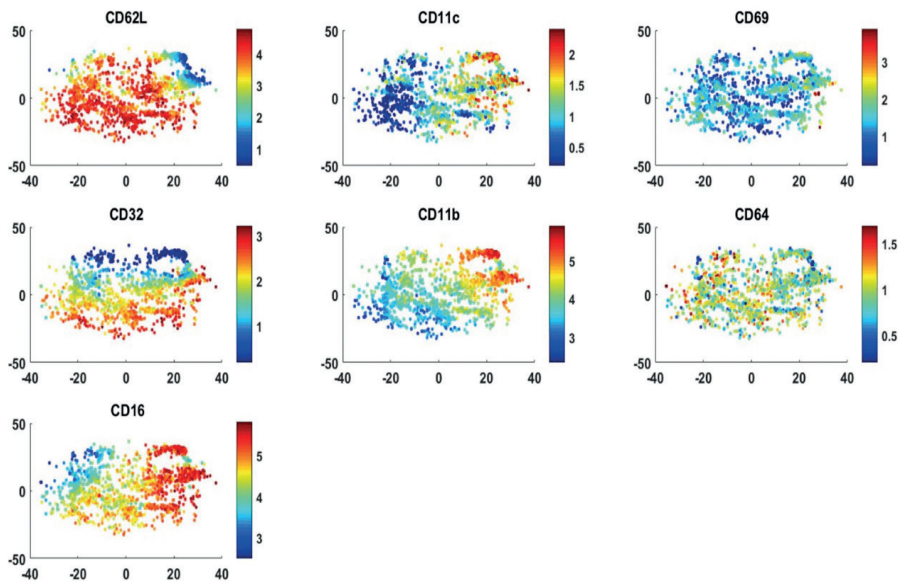


**Figure S11:**

viSNE analysis performed on the LPS data pre-processed with the default arcsinh transformation present in the cyt gui. A) cells in the viSNE map are coloured per different individual; B) cells from the responder individuals, in the viSNE map, are coloured based on control (blue) and responder (red) group. C) cells are coloured based on expression of the single 7 markers.

Figure S12 shows the constructed viSNE map only for the responder individual ID14, for whom we show the ECLIPSE model in the Supplementary material II (Figure S3, upper panel). LPS-specific cells CD62L<sup>+</sup>CD16<sup>+</sup>CD11b<sup>+</sup> (upper right) and CD62L<sup>+</sup>andCD16<sup>+</sup> (upper left) can be observed. These cell populations were clearly distinct also in the ECLIPSE space.

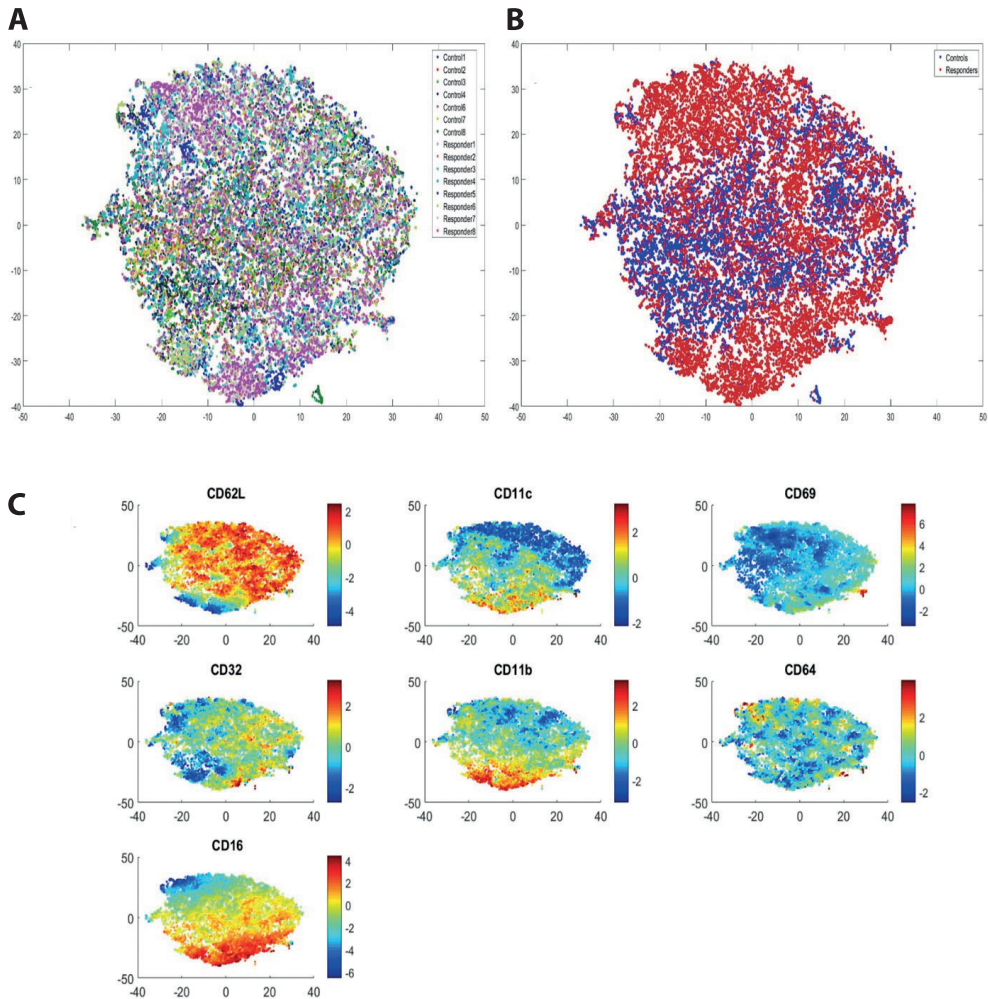
A big overlap is present with the control region. However, in viSNE, we cannot affirm whether these normal-like cells are overrepresented compared to the controls and thus interesting to describe the immune response. The Difference between Densities (DbD, step 4), performed by ECLIPSE, takes into account both situation that might occur because of immune response: deviation from normal cell marker variability and overproduction of normal immune cells.



**Figure S12.**

Cells from the responder individual with ID14 shown on the viSNE map built on all the control and responder individuals. The panels show the marker expression of the measured marker.

Secondly, we performed viSNE on the LPS data pre-processed by the ECLIPSE algorithm; the results can be observed in Figure S13. The viSNE map, coloured per individual, shows how the cells are distributed across the map and no donor-specific clusters are present. In fact, the multiset pre-processing specifically adopted for the LPS study by ECLIPSE (as described in the *Pre-processing* section of Supplementary Material I, Equations S2b and S5b), helps to remove the individual variability of the control individuals. In this case, the individual control variability is biologically not that relevant and introduces noise into the results. It might originate from instrumental variability, for instance.

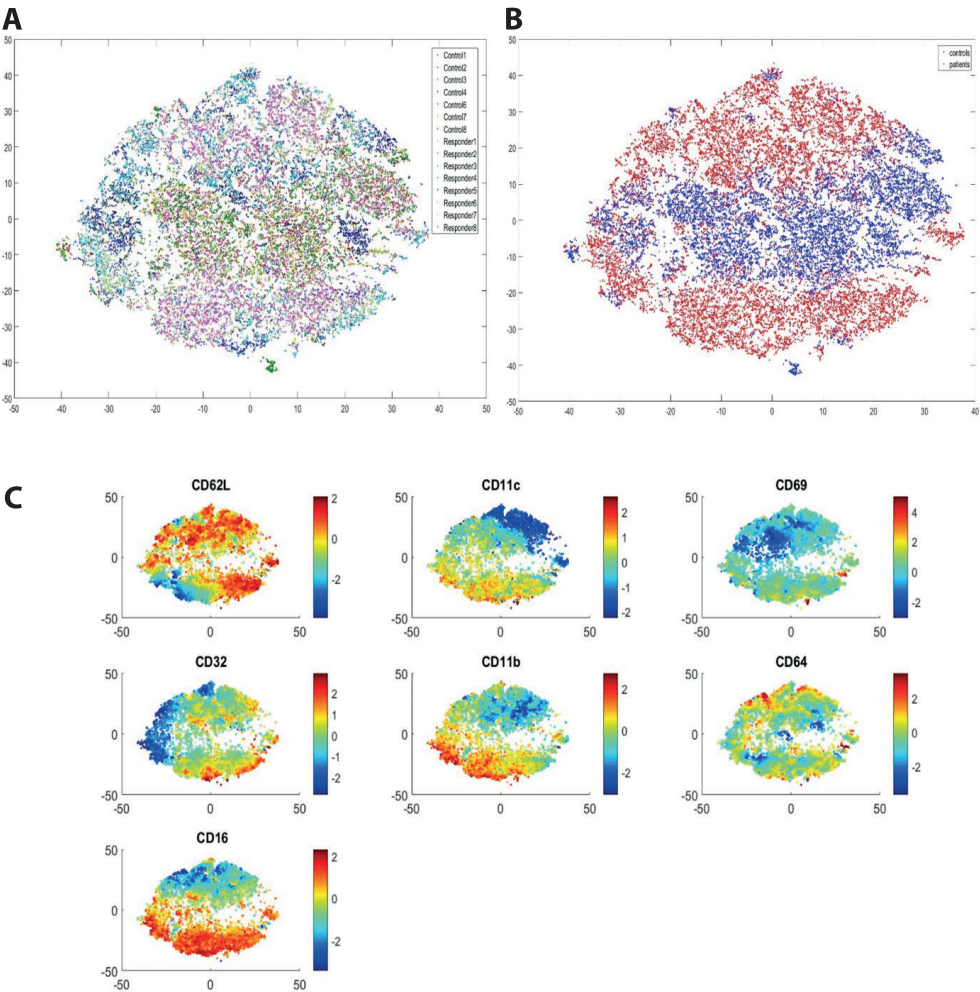


**Figure S13.**

viSNE analysis performed on the LPS data mean-centered and scaled according to the ECLIPSE algorithm. A) cells in the viSNE map are coloured per different individual; B) cells in the viSNE map are coloured based on control (blue) and responder (red) group. C) cells from the responder individuals, in the viSNE map, are coloured based on expression of the single 7 markers.

Finally we applied viSNE to the LPS data pre-processed by ECLIPSE, but after removal of normal cells from the responder individuals, performed by the ECLIPSE algorithm. The results are shown in Figure S14. It is clear how the cells from control individuals (blue, Figure S14b) are placed in the middle of the viSNE map, whereas the cells from the responder individuals (red, Figure S14b) are distributed in the upper and lower region of the map. Very little overlap is present between cells of the two groups and clearly, a lower amount of responding cells is present. The marker expression of these remaining cells is shown

in Figure S14c. ECLIPSE removes mostly the CD16dimCD62L+ and CD16–CD62L– cells, leaving two quite distinctive populations. No particular difference in marker expression is noted in the viSNE analysis, performed on the ECLIPSE subsampled data.



**Figure S14.**

viSNE analysis performed on the LPS data mean-centered and scaled, after removal of normal cells in the responder done by ECLIPSE. A) cells in the viSNE map are coloured per different individual; B) cells in the viSNE map are coloured based on control (blue) and responder (red) group. C) cells from the responder individuals, in the viSNE map, are coloured based on expression of the single 7 markers.



The viSNE algorithm is used mostly for data dimensionality reduction and visualization purposes, while the main goal of ECLIPSE is the selection of interesting cellular information and/or populations based on multiple marker co-expression. In principle, no quantitative parameter can be used for the comparison with the ECLIPSE results on the LPS data. However, we have shown how the viSNE analysis is susceptible to sample-to-sample variation, which has influence on the resulting map. The multiset pre-processing operated by ECLIPSE systematically reduces this variation.

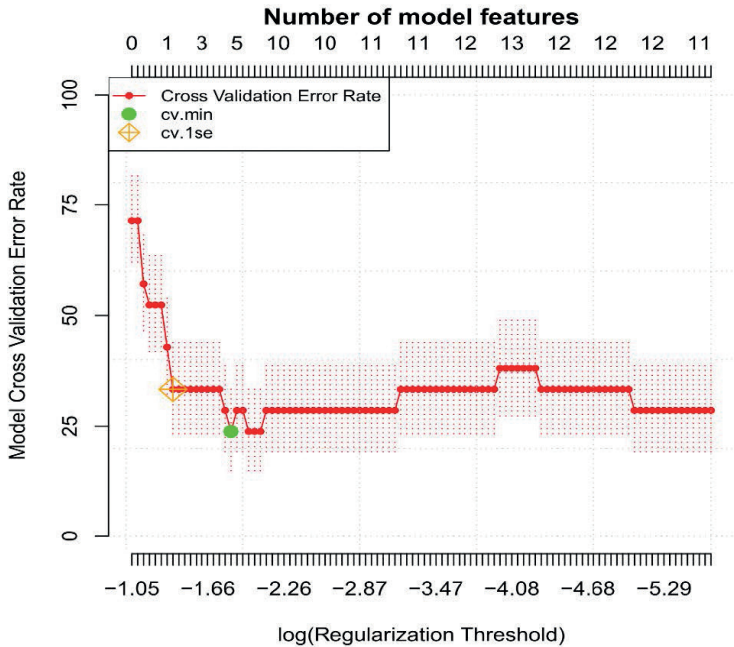
Both ECLIPSE and viSNE performed on the ECLIPSE subsampled data reveal two distinctive populations, which are mostly differentiated by CD16 and CD62L marker expressions. However, an advantage of ECLIPSE is that the co-expression among the markers are displayed in one single biplot. Secondly, ECLIPSE defines response-specific cell populations, also taking into account normal cells that are overrepresented upon an immune response. These cells are not specifically defined when analyzing data with the viSNE algorithm. Next to this, the hallmark of ECLIPSE is the removal of normal cells, leading to a less crowded representation, which enables to distinguish the LPS responding cell populations better, also in the viSNE map (Figure S14).

### **Results of Citrus on the asthma data**

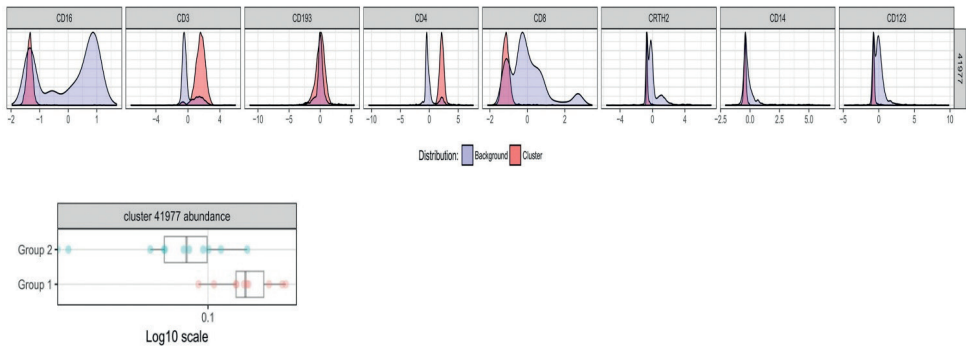
We trained Citrus on the asthma data using the R GUI. The model error rates plot of the built models is shown in Figure S15.

The highest accuracy achievable by the analysis correspond to 25% of misclassified samples. Four features were necessary for this minimal error (Figure S15, cv.min). A model with a cross-validation error 1 std higher than the minimum, corresponding to 30% of misclassification, requires 1 feature (Figure S15, cv.1se). The clusters detected by both models are shown below.

The feature identified by the model cv.1se corresponds to the abundance of cluster 41977, found to be more abundant in the control group (group 1) compared to the asthmatic individual (group 2) (Figure S16). The phenotype indicated that the cluster is mostly characterized by CD4+ T cells, having CD3+CD4+ and CD8–CD16–CD123–CD14–CRTH2– expression levels.



**Figure S15.** The Figure shows the Model Cross-validation Error Rate vs the log(Regularization Threshold) for the classification models constructed on asthma data. The green circle (cv.min) points out the model with the smallest number of features necessary to obtain the lowest cross-validation error, which corresponds to 25% of misclassified samples; the orange diamond (cv\_1se) indicates the model with the smallest number of features associated to cross-validation error 1 std higher than the minimum error, which leads to an error of around 30%.



**Figure S16.** The upper panel shows the phenotype of the most discriminant cluster between control and asthmatic individuals, found by the model associated with cv.1se error (~30%). This cluster is on average more present in control group (group 1) than in asthmatic (group 2), as displayed on the lower panel. Cells in the specified cluster can be recognized as CD4+CD3+ and CD8- T cells.

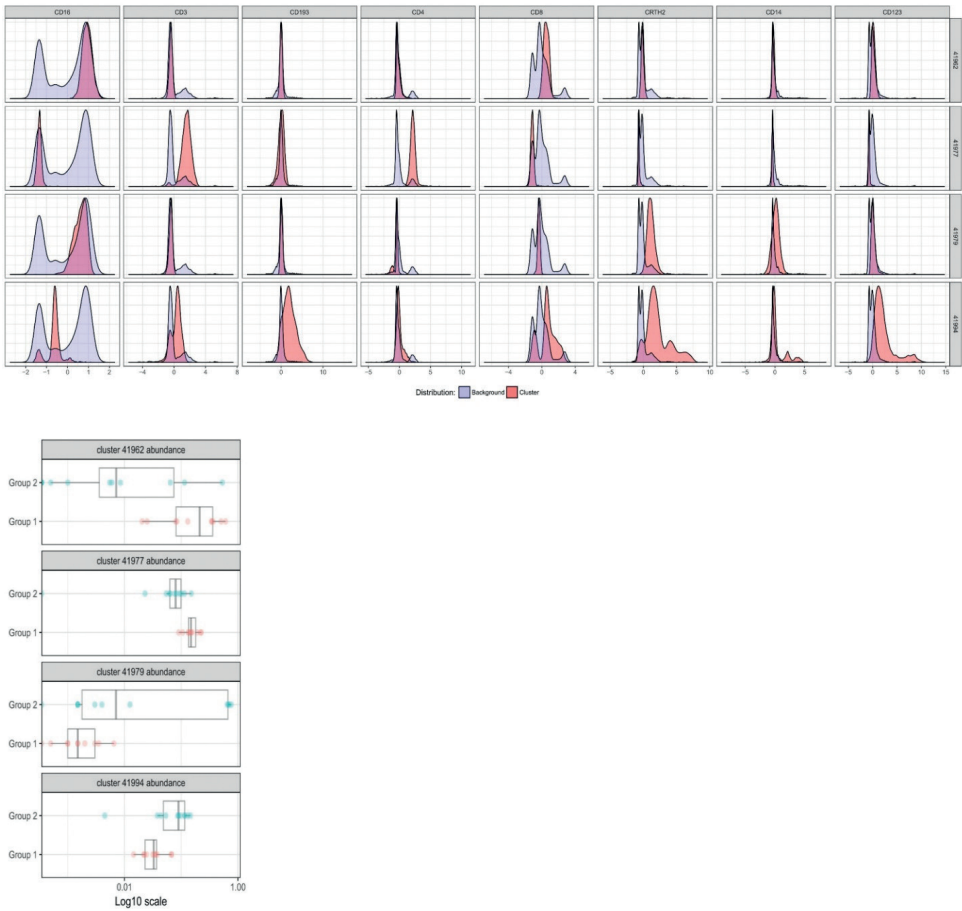
The cell clusters found by the best performing model are shown in Figure S17. Cluster 41962 is characterized by cells with CD16+CD123dimCD8dim and CD3–CD14– expression levels; while cluster 41977 has a CD4+ T cell phenotype (CD3+CD4+). These clusters are more abundant in the control group compared to the asthmatic. This is in contrast to the ECLIPSE results, in which we found CD3+CD4+ cells present in the asthmatic patients after the ECLIPSE elimination step of normal cells.

The last two clusters are more represented in the asthmatic group when compared to the control group. For these clusters, it is harder to identify the cell populations. Cluster 41979 consists of CD16+CD14dimCRTH2+CD123+ cells, which might be identified as monocyte derived cells. As shown by the cluster abundance graph on the right, there is high variability in the occurrence of this cellular cluster among the asthmatic patients. The last cluster 41994 has a very heterogeneous expression pattern, with multiple peaks per marker. | This indicates the presence of multiple cell types within the cluster. Based on the smaller peaks with high CD193, CRTH2 and CD123 expression, basophils and/or eosinophils might be identified within this cluster. In addition, CD3+CD8+ T cells might be present in this cluster. However, co-expression of these markers should be verified to be able to draw these conclusions. When multiple cell types are present within one Citrus cluster, it is impossible to state which cell types are represented, since the single marker histograms do not give information about co-expression of markers. ECLIPSE, on the other hand, allows interpretation of co-expression from the location of the cell populations and the respective orientation of the loadings.

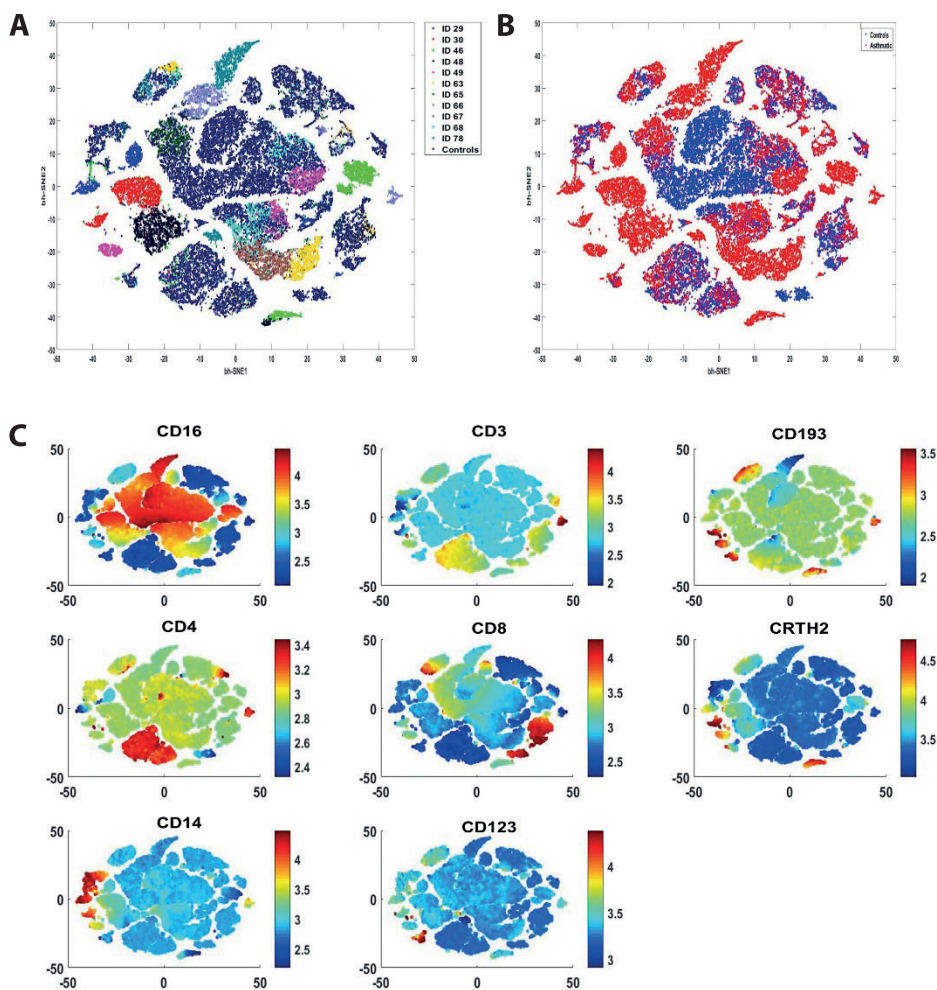
### **viSNE analysis of asthma data**

viSNE was performed on the asthma data. Due to limits in computational power, the original datasets consisting of almost 2 millions cells needed to be downsampled to 3000 cells per individual (48000 cells in total). Figure S18 shows the viSNE analysis performed on the arcsinh-transformed data (cofactor 5), coloured per individual (upper left panel) and for both groups (upper right). Substantial overlap of cells is present between the control and the asthmatic patients. Based on the single-marker expression profile of all cells (lower panel), we might conclude that most of the overlapping regions are associated to CD4+ T-cells; CD8+ T-cells; CD16+ cells, which might be identified as neutrophils; and CD14+ cells, which might be identified as monocytes. Also a few smaller CD193+ clusters were found, which could be eosinophils or basophils.





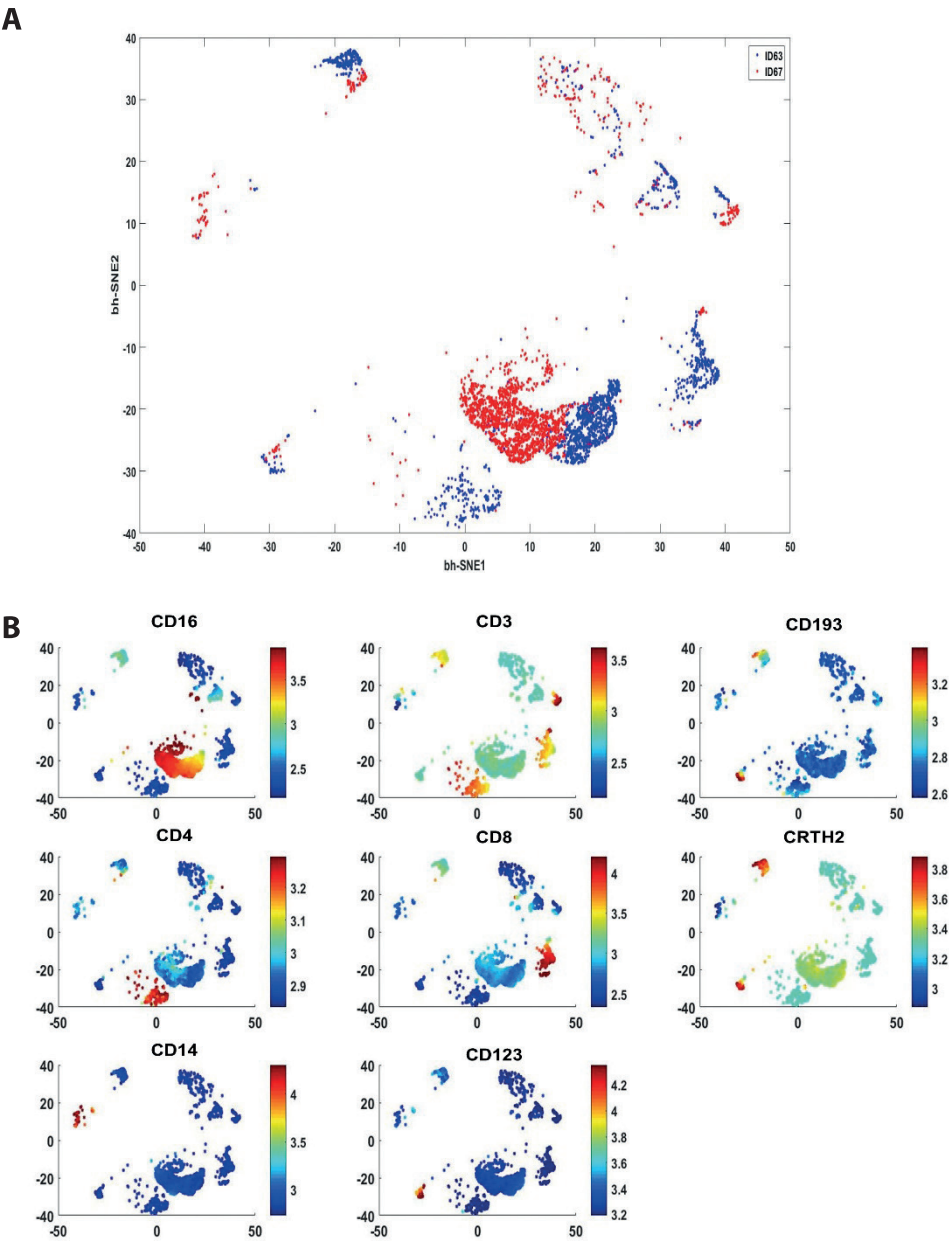
**Figure S17.** Upper panel: Histograms show the phenotype of the discriminant clusters, found by the model associated with smallest cross-validation rate error, corresponding to 25% misclassification. Lower panel: Differential abundance of these clusters between control (group 1) and asthma (group 2).



**Figure S18.**

viSNE analysis performed on the asthma data. A) cells in the viSNE map are coloured per different individual; B) cells, in the viSNE map, are coloured based on control (blue) and responder (red) group. C) cell are coloured based on the expression levels of the single 8 markers.

Considerable heterogeneity can be observed among the asthmatic patients. In this case, maps with distributions of a single or a few patients may enhance focus on subtle and patient-specific cell populations. Figure S19 shows the viSNE map with individuals #63 and #67, which were also grouped and analysed in a partial model by ECLIPSE (Figure 11).



**Figure S19.**

A) viSNE map obtained by the analysis of the asthma data, only the cell distribution of asthmatic individual #63 and #67 are shown. B) cells coloured based on the expression levels of the single 8 markers.

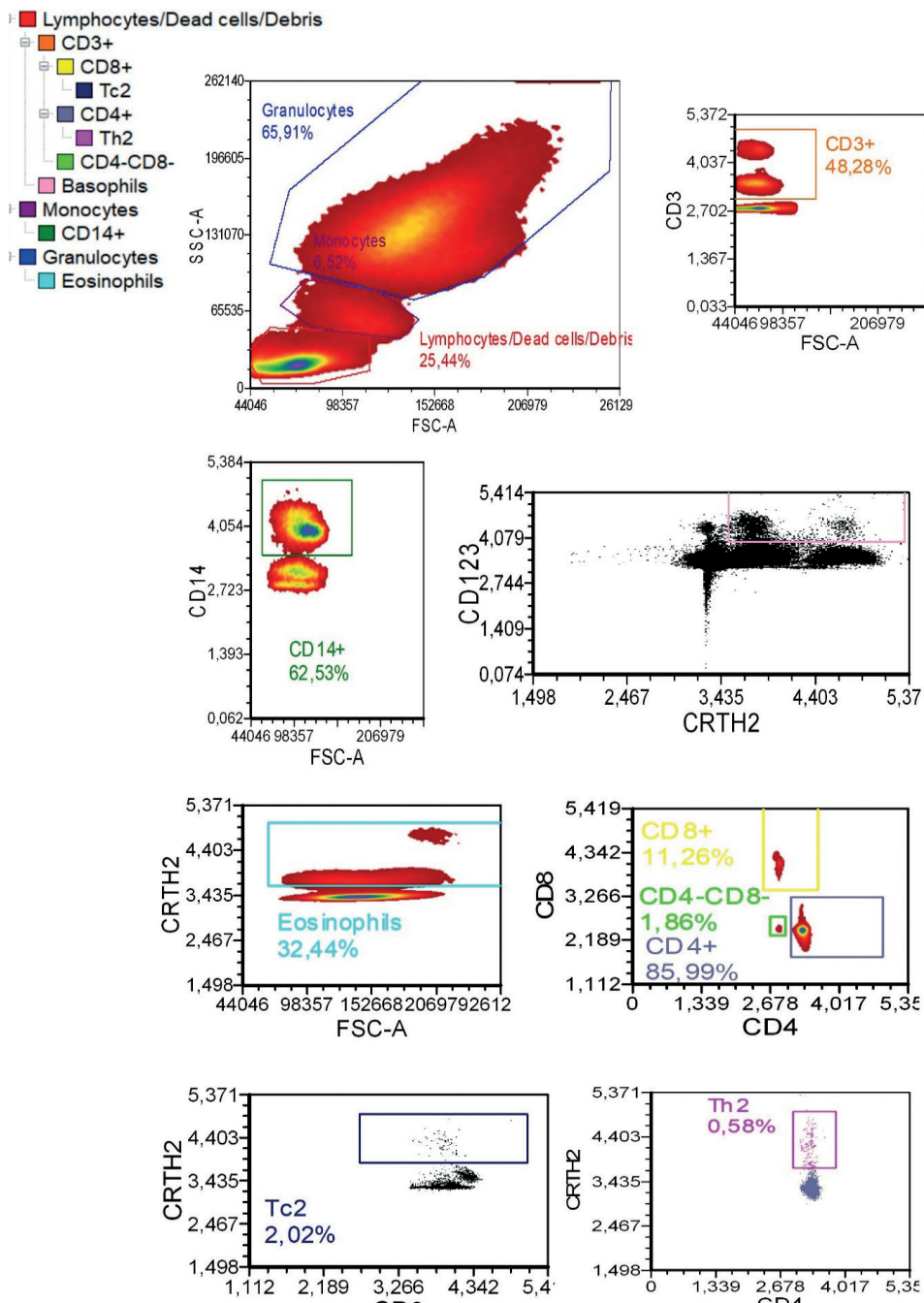
Similar cell subpopulations seem to be present for both individuals. However, for individual #67 only a few cells are present in various regions. This might be a consequence of the down-sampling, which retains the more abundant CD16+ cell population at the expenses of rarer cells. The phenotypical marker pattern observed in Figure S19b is similar to the phenotypes shown in the ECLIPSE partial model. CD16+ cells might be identified as neutrophils; CD123+CD193+CRTH2 cells as basophils; while CD3+CD8+ cells and CD3+CD4+ cells might be assigned to CD8+ T cells and CD4+ T cells, respectively. A key disadvantage for easy identification of the various cell populations is the representation of the marker expressions in single plots per marker.

## Supplementary Material IV

### Manual sequential gating of the asthma red cluster and projection into ECLIPSE space

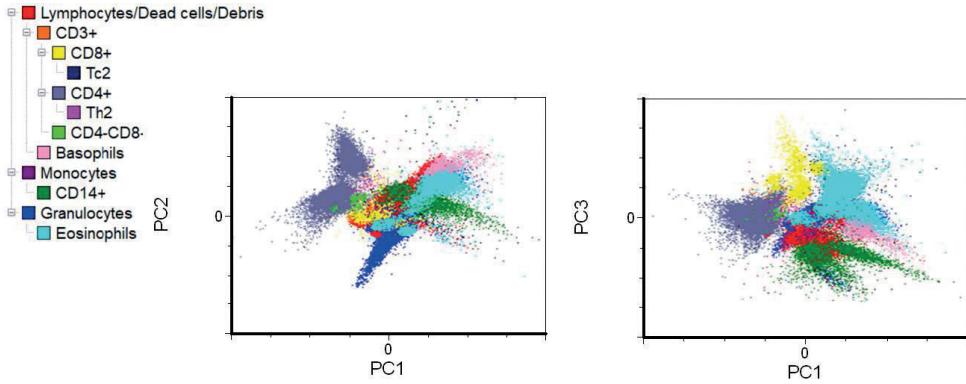
Manual gating of the cells of the individual belonging to the red cluster (Figure 10c) was performed in FCS express 5, according to the surface markers phenotype as shown in Figure S20. The gating demonstrates the absence of double positive CD4/CD8 T cells. Double negative CD4/CD8 T cells are present in a very low percentage. These cells, probably also present in the control individuals, were eliminated by the ECLIPSE algorithm. We can conclude that the mutual direction of CD4 and CD8 markers in the ECLIPSE space represents a big part of the variation of the data well, associated with the presence of different types of T cells (CD4+ and CD8+).

CD8+ T cells (CD3+CD8+CD4-, yellow gate) are not easily visualized in the first two PCs of the ECLIPSE model (Figure S21, left panel, PC1-PC2 plot). The CD8+ T cells are positioned in the middle of the plot, overlapping with other cell populations. Obviously, the other markers are more important to explain the variance of cells this 2PCs model (Figure 10C), since all the other markers show longer loadings when compared to the CD8 loading. If the CD8+ T-cells were more important, the loading would be longer and the cell population would be visible in these first 2 PCs. The investigation of the ECLIPSE space built on PC1 and PC3 made the CD8+ T cells better distinguishable (Figure S21, right panel, PC1-PC3 plot and Figure S22).



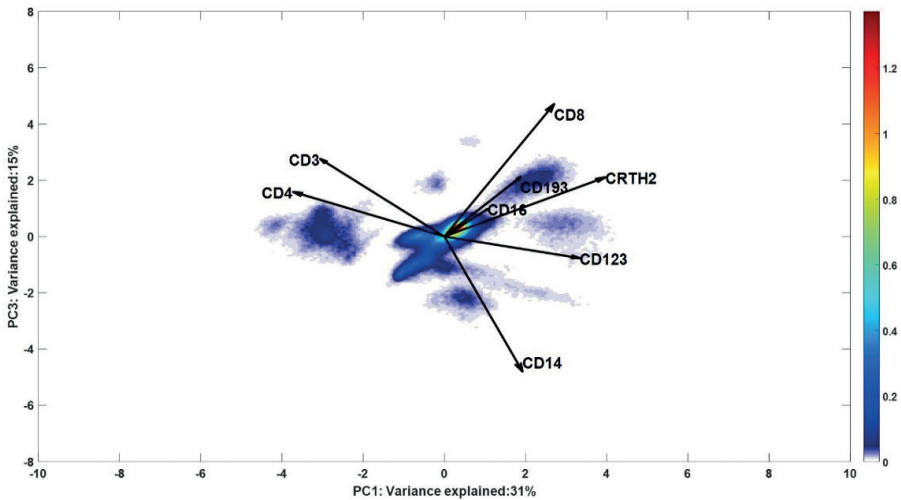
**Figure S20.**

Manual sequential gating of the cells from the individuals belonging to the red cluster, in Figure 10c. The gates are differentially coloured.



**Figure S21.**

ECLIPSE partial model of the red cluster: the left panel shows the space built with PC1-PC2, the left panel shows the space of PC1-PC3. The cell scores in the plots are coloured accordingly to the gates found by the manual gating procedure.

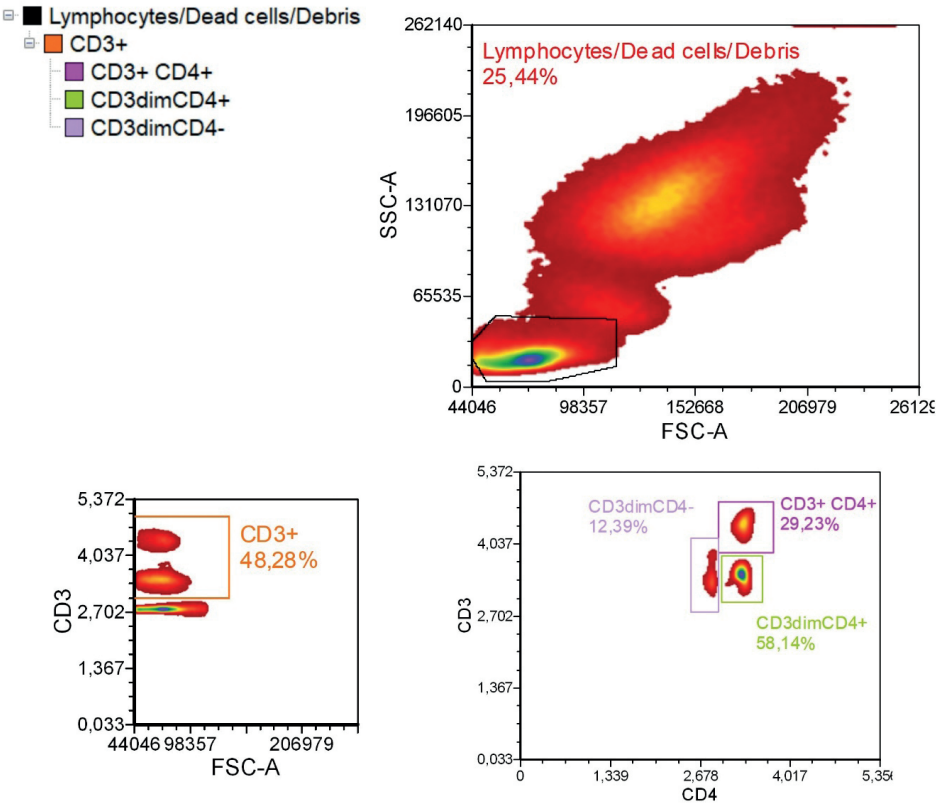


**Figure S22.**

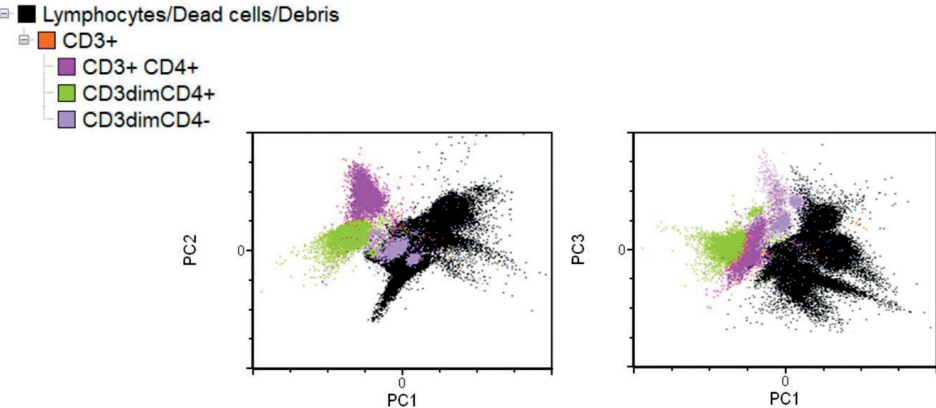
Density representation of the ECLIPSE partial model of the red cluster, showing PC1 and PC3.

CD4<sup>+</sup> T-cells were easily visualized in the left part of the ECLIPSE plot (Figure 10c, ECLIPSE paper), with two main distinctive populations. These populations are different because they express different levels of the marker CD3. Based on the orientation of the loadings in space spanned by PC1 and PC2, we would expect to find CD3<sup>bright</sup>CD4<sup>+</sup> cells situated closer to the CD3 loading; while CD3<sup>dim</sup>CD4<sup>+</sup> cells are expected to lie more to the left, in the direction of the CD4 loading and a bit further away from the CD3 loading. This was confirmed when backgating the cells, as shown in Figure S23 and S24.





**Figure S23.** Manual sequential gating of the cells from the individuals belonging to the red cluster, in Figure 9c. Gates for CD3+ cells are shown and coloured differently.



**Figure S24.** ECLIPSE partial model of the red cluster: the left panel shows the space built with PC1-PC2, the left panel shows the space of PC1-PC3. The cell scores in the plots are coloured according to the expression of the marker CD4 and CD3, in the CD3+ gate.

## Supplementary Material V

### ECLIPSE and Citrus analyses on synthetic datasets

Immunological response to a stimulus or a disease can comprise changes in relative abundance of a particular cell population and/or changes in expression levels of certain markers can also lead to the appearance of new cell phenotypes. The detection of both types of changes is essential for a comprehensive understanding of the etiology of the immune response.

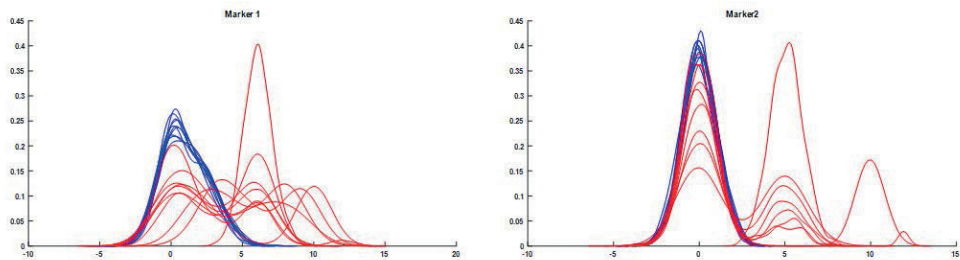
A two-dimensional synthetic dataset was created to show the performance of ECLIPSE and Citrus in identifying cell populations, characterized by differences in terms of abundance and expression levels of different markers. The synthetic dataset consists of 10 controls and 10 responders. The control group presents a single cell population with higher abundance than the response group. Heterogeneity has been introduced into the response group as the samples have diverse cell subpopulations, one of which consists of a small subset of 20 cells present only in one responder. This heterogeneity can be observed in the 2D scatter plot, Figure S25 (the small cell population in yellow in the top right corner belongs to Responder 10), and in the histograms shown in Figure S26; contour plots showing the differing relative abundance of cell populations between controls and responders are shown in Figure S27.



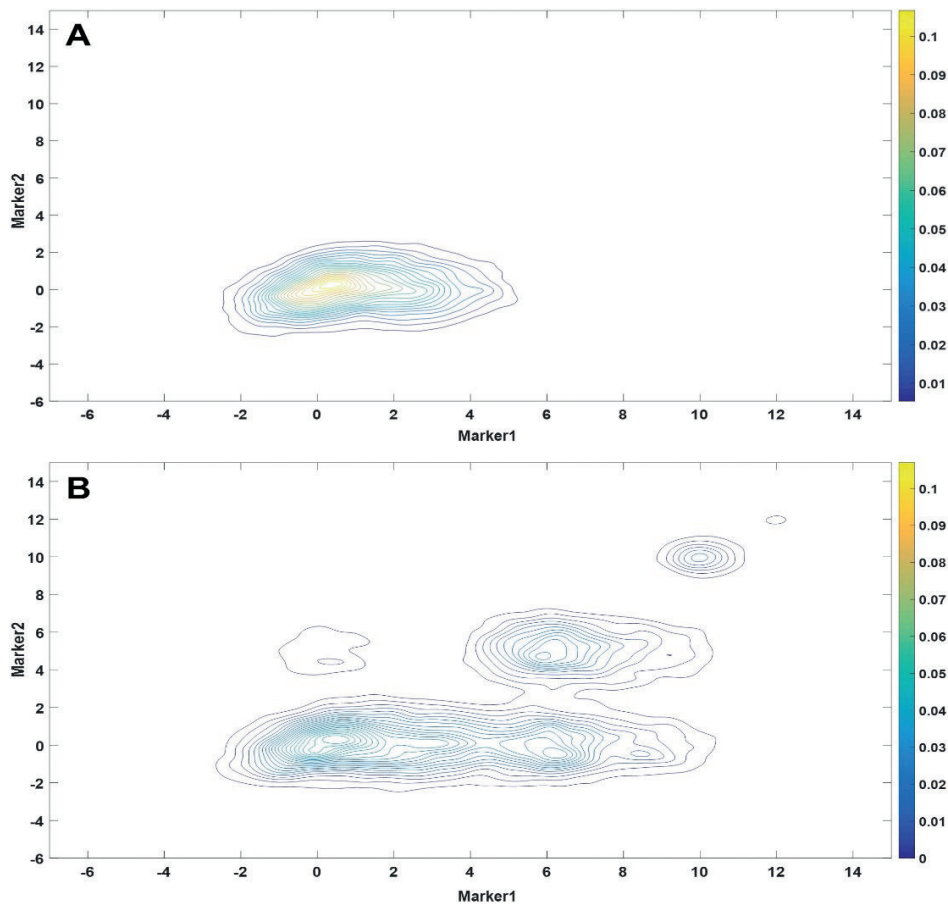
**Figure S25.**

2D scatter plot of the synthetic data. Cells from the control individuals are coloured in blue, while cells from the responders are differently coloured per individual.



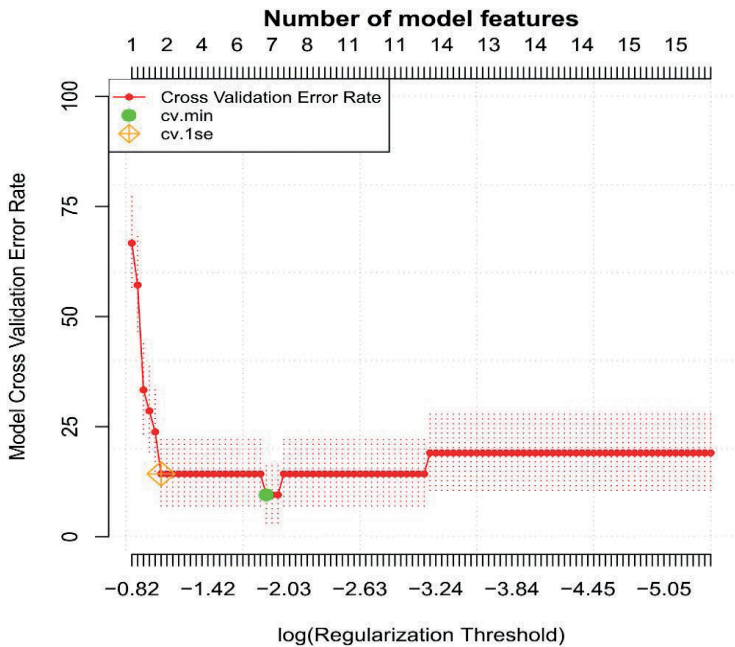


**Figure S26.** Histograms showing the expression of the 2 markers included in the synthetic data. Histograms of the control individuals are depicted in blue, while the histograms of the responders are in red.



**Figure S27.** A) contour plot of the cells distribution of control individuals. A single cell population is present for all the controls; B) contour plot of cells distribution for response individuals.

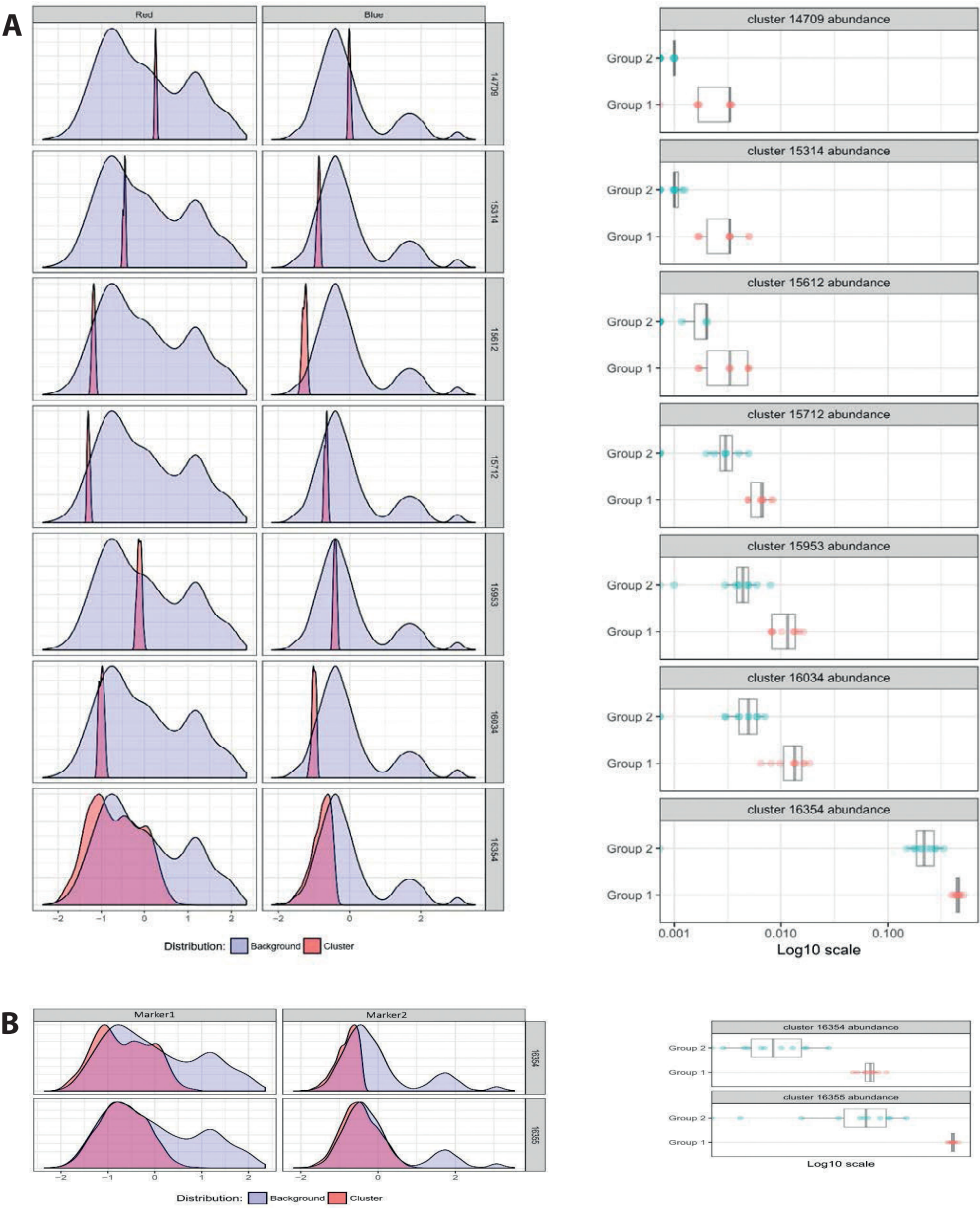
Citrus was applied to the dataset with a Minimum Cluster Size Threshold (MCST) corresponding to 10 cells, so that the method could identify the rare subset. The accuracy of the constructed models is shown in Figure S28, which reports the model cross-validation error rate versus the regularization threshold associated to the number of features identified. Seven features were detected by the model with the minimum error (cv.min), while the model with a cross validation error 1 std higher than the minimum (cv.1se) identified two most discriminant features.



**Figure S28.**

The Figure shows the Model Cross-validation Error Rate vs the log(Regularization Threshold) for the classification models built on the simulated data. The green circle (cv.min) points out the model with the smallest number of features necessary to obtain the lowest cross-validation error, which corresponds to 10% of misclassified samples; the orange diamond (cv\_1se) indicates the model with the smallest number of features associated to cross-validation error 1 std higher than the minimum error, which leads to an error of around 15%.

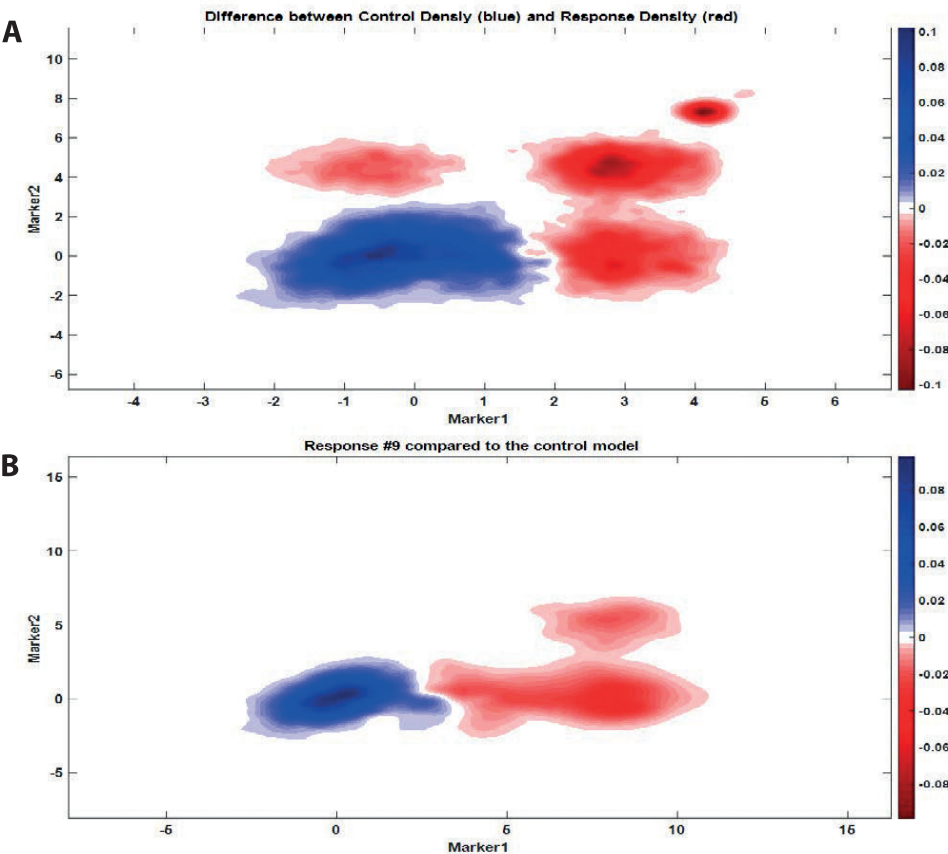
The features returned by both models are visualized in Figure S29. In both cases, the relative abundance of the identified cell clusters is higher in the control group compared to the response. These clusters correspond to the population with lower expressions of *Marker1* and *Marker2*, created as more abundant for the control samples. Although the information of differential cell population abundance might be relevant for describing an immune response, evidence about such relevant response-specific cell populations is missing from the optimally parsimonious Citrus model.



**Figure S29.**

A) Histograms show the phenotype of the discriminant clusters, found by the model associated with cv.min error (10% of misclassification). Differential abundance of these clusters between control (group 1) and response group (group 2) can be observed in the right panel: these are all found more abundant in the control group. B) Histograms show the phenotype of the discriminant clusters, found by the model associated with cv.1se error (15% of misclassification). Differential abundance of these clusters between control (group 1) and response group (group 2) can be observed in the right panel: they are more abundant in the control group.

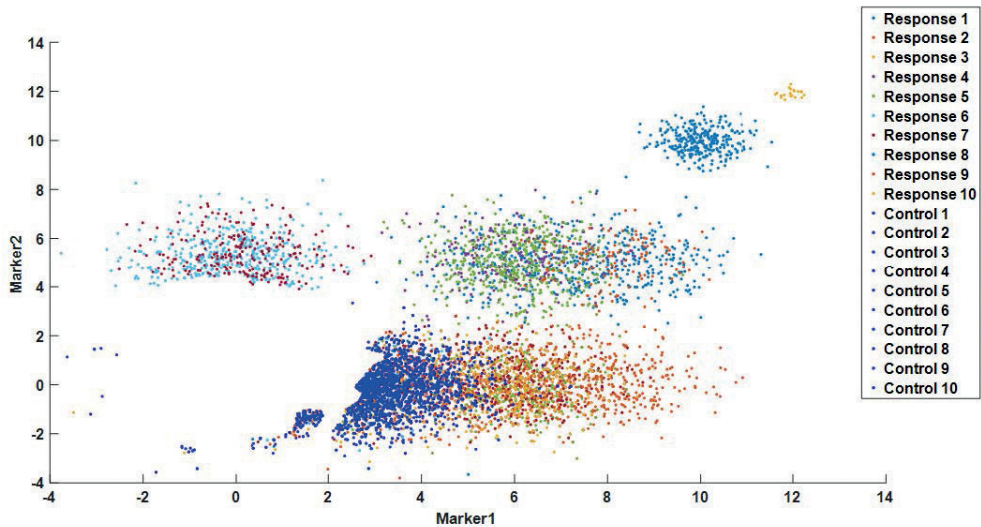
Secondly, we performed an ECLIPSE analysis on the synthetic dataset. The Difference between Densities (DbD) plot (Figure S30) clearly shows the difference between the two groups. Advantageous is the possibility to estimate the difference between densities of a single responder against the control group estimate (Figure S30B). This will enhance the resolution on a specific individual and it is helpful when one sample is available for the response class/classes.



**Figure S30.**

A) Difference between Densities plot, obtained by subtracting the KDE cell distribution of 10 responders from the cumulative KDE of the control cells distribution. The negative intensity (red) indicates the location where cells over-produced in the responder are more likely to be present; the positive intensity (blue) indicates the location where control cells are more likely to be present. B) Difference between Densities plot of the KDE of cell distribution of a responder (ID #9) subtracted from the cumulative KDE of the control cells distribution.

Response cells overlapping with the control marker variability were removed from the dataset and the remaining cells are displayed in Figure S31. The small population existing of 20 cells is still visualized (top right corner). Few cells are retained also for the control group, due to the individual variability we introduced between the control samples.



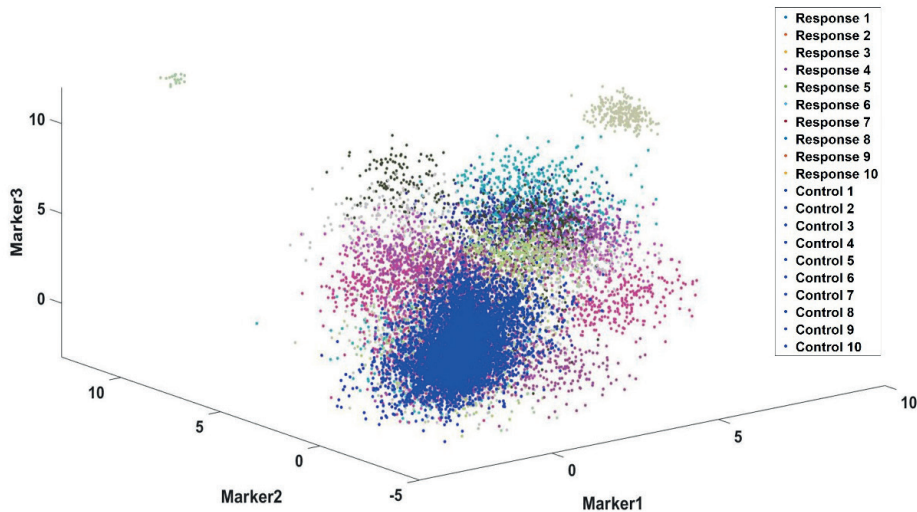
**Figure S31.**

2D scatter plot of the synthetic data, after the elimination of normal cells performed by ECLIPSE. Cells left for the control individuals, due to personal variability, are coloured in blue, while cells from the responders are differently coloured per individual.

The results of the analyses on the synthetic data showed how ECLIPSE outperforms Citrus in presence of a heterogeneity within the response class. In fact, Citrus is a two-class classifier method which needs numerous samples with similar phenotypically properties within both groups to find their signature features. Contrarily, ECLIPSE is a one class classifier method that requires only a group of control individuals. This is necessary to define a reference, against which a single patient or a group of patients, even highly heterogeneous in their response, can be compared. The removal of normal cells will put more focus on the response specific subpopulations of each individual, including rare cell subsets.

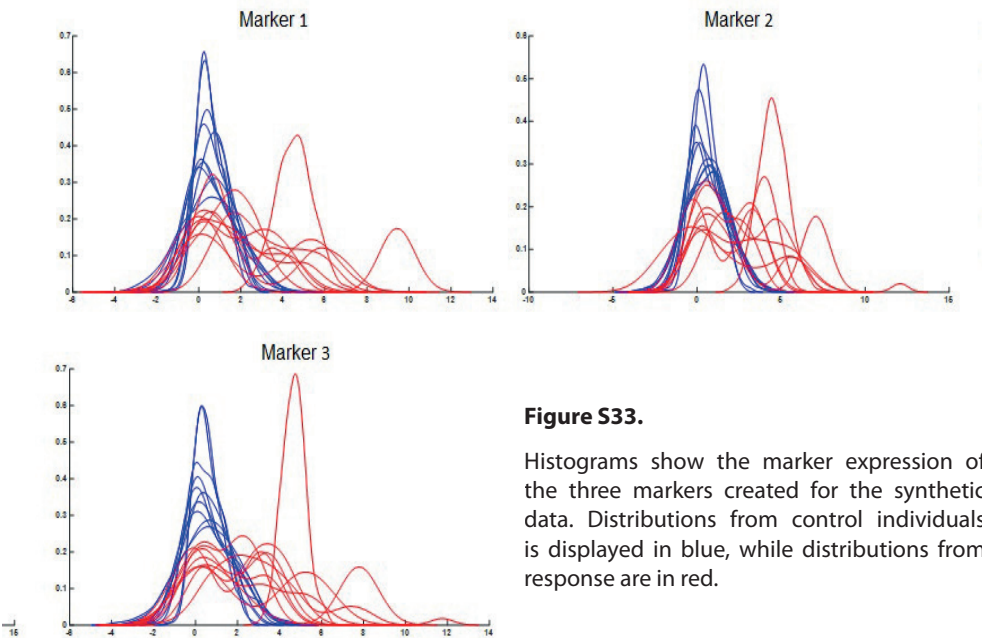
In this example, the dimensional reduction step of ECLIPSE was not needed because the data had only two dimensions. In order to show that the SCA-based transformation will not affect the discovery of cells subpopulations we performed the analyses on a 3D datasets, obtained from the first one after specific transformations.

A third random aspect was added to the first dataset; the matrix obtained was rotated by using orthogonal Procrustes rotation<sup>8</sup>, which allowed to introduce a more realistic correlation across all the three variables. A 3D scatter plot of the new dataset and histograms of the three marker expression levels are shown below (Figure S32 and S33).



**Figure S32.**

3D scatter plot of the 3 dimensional synthetic data. Cells from the control individuals are coloured in blue, while cells from the responders are differently coloured per individual.



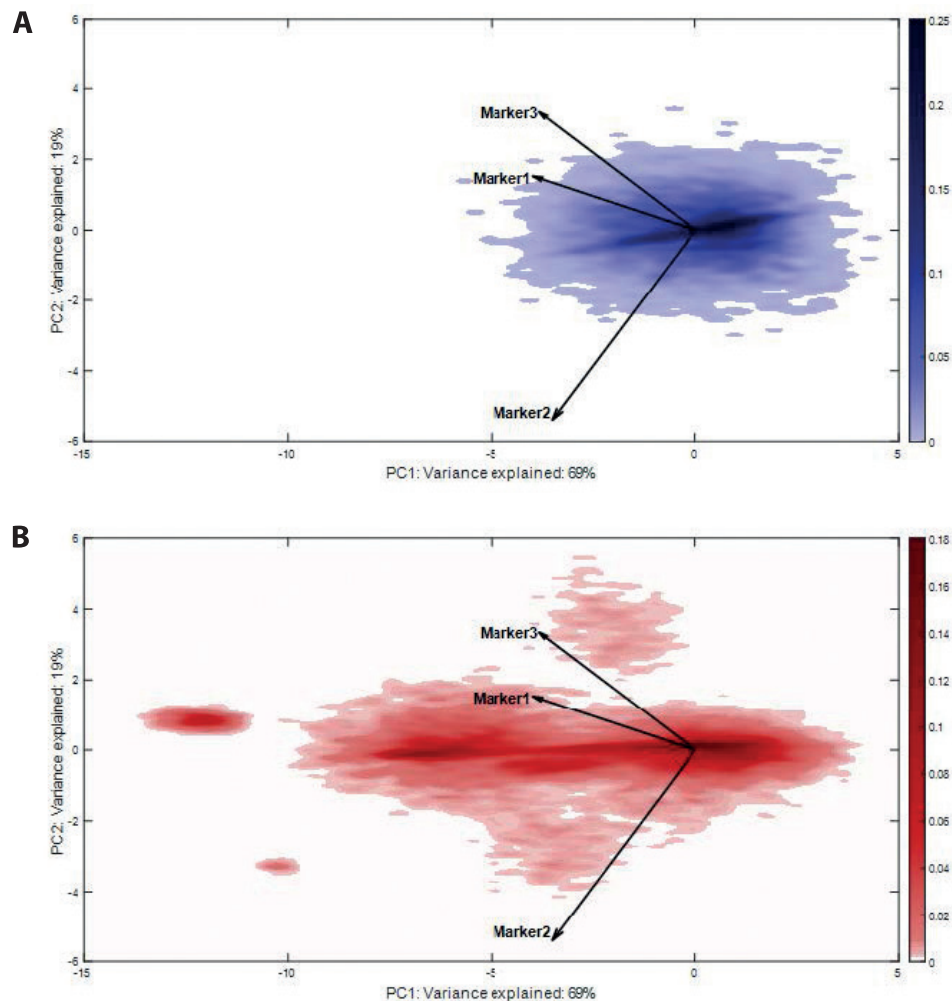
**Figure S33.**

Histograms show the marker expression of the three markers created for the synthetic data. Distributions from control individuals is displayed in blue, while distributions from response are in red.





ECLIPSE analysis was performed on the three-dimensional dataset, whose dimensionality was reduced with Simultaneous Component Analysis to two components. Figure S36 shows the Control Model with the density estimation of cell score distributions for the control (blue) and response group (red), together with the marker loadings. The axis show the variance explained by the model. The deviation of the responders (Figure S36B) from the reference group principally occurs along the first component PC1 and is mainly described by Marker1. Marker2 and Marker3 mostly explain the specific variability of cells along the second component PC2.

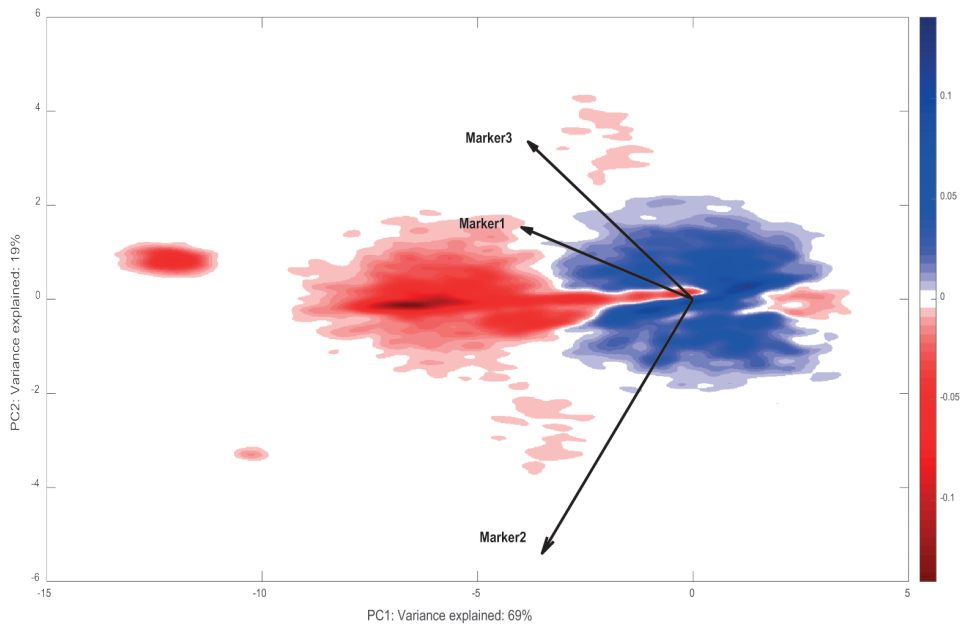


**Figure S36.**

A) shows the density estimate of the control cells scores distribution. B) shows the KDE estimates of the response cells scores distribution. Biplots of the KDE estimates in the Control Model, built on the variability of the control individuals. The loadings of Control Model are plotted as vectors:



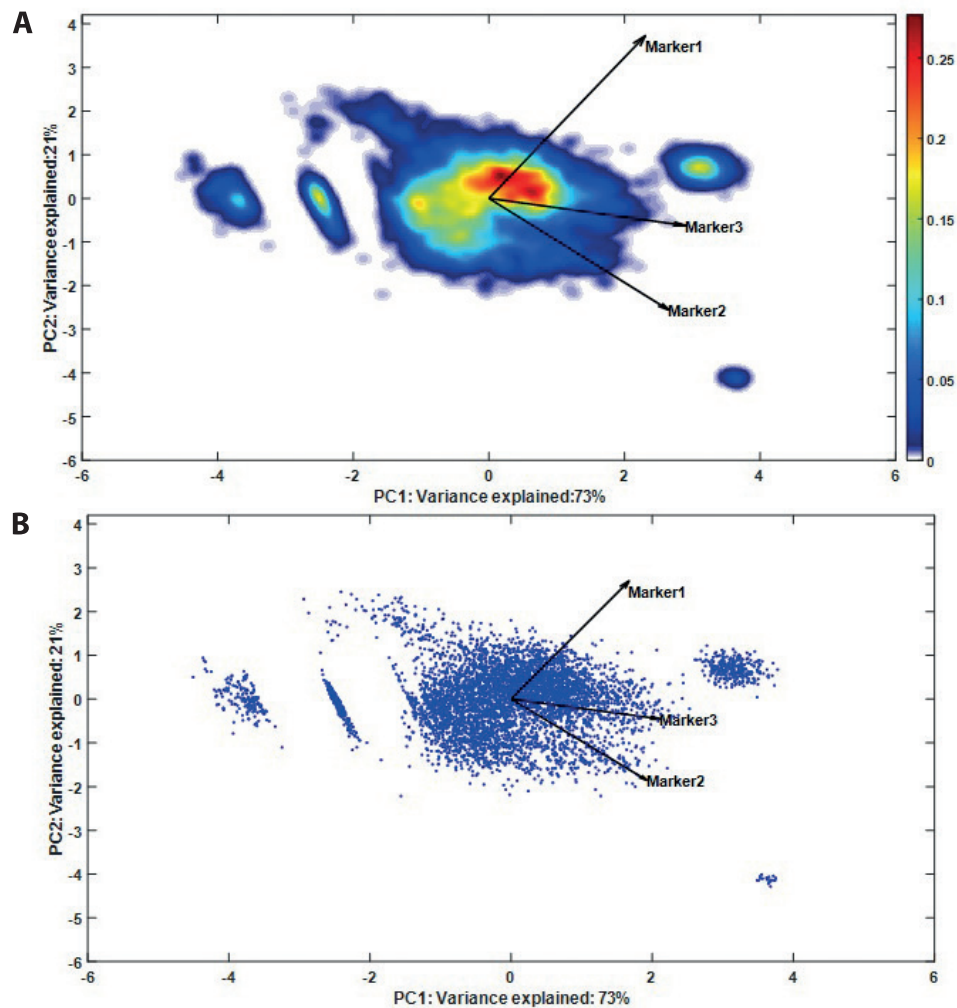
their length indicates the contribution of each marker to the cell-to-cell variability; the mutual directions suggest a positive (same direction) or negative (opposite direction) co-expression.



**Figure S37.**

Density estimate of the cells scores distribution from the responder individuals is subtracted from the density estimate of the control individuals. The negative intensity (red) specifies the location where cells over-produced in the responder are more likely to be present, while the positive intensity (blue) indicates the location where healthy cells are more likely to be present. The white area between the red and blue areas corresponds to a value of KDE=0, which can indicate bins with no cells or equal intensity of control and responder estimates.

Responder cells with a marker profile variability overlapping with the control variability were removed. A new SCA-based space is built on the variability of cells left after this removal and it specifically focuses only on the marker expression specific for the responder cells. The resulting ECLIPSE space is shown in Figure S38, which does show the rare cell population in down right corner. Moreover, the loadings show the correct co-expression between Marker 2 and 3 for this rare cell population which was masked in Figure S37, due to the high variability in the control cells of all markers.



**Figure S38.**

ECLIPSE model built on the cells of the responder individuals, after removal of normal cells (KDE representation, panel A; single cells representation panel B). The loadings show the marker co-expression specific for the response-specific cell subsets. The small population, consisting of 20 cells, is visible on the bottom right corner.

## References

- 1 Engel, J. *et al.* Breaking with trends in pre-processing? TrAC Trends in *Analytical Chemistry* 50, 96-106, doi:<https://doi.org/10.1016/j.trac.2013.04.015> (2013).
- 2 Roederer, M. Compensation in flow cytometry. *Current protocols in cytometry* Chapter 1, Unit 1.14, doi:[10.1002/0471142956.cy0114s22](https://doi.org/10.1002/0471142956.cy0114s22) (2002).
- 3 Finak, G., Perez, J. M., Weng, A. & Gottardo, R. Optimizing transformations for automated, high throughput analysis of flow cytometry data. *BMC bioinformatics* 11, 546, doi:[10.1186/1471-2105-11-546](https://doi.org/10.1186/1471-2105-11-546) (2010).
- 4 Timmerman, M. E., Hoefsloot, H. C. J., Smilde, A. K. & Ceulemans, E. Scaling in ANOVA-simultaneous component analysis. *Metabolomics* 11, 1265-1276, doi:[10.1007/s11306-015-0785-8](https://doi.org/10.1007/s11306-015-0785-8) (2015).
- 5 Bro, R. & Smilde, A. K. Centering and scaling in component analysis. *Journal of Chemometrics* 17, 16-33, doi:[10.1002/cem.773](https://doi.org/10.1002/cem.773) (2003).
- 6 Bruggner, R. V., Bodenmiller, B., Dill, D. L., Tibshirani, R. J. & Nolan, G. P. Automated identification of stratifying signatures in cellular subpopulations. *Proceedings of the National Academy of Sciences of the United States of America* 111, E2770-E2777, doi:[10.1073/pnas.1408792111](https://doi.org/10.1073/pnas.1408792111) (2014).
- 7 Amir el, A. D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology* 31, 545-552, doi:[10.1038/nbt.2594](https://doi.org/10.1038/nbt.2594) (2013).
- 8 Ten Berge, J. M. F. Orthogonal procrustes rotation for two or more matrices. *Psychometrika* 42, 267-276, doi:[10.1007/bf02294053](https://doi.org/10.1007/bf02294053) (1977).

# 6

## CHAPTER 6

# Unsupervised identification of rare cell subsets by multicolor flow cytometry applying the multivariate algorithm ECLIPSE

Selma van Staveren<sup>1,2\*</sup>, Rita Folcarelli<sup>3\*</sup>, Andries C. Bloem<sup>4</sup>, Geert Postma<sup>3</sup>, Gerjen H. Tinnevelt<sup>2,3</sup>, Nienke Vrisekoop<sup>1</sup>, Lutgarde M. C. Buydens<sup>3</sup>, Jeroen J. Jansen<sup>3§</sup>, Leo Koenderman<sup>1§</sup>

1 Department of Respiratory Medicine, Laboratory of translational immunology, University Medical Center Utrecht, Heidelberglaan 100, 3584CX, Utrecht, The Netherlands

2 TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands

3 Radboud University, Institute for Molecules and Materials, Analytical Chemistry, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands

4 Department of Immunology, University Medical Center Utrecht, Heidelberglaan 100, 3584CX, Utrecht, The Netherlands

\* and § these authors equally contributed to this work

The detection of rare cells in patient samples by multicolour flow cytometry (MFC) is becoming increasingly important. For instance, in haematological malignancies the detection of Minimal Residual Disease (MRD) is the key clinically relevant parameter for the evaluation of treatment effectiveness and risk of clinical relapse. MRD is the persistence of rare aberrant cells in patients treated for haematologic malignancy, such as Multiple Myeloma (MM). Multicolour Flow Cytometry (MFC) is often used for MRD monitoring as it is simple, fast, cost-effective and sensitive. However, detection of MRD by conventional analysis based on manual gating is highly subjective, labour-intensive and requires expertise to identify rare cells among an overwhelming majority of 'normal' cells. Here we show that a multivariate method we recently developed, called ECLIPSE, is particularly suited to detect rare disease-specific cell populations as in MM MRD described in this study, since the key feature of the method is an elimination step of normal cells in patient samples. We applied ECLIPSE on 8-color MFC data from bone marrow samples of 7 MM patients and control bone marrow samples. ECLIPSE enables fast and automated identification of MRD in all MM patients with a lower limit of detection of 0.001%. Moreover, in 5 out of 7 cases ECLIPSE could detect a larger number of monoclonal plasma cells than conventional manual gating. In conclusion, ECLIPSE is more sensitive, faster and requires considerably less expert effort than conventional rare cell detection with MFC.

## Introduction

Flow cytometry is a widely applicable high-throughput technique which measures the expression of proteins in and on single cells.<sup>1</sup> The most recently developed generation of flow cytometers can measure up to 50 parameters per cell.<sup>2</sup> This increasing number of parameters leads to the discovery of more and more cell subsets that occur in different abundances.

Studying rare cell types involves various challenges, primarily due to an overwhelming abundance of other cells. Rare event analysis requires measurement of at least 500,000 events per sample<sup>3</sup>, from which only a minute fraction is of interest. Conventional gating analysis highly relies on specific expertise of the analyst, and involves inspecting combinations of 1 or 2 markers sequentially. Cells of interest characterized by more than two markers may be lost, missed or not well discriminated in every sequential gating step. Although gating may be well-protocolled, it will always suffer from analyst error like every manual assessment.

Several multivariate analysis (MVA) methods<sup>4-10</sup> have been developed for fast, unbiased and reproducible analysis of MFC data. Although all these techniques overcome the limited observation of two markers in manual gating, they are not specifically suited for identification and characterization of rare cells as they simultaneously describe all cells within the sample, which results in a bias towards describing the most abundant cell populations.<sup>4-8</sup> We recently introduced Elimination of Cells Lying in Pattern Similar to Endogeneity (ECLIPSE)<sup>9</sup>, a method that specifically identifies disease-related cells by eliminating the more abundant healthy-like cells that also occur in multiple control samples.<sup>9</sup> Firstly, the ECLIPSE algorithm consists of building an SCA model<sup>11</sup> to describe the common variance present in the control samples. Secondly, the normal-like cells are removed from the dataset (both for patient and control samples) and subsequently the variance of the remaining disease-specific cells is described by the ECLIPSE model. The resulting model consists of a map of multidimensional surface marker co-expressions of cells found only within diseased samples. This resolves disease-specific cell populations to enable their count and characterization. Here we show that ECLIPSE is very well suited for the identification of rare cell subsets, using a proof of principle dataset: minimal residual disease in patients with Multiple Myeloma.

Multiple Myeloma (MM) is the lymphoproliferative, clonal expansion of plasma cells in the bone marrow. MM treatment has changed in the past two decades to result in higher rates of complete response (CR).<sup>12</sup> However approximately two-thirds of patients that reached CR still relapse within 10 years.<sup>13</sup> As in other haematological malignancies, in MM the extend of response to therapy is correlated to the clinical outcome.<sup>14</sup> Presence of Minimal

Residual Disease (MRD) shortens progression-free survival and overall survival.<sup>15</sup> The most widely used technique to detect MRD in MM patients is multi-colour flow cytometry (MFC), because of its applicability to all patients, high analysis throughput rate, relatively low costs and multi-parameter capabilities.<sup>14,16,17</sup> Current practice involves laborious manual gating by specialist technicians, such that validated automation of this step promises changes in resource-efficiency and diagnostic robustness.

We first tested the ability of ECLIPSE in identifying rare cell populations in a synthetic dataset having an increasing percentage of MRD-like cells within control bone marrow cells. This sensitivity test showed that ECLIPSE enabled to successfully detect a percentage of MRD-like cells of 0.001%. After this, we applied ECLIPSE to a real dataset of MM diagnosis and MRD-containing follow-up samples of seven MM patients, against 13 control bone marrow samples. We compared performance against manual gating. ECLIPSE was capable of detecting and quantifying MRD cells using the patients' diagnosis sample containing the primary MM with a procedure that is automated, intuitive and faster than conventional gating. ECLIPSE also outperforms Hierarchical SNE (HSNE)<sup>10</sup>, reported to be specifically suited for the discovery of rare cell populations. In conclusion, ECLIPSE is a promising MVA method for automated detection of rare disease-specific cell populations in flow cytometry data.



## Methods

### Spiking control bone marrow sample with clonal plasma cells

We simulated data of 10 bone marrow samples containing 10%, 5%, 2%, 1%, 0.5%, 0.1%, 0.05%, 0.01%, 0.005% and 0.001% clonal plasma cells among 200.000 cells in total. This was done by adding manually gated clonal plasma cells from one diagnosis sample to one control bone marrow samples (FlowJo analysis software, Tree Star Inc., Ashland, Oregon).

### Multiple Myeloma dataset

Bone marrow samples were analysed using an eight-color-panel for clonal plasma cells by MFC on a FACSCanto II (Becton Dickinson). The panel includes six surface markers: CD45, CD38, CD138, CD27, CD19, CD56 and 2 cytoplasmic markers: Cylgk and Cylgλ.<sup>18,19</sup> The compensated FCS files of seven MM patients at diagnosis and at a follow-up measurement (ranging from 6 months to 3 years after diagnosis) were included in this study. At follow-up, for all seven patients we manually gated for the presence of clonal plasma cells (ranging from 0.005-0.5 % of total cells). We also negated presence of such cells in bone marrow samples from 13 samples from individuals with no indication for presence of haematological malignancies or any other abnormality (control individuals). The number of events recorded per sample ranged from approximately  $10^5$ - $10^6$  per sample. All 27 FCS files were gated on cells based on FSC/SSC, excluding debris (see Figure S1 for gating strategy, Online Supplementary Material).

### Manual gating strategy of clonal plasma cells

For all diagnosis and follow-up files we manually gated clonal plasma cells according to the scheme in Figure S2 in the Online Supplementary Material I. We compared the percentages of the identified manually gated cells to those found by ECLIPSE.

### Elimination of Cells Lying in Pattern Similar to Endogeneity (ECLIPSE)

The ECLIPSE method and underlying algorithm<sup>9</sup> first pre-processes the data as described in Supplementary Material I, slightly adapted from the original method<sup>9</sup> to accommodate rare cell discovery in clinical monitoring. The method then reduces dimensionality of the MFC data by Simultaneous Component Analysis with equal pattern (SCA-P)<sup>11</sup>, the extension of Principal Component Analysis to multiple sets. The key feature of ECLIPSE is the initial identification of normal-like cells in diseased samples by quantitative comparison to the 13 control individuals and the subsequent elimination of these cells from the diagnosis MM samples (Supplementary Material I). A second dedicated ECLIPSE model, then describes

the multi-cellular variability of those cells not previously eliminated, either *per* MM patient or within a single model for all MM patients combined respectively. Next, the normal-like cells eliminated from a presented follow-up sample. The remaining cells from the follow-up sample are then projected in the ECLIPSE space based on the clonal MM cells from the diagnosis sample described before. This results in a map of the cells within the follow-up sample *not* found in the control samples, using the most prominent multidimensional marker co-expressions of MM-specific cells of that individual. Clonal plasma cells can be identified using the surface marker loadings that serve as a compass-like representation of their marker co-expression. MRD cells should not overlap with donor-specific control bone marrow cells that were not eliminated before the ECLIPSE model, which also may be projected into ECLIPSE. Clonal plasma cells identified in the ECLIPSE space were gated and the number of events associated to MRD cells was compared to the number found by sequential manual gating.

### Recovery of manual gated cells in ECLIPSE

We found MRD populations between 0.005% and 0.5% by manual gating , with absolute cell counts ranging from as few as 19 cells to one individual with around 4000 cells. We estimated the ability of ECLIPSE to detect the cells found by manual gating through the recall in eq.1 <sup>20</sup>, which quantifies how many relevant cells the model allows to find within the datasets and it can be used as measures of how well the manual gated MRD cells are retrieved by ECLIPSE.

$$recall_{i_g} = \frac{\#cells (manual\ gating_{i_g} \cap ECLIPSE\ gating_{i_g})}{\#cells\ manual\ gating_{i_g}} \quad \text{Eq. 1}$$

Where the numerator represents the intersect of the cells found by manual and ECLIPSE gating and the denominator represents the total number of cells identified by manual gating;  $i_g$  indicates the individual within the follow-up group  $g$ .

## Results

### ECLIPSE detects MRD as low as 0.001%

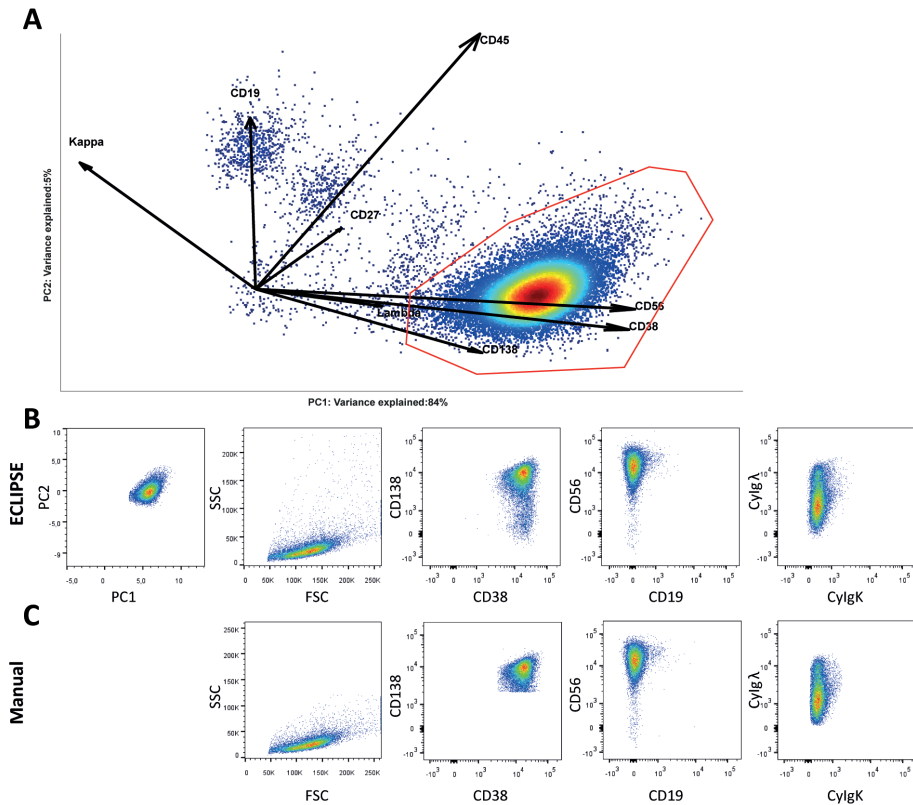
Using a single eight-color MFC panel and manual gating, the lower detection limit for MRD lied between to 0.01% to 0.001% of total bone marrow cells.<sup>21,22</sup> We used a synthetic dataset to test the detection limit of ECLIPSE by virtually spiking a control bone marrow sample with different numbers of gated monoclonal plasma cells (see Methods). After automated removal of control bone marrow cells by ECLIPSE, we could identify clonal plasma cells in all the spiked samples, also for the very rare MRD populations comprising of 20, 10 or even 2 cells (corresponding to 0.01%, 0.005% and 0.001% of total bone marrow cells), as shown in Supplementary Material II.

### The MM-specific ECLIPSE model identifies rare MRD cells in MM patient samples

Next to virtually spiked samples, we evaluated the detection of MRD cells in samples of 7 MM patients using ECLIPSE. All quantitative preprocessing steps are shown in Supplementary material III.

The projection of the diagnosis samples into the control model showed considerable heterogeneity of the disease-related cells between patients (Supplementary Material Figure. S8). In order to individually interpret the phenotype of the disease-related cells, we separately modelled each patient in a dedicated ECLIPSE model after elimination of control-like cells.

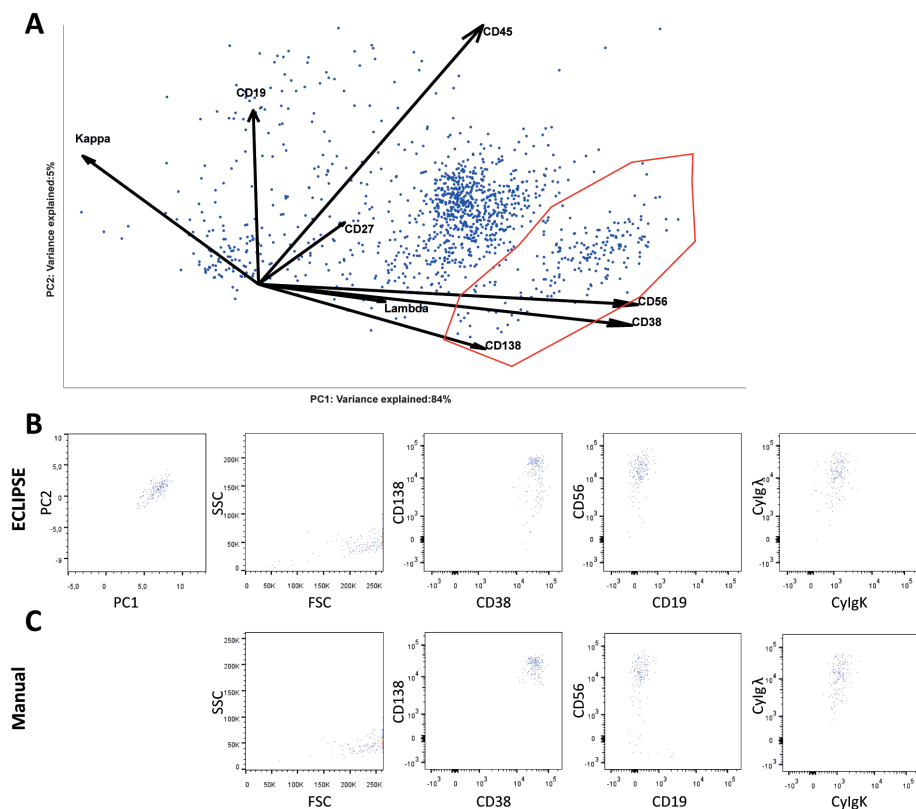
After elimination of normal-like cells, we retained around 14% of the original amount of cells from Diagnosis sample ID1, being distinct from the surface marker expression of the control bone marrow cells. This down-sampling increased focus on the MM cells in the diagnosis-specific model (Figure 1A), compared to the representation of the full sample projected in the Control model (Supplementary Material IV). In Figure 1A, the major cell population (gated in red) is characterized by higher (co-)expression of the markers CD138, CD38, CD56 and Cylgλ. The opposite orientation of the CylgK loading confirms clonality of Cylgλ, validating identification of monoclonal plasma cells. We compared the cells gated by ECLIPSE to the phenotype of the manually gated cells (Figure 1B+C). The identical phenotypes of ECLIPSE and manually identified malignant plasma cells confirm validity of the ECLIPSE results for MM.



**Figure 1.**

A) ECLIPSE model of the diagnosis sample ID1. The control bone marrow-like cells have been eliminated from the analysis and a PCA space is computed on the ‘abnormal’ cells (14% of the original cells), after mean centering and scaling using the FMO control samples, the loadings represent the marker co-expression specific for the ‘abnormal’ cells. Based on the marker co-expression the MM cells were identified (gated in red). B) The ECLIPSE gated cells were exported and their phenotype was evaluated in various biplots, from left to right: ECLIPSE PC1/PC2; FSC/SSC; CD38/CD138; CD19/CD56; Kappa/Lambda. C) FSC/SSC; CD38/CD138; CD19/CD56; Kappa/Lambda biplots for evaluation of the phenotype of the manually gated monoclonal plasma cells.

We then projected cells of the follow-up sample that were identified as “abnormal” onto the ECLIPSE model of the corresponding individual (Figure 2A). This retained only 0.8% of cells from the complete ID1 follow-up sample. The compass-like representation of the cellular marker loadings can be used to characterize the cells. Since treatment might induce changes in marker expressions,<sup>23</sup> we placed an MRD-specific gate around the cell population with high CD138, CD38, CD56 and CylgA expressions. We could also confirm identification of monoclonal plasma cells in the follow up samples by ECLIPSE as their surface marker expression overlaps with the phenotype of the manually gated monoclonal plasma cells (Figure 2B+C).

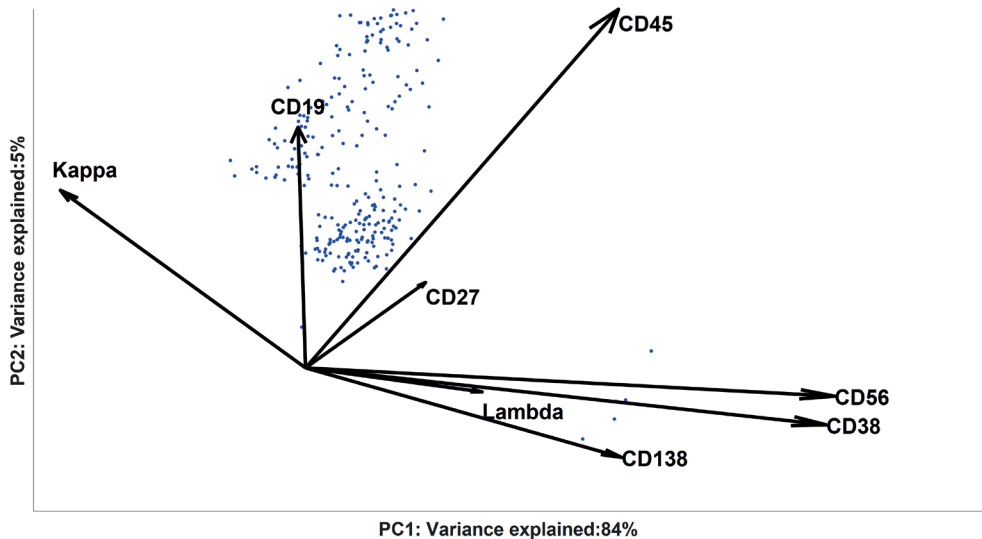


**Figure 2.**

A) Projection of the cells of the follow-up sample (ID1), after elimination of control bone marrow associated cells, onto the corresponding diagnosis sample-specific ECLIPSE model. Based on the marker co-expression the MRD cells, gated in red, can be identified. B) The ECLIPSE gated cells were exported and their phenotype was evaluated in various biplots, from left to right: ECLIPSE PC1/PC2; FSC/SSC; CD38/CD138; CD19/CD56; Kappa/Lambda. C) FSC/SSC; CD38/CD138; CD19/CD56; Kappa/Lambda biplots for evaluation of the phenotype of the manually gated monoclonal plasma cells.

The earlier built Control Model may not describe all donor-specific cells within the control bone marrow samples equally well, because the phenotype of some cells may only be expressed in a limited donor subset. We projected these donor-specific benign cells into each (individual) diagnosis-specific ECLIPSE map. Figure 3 shows these cells for all 13 control bone marrow samples in the ECLIPSE model of the first MM patient described above (ID1). We found very low fractions of donor-specific control bone marrow cells (0.001%-0.02%) to overlap with non-myeloma cells in the patient samples (Supplementary Figure S10A-C). These cells express CD45 and CD19 (Supplementary Figure S10D), corresponding to a B cell phenotype. The diagnosis and follow-up samples of ID1 show extra cell populations that do not overlap with control-specific cells and not with the previously gated MM/MRD cells. We evaluated the single-marker expressions of these additional populations

(Supplementary Figure S10D) and concluded these cells were neither monoclonal plasma cells nor B cells, since they did not express CD38, CD138 and CD19. Further identification beyond B or plasma cells is hampered by the limited panel.



**Figure 3.**

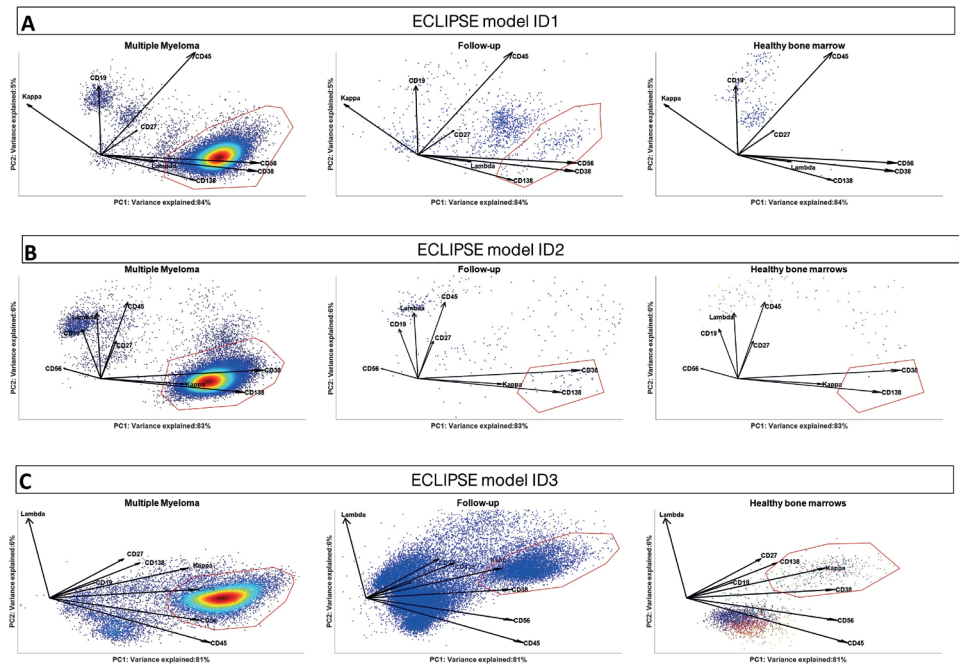
Projection of cells from the 13 control bone marrow samples after elimination of the majority of normal cells into the diagnosis sample (ID1) specific ECLIPSE model. These cells may be associated to donor-specific variability not explained by the Control Model.

ECLIPSE models may vary considerably between individuals as shown in Figure 4 for ID1-3 and Suppl Figure S11 for ID4-7. Nevertheless, the direction of the loadings led to the quick identification of clonal plasma cells, gated in red.

### False positive discovery rate

False positive discovery was investigated by projecting the donor-specific cells of control bone marrow samples in the diagnosis-specific model and the amount of these cells within the MRD gate was estimated. This was empty or with a lower percentage of control bone marrow cells than the sensitivity limit considered to diagnose relapse ( $<0.001\%$ ), except for ID3 and ID7. In the latter two patients there was an average percentage of  $0.004\%$  and  $0.01\%$  of cells from the 13 control bone marrow overlapping with the region of the MRD clone, respectively (Figure 4C and Suppl Figure S11). Further investigation identifies these control bone marrow cells as CD38+CD138+ plasma cells, all expressing Cylgk (Suppl Figure S12). In both cases ECLIPSE eliminated the Cylgk+ plasma cells from the control

bone marrow samples, only the Cylgk+ plasma cells remained. Since the myeloma cells of ID3 and ID7 were Cylgk+ and (part of) the myeloma cells had a normal plasma cell phenotype, these MM cells may overlap with the Kappa+ plasma cells of controls. In the case of ID3, control bone marrow cells were indeed also overlapping with the MM-specific area, suggesting that the phenotype at diagnosis may slightly resemble the normal plasma cells of control bone marrow samples. This may be a case when ECLIPSE and any other uni- or multivariate analysis method would fail in detecting MRD positivity, as we expect a 'different from normal phenotype'.



**Figure 4.**

ECLIPSE models build on ID1 (A), ID2 (B) and ID3 (C) Samples are depicted after ECLIPSE elimination of normal cells from 13 healthy controls. The left graph shows the ECLIPSE model with the cells of the diagnosis sample, the middle graph shows the cells of the follow-up sample projected inside the diagnosis ECLIPSE model. The right graph shows the donor-specific cells of all control bone marrow samples projected in the diagnosis ECLIPSE model (the cells are coloured per control bone marrow sample).

**ECLIPSE may identify higher percentage of MRD cells than manual gating**

We compared the overlap between MRD cells identified in the ECLIPSE-based gate vs those found by the manual gating benchmark (Table 1). The recall statistics (see Methods) shows the ability of the ECLIPSE algorithm in retrieving manual gated MRD cells. Recall ranges between 0 and 1: the higher the recall value, the better is the reproducibility of the manual gating findings by the ECLIPSE algorithm.

We obtained recalls equal to or higher than 0.9 for four follow-up samples. The analyses on the remaining three samples provided lower recalls, going from the minimum recall of 0.53 for ID2 to suboptimal recalls of 0.75 and 0.77 for ID4 and ID5, respectively. For the recall statistics, manual gating is the reference method. It can indeed happen that ECLIPSE may not retrieve cells found by manual gating which might be slightly more similar to a control cell phenotype, resulting in a low recall. On the other hand, it should also be evaluated whether ECLIPSE identified additional myeloma cells which were missed with manual gating. This is quantified by the fold-increase statistics in Table 1. For most of the individuals (5 out of 7 samples, with only ID5 and ID7 as exceptions), ECLIPSE detected more monoclonal plasma cells than manual gating (fold-increase >1). In the most critical case of ID2 ECLIPSE retrieved only around half the amount of the MRD cells from the manual gate (recall of 0.53). However, the algorithm identified a considerable higher amount of total monoclonal cells corresponding to 0.011% MRD cells of the total follow-up sample which is a doubled percentage compared to manual gating. Also for ID4, despite the suboptimal recall of 0.75, the amount of cells identified by ECLIPSE is 2.25 times higher compared to manual gating.

More challenging were the results for ID5, having suboptimal recall 0.77 and a fold-increase of 0.94. For both the diagnosis sample (not shown) as well as the follow-up ID5, ECLIPSE found a lower percentage of MRD cells than manual gating. Here investigating the phenotype of the manual gated cells displayed that some of the manually identified malignant cells were similar to control bone marrow cells which is most likely the reason why they were not found in the diagnosis-specific ECLIPSE model, since they were eliminated in the earlier Control Model.

In conclusion, ECLIPSE was able to automatically identify MRD cells in all 7 cases, with the same number or even higher numbers of cells when compared to manual gating (fold increase 0.94-2.25). For all the individuals, fold-increase was calculated excluding the presence of false positive discoveries. For this we thoroughly investigated the phenotype of the additional cells identified by ECLIPSE. The additional monoclonal plasma cells identified by ECLIPSE which were not found by manual gating, had higher SSC intensities and/or lower expressions of CD138 compared to the phenotype of the manually gated



cells (Figure S9). Manual gating (Supplementary Material I) excluded the cells with higher SSC since these cells fall outside the FSC/SSC lymphocyte gate. Plasma cells were then gated for their CD38<sup>high</sup>CD138<sup>high</sup> phenotype, thereby excluding the CD138<sup>dim</sup> cells. Multiple reports have indeed shown the presence of CD138<sup>low</sup> monoclonal plasma cells in MM cell lines as well as patient samples.<sup>24,25</sup> It is thought that the micro-environment might influence the clonogenicity and stimulates the de-differentiation of CD138<sup>high</sup> monoclonal plasma cells to a CD138<sup>low</sup> phenotype.<sup>26,27</sup>

**Table 1.**

Comparison of the overlap of cells found in the ECLIPSE gate with the manual gate. Recall statistics are calculated per follow-up sample of each individual quantifying the number of manually gated cells also found by ECLIPSE. Fold-increase summarizes the number of cells in the ECLIPSE gate compared to the manual gate.

Note: the fold-increase is calculated using the absolute number of cells as described in Equation 1.

	% monoclonal plasma cells in ECLIPSE gate	% of cells in ECLIPSE gate overlapping with cells in manual gate	% monoclonal plasma cells in manual gate	Recall	Fold-increase ECLIPSE
ID1	0.13	0.095	0.11	<b>0.89</b>	<b>1.18</b>
ID2	0.011	0.0030	0.0057	<b>0.53</b>	<b>2.00</b>
ID3	0.76	0.62	0.65	<b>0.98</b>	<b>1.17</b>
ID4	0.0076	0.0025	0.0034	<b>0.75</b>	<b>2.25</b>
ID5	0.40	0.33	0.43	<b>0.77</b>	<b>0.94</b>
ID6	0.30	0.28	0.30	<b>0.98</b>	<b>1.02</b>
ID7	0.43	0.40	0.44	<b>0.91</b>	<b>0.98</b>

To compare the performance of ECLIPSE to another multi-dimensional method, we analysed the same samples (ID1-3 ) with HSNE<sup>10</sup>. These results are shown in the Online Supplementary Material VI. In all the HSNE analyses, expertise was needed to choose which clusters to zoom in on until the single cells level to allow the identification of MRD cells. Various tentative gating steps led to identify MRD cells with a similar recall for ID1-3 and higher recall for ID2 than ECLIPSE. However the false discovery was considerable higher than ECLIPSE because a significant percentage of non-plasma cells from diagnosis, follow-up and control samples were found in the same cluster associated to the MRD cells.

## Discussion

Analyses based on conventional manual gating are labour intensive and require dedicated expertise, because a certain order of several possible pairwise combination of markers need to be thoroughly investigated in order to find the cells of interest. Additionally, intrinsic drawbacks of the two-marker perspective of manual gating become critical. This holds true especially for studying small cell populations such as MRD cells, because their identification may be hampered by the preponderant majority of normal-like cells. ECLIPSE provides a higher resolution compared to current conventional algorithms on disease-specific cells because it eliminates the predominant normal cells.<sup>9</sup> Here we show that this key advantage indeed facilitates the quantitative detection of small disease-specific cell populations.

As a first proof-of-concept, we run ECLIPSE on a synthetic combination to test the detection limit of the algorithm by reproducing MRD-like samples. ECLIPSE identified rare cells in all the spiked samples, also when very few MRD-like cells were present (minimum percentage in this dataset of 0.001%).

Next, we applied ECLIPSE to a real dataset containing bone marrow samples of MM patients at diagnosis and at follow-up characterized by MRD positivity. This was compared with manual gating. ECLIPSE detected MRD cells for 7 out of 7 cases, in a fast, unbiased, and data-driven approach. Projecting control bone marrow samples, containing non-MM cells into the ECLIPSE model serves to calculate the fraction of 'control-like' cells identified as MRD and thus provides indication of false positive discovery by the algorithm. For most of the cases (5 out of 7 analyses) no control bone marrow cells or an amount below limit of significance ( $< 0.00001\%$ ), were found in the ECLIPSE-based MRD gates. On the other hand, the ECLIPSE MRD gates of IDs 3 and 7 contained a clinical relevant fraction of control plasma cells (average of  $\sim 0.01\%$  for the 13 control bone marrow samples) in the established MRD gate. This could be due to the control group of control bone marrow samples ( $n=13$ ) being too limited to describe a control reference plasma cell population sufficiently well. Increasing the number of control bone marrow samples and possibly adding reference samples of regenerating bone marrow to build the Control Model would reduce the false positive discoveries. This will enable an accurate definition of the variability of the control plasma cells and remove more control-like cells (normal and regenerating bone marrow cells) in both control and patient individuals.

The ECLIPSE algorithm results were then benchmarked against the manual gating standard. Recall and the fold-increase statistics were calculated. High recall statistics were obtained for 6 out of 7 analyses, ranging from 0.75 to 0.99, indicating that the ECLIPSE algorithm sufficiently allowed to reproduce the findings of the conventional analysis. A

recall  $< 1$  can be seen as a measurement of (high) false negative rate. However, ECLIPSE identified a higher percentage of monoclonal cells compared to manual gating (fold-increase, Table 1). The high fold-increase highlights one of the main advantages of ECLIPSE compared to the dimensionality-limited and subjective approach of manual gating. In fact, the multivariate and unbiased ECLIPSE method detected malignant cells which were often challenging to identify using manual gating, such as CD138<sup>low</sup> monoclonal plasma, because they overlapped with other CD138<sup>low</sup> bone marrow cells in a CD38/CD138 biplot.

Analysis of MRD with ECLIPSE is a very relevant example of ‘personalized healthcare’: the ECLIPSE model serves as a personalized fingerprint that can be used as a digital benchmark for checkups and diagnoses in a clinical long-term monitoring setting. In clinical practice, the diagnosis and follow-up samples of one patient will generally be measured with the same panel. This well matches with the workflow of ECLIPSE in building a diagnosis-specific ECLIPSE model. When the diagnosis sample is measured in the clinic, the patient specific ECLIPSE model can be generated and sparse monoclonal plasma cells can be identified by projecting incoming follow-up samples. Subsequently it should be evaluated whether the monoclonal plasma cells of the patient can be identified by ECLIPSE and whether there are control bone marrow cells overlapping with the MM clone. If the MM-specific area is empty in the control samples, the patient’s responsiveness to therapy can be followed-up over time by projecting newly measured follow-up samples into the ECLIPSE model. In the ID3-specific ECLIPSE analysis control plasma cells of control BM samples were found overlapping with the MM-specific area, and in this case it is not possible to check for MRD positivity using ECLIPSE. This would already be known at the time of diagnosis and this patient should be selected for follow-up using manual gating, since for manual gating this problem will not occur as control bone marrow samples are not used to find MRD.

Next to MM, the treatment responsiveness of other haematological malignancies are also monitored using flow cytometry.<sup>28,29</sup> These malignancies have different characteristics regarding their clonality and the appearance of aberrant markers on the cell surface over time.<sup>30,31</sup> For Acute Myeloid Leukaemia the aberrant markers on the MRD cell surface frequently differ from the aberrant markers present on the initial AML cells at the time of diagnosis. Thus the ECLIPSE space based on the diagnosis sample might not be optimal to distinguish MRD at follow-up. Also multiple phenotypically different AML populations can be present within one patient.<sup>32,33</sup> These aspects challenge detection of MRD cells, but these may be mitigated by further extensions of ECLIPSE. The potential of ECLIPSE to detect monoclonal MRD reported here is an essential prerequisite for this.

In conclusion, the ECLIPSE method is well suited for the identification of disease-specific rare cells, as has been shown for MRD cells in MM patients. The removal of healthy cells enhances the focus on cells associated to the disease studied. Generating a patient-

specific ECLIPSE space based on the diagnosis sample and projecting future follow-up samples within this space allows straightforward identification of MRD cells in or near the MM region. The fact that the number of monoclonal plasma cells found by ECLIPSE was higher compared to manual gating is a promising finding. This increased sensitivity could be of importance for diagnostic purposes. The next step will be the verification of the performance of ECLIPSE in a prospective study with a large control reference group and a large group of MM patients undergoing therapy.

## References

- 1 Robinson, J. P. & Roederer, M. Flow cytometry strikes gold. *Science* 350, 739-740, doi:10.1126/science.aad6770 (2015).
- 2 Chattopadhyay, P. *et al.* Toward 40+ parameter fluorescence flow cytometry. (2014).
- 3 Stetler-Stevenson, M. *et al.* Consensus guidelines for myeloma minimal residual disease sample staining and data acquisition. *Cytometry Part B: Clinical Cytometry* 90, 26-30, doi:10.1002/cyto.b.21249 (2016).
- 4 Amir el, A. D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology* 31, 545-552, doi:10.1038/nbt.2594 (2013).
- 5 Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nature biotechnology* 29, 886-891, doi:10.1038/nbt.1991 (2011).
- 6 Van Gassen, S. *et al.* FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* 87, 636-645, doi:10.1002/cyto.a.22625 (2015).
- 7 Liu, Y. J. *et al.* Angle distribution of loading subspaces (ADLS) for chemical rank estimation in three-way analysis. *Chemometrics and Intelligent Laboratory Systems* 152, 146-156, doi:10.1016/j.chemolab.2015.11.006 (2016).
- 8 Tinnevelt, G. H. *et al.* Novel data analysis method for multicolour flow cytometry links variability of multiple markers on single cells to a clinical phenotype. *Scientific Reports* 7, 5471, doi:10.1038/s41598-017-05714-1 (2017).
- 9 Folcarelli, R. *et al.* Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm. *Scientific Reports* 8, 10907, doi:10.1038/s41598-018-29367-w (2018).
- 10 van Unen, V. *et al.* Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. *Nature Communications* 8, 1740, doi:10.1038/s41467-017-01689-9 (2017).
- 11 Timmerman, M. E. & Kiers, H. A. L. Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika* 68, 105-121, doi:10.1007/BF02296656 (2003).
- 12 Moreau, P., Attal, M. & Facon, T. Frontline therapy of multiple myeloma. *Blood* 125, 3076-3084, doi:10.1182/blood-2014-09-568915 (2015).
- 13 Barlogie, B. *et al.* Curing myeloma at last: defining criteria and providing the evidence. *Blood* 124, 3043-3051, doi:10.1182/blood-2014-07-552059 (2014).
- 14 Paiva, B., van Dongen, J. J. M. & Orfao, A. New criteria for response assessment: role of minimal residual disease in multiple myeloma. *Blood* 125, 3059-3068, doi:10.1182/blood-2014-11-568907 (2015).
- 15 Landgren, O., Devlin, S., Boulad, M. & Mailankody, S. Role of MRD status in relation to clinical outcomes in newly diagnosed multiple myeloma patients: a meta-analysis. *Bone marrow transplantation* 51, 1565-1568, doi:10.1038/bmt.2016.222 (2016).

- 16 Schinke, C. *et al.* The prognostic value of the depth of response in multiple myeloma depends on the time of assessment, risk status and molecular subtype. *Haematologica* 102, e313-e316, doi:10.3324/haematol.2017.165217 (2017).
- 17 Innao, V. *et al.* Standardisation of minimal residual disease in multiple myeloma. *European Journal of Cancer Care* 26, e12732, doi:doi:10.1111/ecc.12732 (2017).
- 18 van Velzen, J. F., van den Blink, D. & Bloem, A. C. Inability of a monoclonal anti-light chain antibody to detect clonal plasma cells in a patient with multiple myeloma by multicolor flow cytometry. *Cytometry Part B: Clinical Cytometry* 84B, 30-32, doi:doi:10.1002/cyto.b.21044 (2013).
- 19 Krejci, J. *et al.* Monocytes and granulocytes reduce CD38 expression levels on myeloma cells in patients treated with daratumumab. *Clinical cancer research : an official journal of the American Association for Cancer Research* 23, 7498-7511, doi:10.1158/1078-0432.CCR-17-2027 (2017).
- 20 Davis, J. & Goadrich, M. in *Proceedings of the 23rd international conference on Machine learning* 233-240 (ACM, Pittsburgh, Pennsylvania, USA, 2006).
- 21 Paiva, B. *et al.* Minimal residual disease monitoring and immune profiling in multiple myeloma in elderly patients. *Blood* 127, 3165-3174, doi:10.1182/blood-2016-03-705319 (2016).
- 22 Arro, M. *et al.* Consensus guidelines on plasma cell myeloma minimal residual disease analysis and reporting. *Cytometry Part B: Clinical Cytometry* 90, 31-39, doi:doi:10.1002/cyto.b.21228 (2016).
- 23 Spears, M. D., Olteanu, H., Kroft, S. H. & Harrington, A. M. The immunophenotypic stability of plasma cell myeloma by flow cytometry. *International Journal of Laboratory Hematology* 33, 483-491, doi:10.1111/j.1751-553X.2011.01317.x (2011).
- 24 Matsui, W. *et al.* Characterization of clonogenic multiple myeloma cells. *Blood* 103, 2332-2336, doi:10.1182/blood-2003-09-3064 (2004).
- 25 Paino, T. *et al.* Phenotypic, Genomic and Functional Characterization Reveals No Differences between CD138(++) and CD138(low) Subpopulations in Multiple Myeloma Cell Lines. *PLoS ONE* 9, e92378, doi:10.1371/journal.pone.0092378 (2014).
- 26 Kukreja, A. *et al.* Enhancement of clonogenicity of human multiple myeloma by dendritic cells. *The Journal of Experimental Medicine* 203, 1859-1865, doi:10.1084/jem.20052136 (2006).
- 27 Fuhler, G. M. *et al.* Bone marrow stromal cell interaction reduces Syndecan-1 expression and induces kinomic changes in myeloma cells. *Experimental Cell Research* 316, 1816-1828, doi:https://doi.org/10.1016/j.yexcr.2010.03.013 (2010).
- 28 Ikoma, M. R. V. *et al.* Proposal for the standardization of flow cytometry protocols to detect minimal residual disease in acute lymphoblastic leukemia. *Revista Brasileira de Hematologia e Hemoterapia* 37, 406-413, doi:10.1016/j.bjhh.2015.07.012 (2015).
- 29 Schuurhuis, G. J. *et al.* Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* 131, 1275-1291, doi:10.1182/blood-2017-09-801498 (2018).
- 30 Zeijlemaker, W., Gratama, J. W. & Schuurhuis, G. J. Tumor heterogeneity makes AML a "moving target" for detection of residual disease. *Cytometry Part B: Clinical Cytometry* 86, 3-14, doi:doi:10.1002/cyto.b.21134 (2014).

- 31 Chen, X. & Wood, B. L. Monitoring minimal residual disease in acute leukemia: Technical challenges and interpretive complexities. *Blood Reviews* 31, 63-75, doi:<https://doi.org/10.1016/j.blre.2016.09.006> (2017).
- 32 Baer, M. R. *et al.* High frequency of immunophenotype changes in acute myeloid leukemia at relapse: implications for residual disease detection (Cancer and Leukemia Group B Study 8361). *Blood* 97, 3574-3580, doi:10.1182/blood.V97.11.3574 (2001).
- 33 Langebrake, C. *et al.* Immunophenotypic differences between diagnosis and relapse in childhood AML: Implications for MRD monitoring. *Cytometry Part B: Clinical Cytometry* 63B, 1-9, doi:doi:10.1002/cyto.b.20037 (2005).

## Supplementary Material

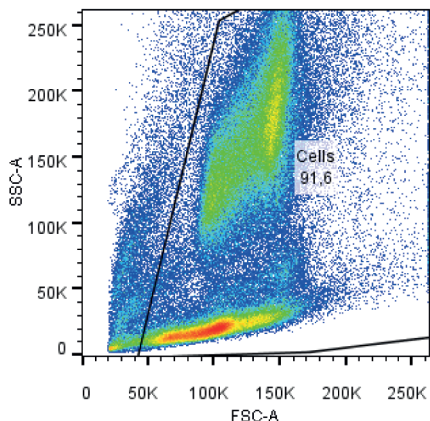
### Contents

Supplementary material I – Methods	253
Excluding debris by manual gating	253
Manual Gating strategy	254
Transformation	255
Fluorescence Minus One centering	256
ECLIPSE Control Model and removal of control cells	258
Supplementary Material II – ECLIPSE results of a synthetic dataset	259
ECLIPSE analysis of the synthetic dataset with decreasing numbers of clonal plasma cells	259
Supplementary material III – Control Model Results	262
The ECLIPSE Control Model describes healthy bone marrow cells	262
Healthy bone marrow cells are eliminated by ECLIPSE	265
Supplementary Material IV – ECLIPSE model Results	269
ECLIPSE model	271
Supplementary Material V – HSNE results	277
Hierarchical Stochastic Neighbour Embedding (HSNE) comparison	277
References	279



## Supplementary material I – Methods

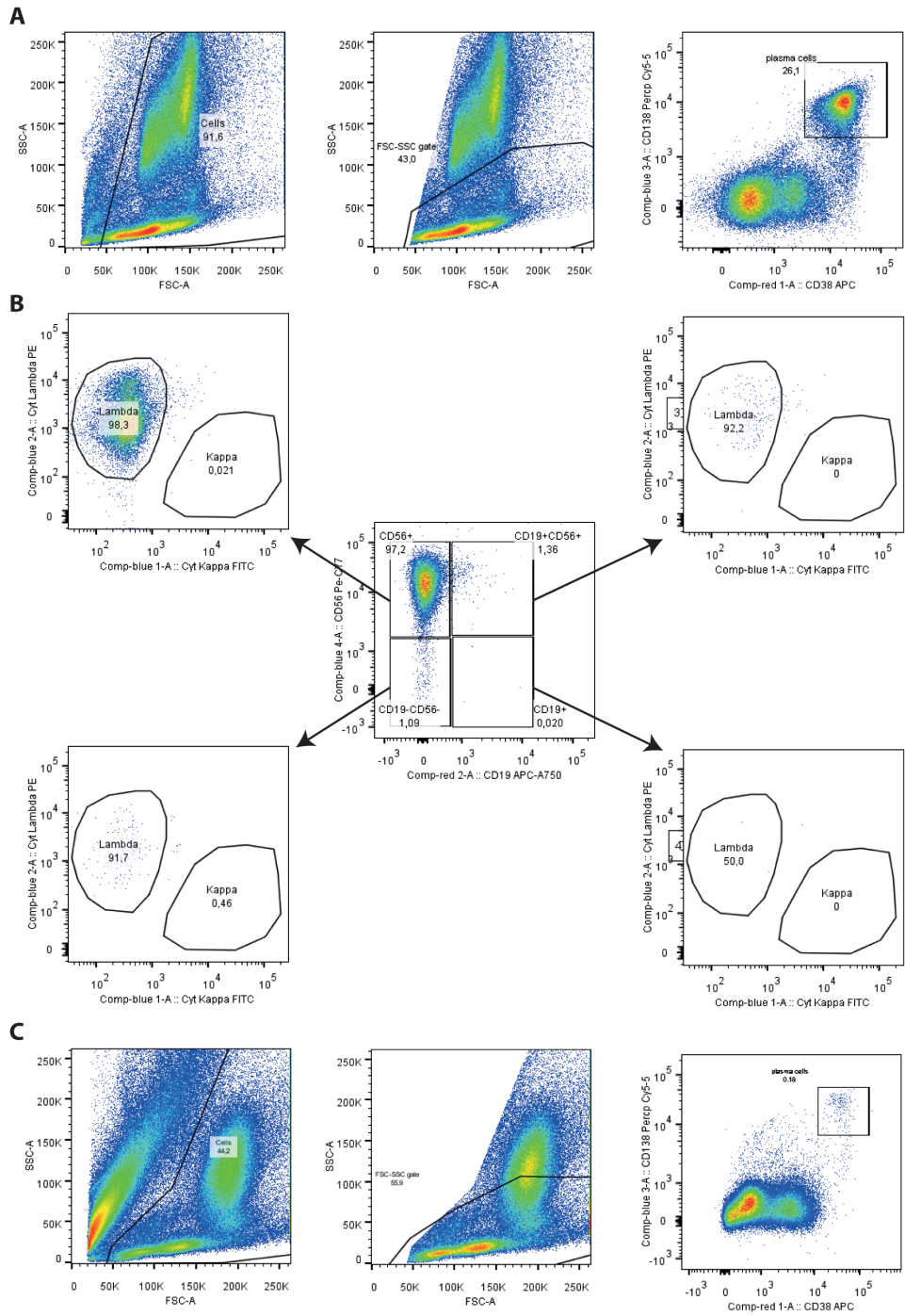
### Excluding debris by manual gating



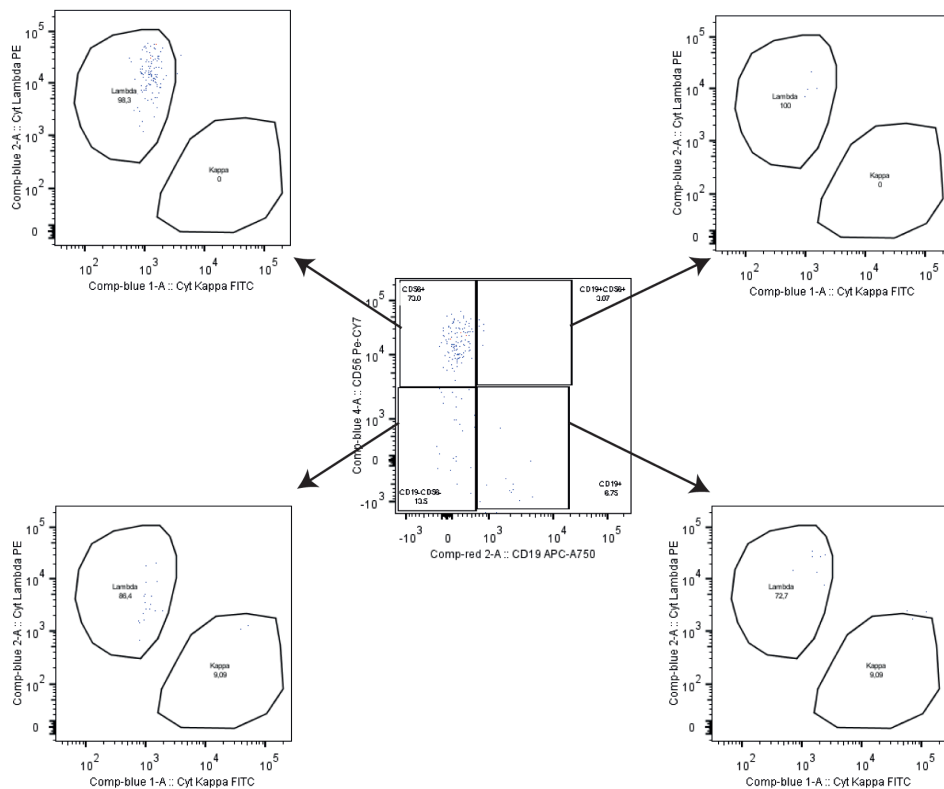
**Figure S1. Excluding debris by manual gating.**

This is a representative plot of the manual gating of cells based on FSC/SSC, excluding debris. All fcs files were gated, and the events within this gate were exported as new fcs files and imported for subsequent ECLIPSE analysis.

Manual Gating strategy



D



**Figure S2. Manual gating of clonal plasma cells.**

All fcs files were examined by manual gating. This Figure is a representative manual gating result of ID1 diagnosis sample (A, B) and the follow-up samples of the same patient (C, D). A, C) First debris was excluded based on low FSC (left plot). Subsequently cells with a low SSC were gated (middle plot). In the right graph plasma cells were gated based on their double positivity for CD38 and CD138 expression. B, D) middle panel: CD19 and CD56 expression of the plasma cells was examined in a biplot, gating CD19-CD56+, CD19+CD56+, CD19-CD56- and CD19+CD56- separately. Cytg $\lambda$  monoclonality was confirmed for CD19-CD56+, CD19+CD56+ and CD19-CD56- plasma cells, in both the diagnosis and the follow-up sample.

## Transformation

We transformed the raw MFC data by using hyperbolic inverse sine (arcsinh) with cofactor 150<sup>3</sup>. When building the Control Model, the arcsinh-transformed data were mean centered and scaled based on variability within the control group. This pre-processing selectively emphasizes deviations from the 'normal' cell variability observed in the responder individuals.<sup>2</sup>

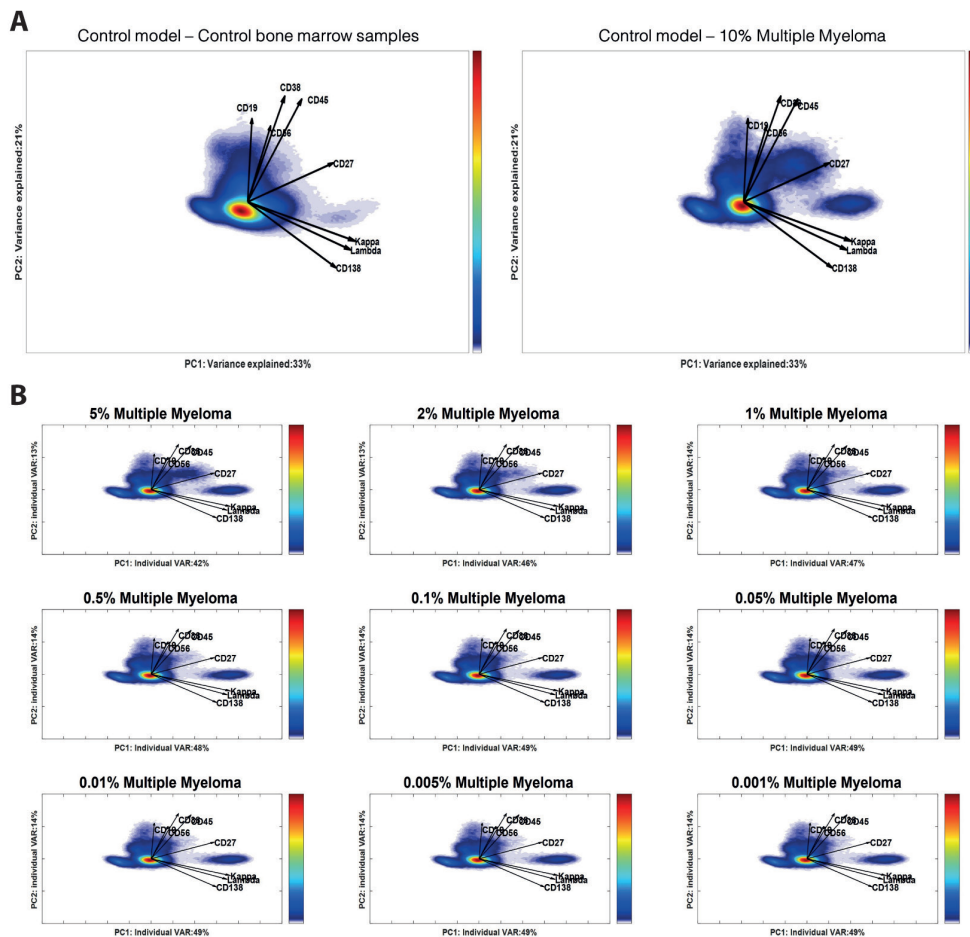
### Fluorescence Minus One centering

To improve the interpretability of the diagnosis-specific ECLIPSE plot, we developed a different and dedicated type of centering and scaling, based on Fluorescence Minus One (FMO)<sup>4</sup> samples. As the name suggests, an FMO control sample<sup>4</sup> is obtained by staining the cells with all the experimental fluorescent conjugates of the MFC panel except one. When setting up a new MFC panel, FMO control samples are commonly measured for each marker used. The goal is to assess the contribution of spillover signal from all the other markers into the channel of the missing fluorescent conjugate. With an accurate compensation applied, no high intensities should be detected in the specific channel of the FMO control. Consequently, in the fully stained sample the 'real' positive population may be accurately identified. In other words, the FMO is the true negative boundary to distinguish populations positive or negative for a certain marker. Here we propose an alternative purpose for the FMO control samples: centering and scaling the data based on the FMO controls when building the diagnosis-specific ECLIPSE model.

The type of centering (and scaling) influences the orientation of the loadings. When the model is built on the data pre-processed based on the FMOs intensities, the orientation of the resulting loadings allow a better characterization of the phenotype of the clonal plasma cells, compared to the model built on data mean-centered based on the diagnosis sample. The effect of the different pre-processing on the loadings is shown in the Figure S5a-b. For each FMO, the mean fluorescence in the channel of the missing fluorescent conjugate was calculated and then subtracted from the corresponding variable in the samples used for the ECLIPSE model. The resulting FMOs-centered matrix is then divided by the standard deviation calculated on the FMO samples. In this way, the ECLIPSE model built on the FMOs-pre-processed data will focus on describing the deviation of single cells from the true negative expression of the markers, corresponding to the origin of the loadings.

SCA-based dimensionality reduction techniques are generally applied after centering (or autoscaling) the data matrix.<sup>5</sup> Centering flow cytometry data consists of subtracting the average marker expression from each variable. In a plot of an SCA model, this geometrically translates in a shift of the loadings axis origin to the mean (center of mass) of the flow cytometry data without altering its structure. The first principal component of the built model then corresponds to the direction with the largest variation, which, after mean centering, is represented by the deviation from the center of mass. When multiple cellular populations are present, the center of mass coincides with the center of the most abundant population. This turned out to be suboptimal when building the diagnosis-specific ECLIPSE model, aimed to detect and describe the clonal plasma cells. When the majority of abnormal cells identified by the Control model are clonal plasma cells, these

cells are the center of mass and, after centering, the related scores will be placed in the center of the ECLIPSE model. This means that no relevant information about the marker variability of clonal plasma cells can be extracted from the model loadings, as shown in Figure S3a. When using FMO centering the resulting model focuses on describing the deviation of single cells from the true negative expression of the markers (Figure S3b).



**Figure S3. ECLIPSE model built on the cells of patient ID1 at diagnosis.**

The ECLIPSE model was built after elimination of normal cells by the ECLIPSE algorithm. The major cell population in the plots corresponds to the multiple myeloma cells. A) the model is built on the data pre-processed with mean and standard deviation calculated on the abnormal cells. The origin of the model loadings moves to the data center of mass which coincides with the multiple myeloma clone. As result, the direction of the loadings indicates an average expression of all the markers by the monoclonal plasma cells; B) the model was built on the data pre-processed with mean and standard deviation calculated on the FMO control samples. The model loadings origin moves to the real 'negative' expression for all the markers. By looking at the new direction of the loadings we can observe CD38+CD56+CD138+CytIgλ+ typical expression for the multiple myeloma cells.

### **ECLIPSE Control Model and removal of control cells**

A control model describes the cellular marker variability within the control (reference) group. SCA explicitly models similarities among all control individuals, of which cellular marker (co)-expressions are described by a set of common loadings.

In the case of the MM study, the diagnosis and follow-up individuals were projected onto the Control Space. The distance of each cell from the fitted Control space may be calculated as sum of squared residuals or Squared Prediction Error (SPE). SPE values of the projection are evaluated against the control SPE confidence limit. Cells in the diagnosed individuals of which the SPE exceed the control limit are considered 'abnormal' in their cellular marker co-expression, as it does not overlap with the variability modelled by the Control Model. These cells are more likely related to the haematological malignancy, while cells with lower SPE values are more similar to the control bone marrow cells. These latter cells are eliminated, because they are considered irrelevant. This elimination improves focus on (rare) aberrant cells, which are subsequently characterized in the ECLIPSE analysis.

The number of components needed to model the multi-cellular variability is estimated by the cross-validated Predicted Residual Error Sum of Squares (PRESS) value<sup>1</sup>, calculated by leaving-one control individual-out each iteration. Control-specific variability distinctive for specific control individuals, should remain unexplained by the model. This 'Residual' variability can be found in the orthogonal contribution to the Control space. The distance of each cell from the fitted Control space may be calculated as sum of squared residuals or Squared Prediction Error (SPE). This value serves as measure of how well each cell conforms to the model. When SPE values increase, the cell variability is less well described by the Control Model and thus the cell is considered 'abnormal'. A confidence interval (95%) can be established for the SPE statistics of control/healthy cells. Cells which exceed this control-specific limit show an 'abnormal' marker variability and thus are worth to investigate further. For mathematical details, we refer to our previous article.<sup>2</sup>

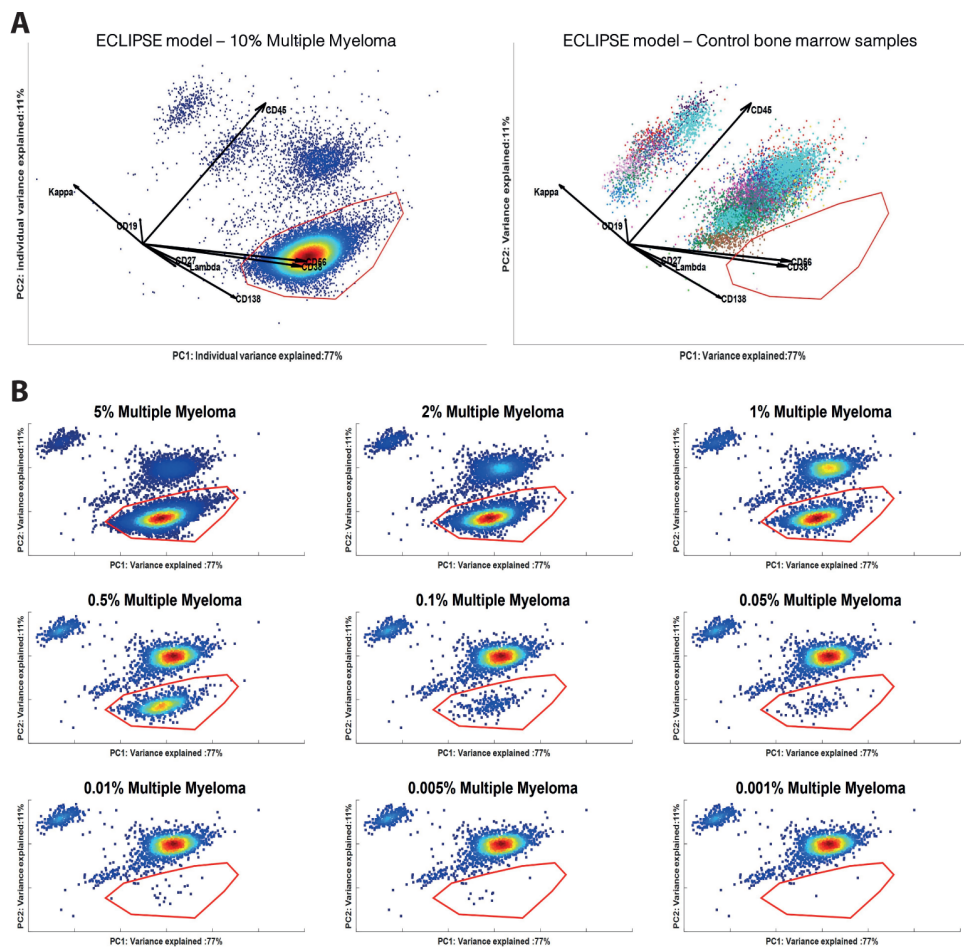
## Supplementary Material II – ECLIPSE results of a synthetic dataset

### ECLIPSE analysis of the synthetic dataset with decreasing numbers of clonal plasma cells

To test the sensitivity of the ECLIPSE method for the detection of MRD cells a synthetic combination was used. We virtually spiked a control bone marrow sample with gated multiple myeloma cells from a patient sample measured at diagnosis. In this way we created 10 synthetic samples with various quantity of spiked clonal plasma cells. Each sample comprised a total of 200.000 events with a percentage of spiked clonal plasma cell populations ranging from 10% to 0.001%. The control bone marrow sample used in the synthetic combination was excluded from the control reference group, leaving 12 control bone marrow samples to build the control model (Figure S3a). First, the spiked samples were projected in the resulting control space (Figure S3a-3b).

From the control model it was difficult to visually identify cells phenotypically distinct from the normal bone marrow cells. Also differences in percentage of clonal plasma cells among the spiked samples could not be easily appreciated because the normal cells overlap the clonal plasma cells. Automated removal of normal bone marrow cells by ECLIPSE resulted in a model focussed on the clonal plasma cells. The sample with 10% of clonal plasma cells was used as diagnosis sample, to create the ECLIPSE space, while the samples with lower percentages of spiked plasma cells were projected into the ECLIPSE space, like previously was done for the follow-up samples. The clonal plasma cells can be discerned from other 'abnormal' cells by looking at the orientation of the loadings and based on minimum overlay with donor-specific cells left in the control bone marrow samples, projected in the multiple myeloma-specific model (Figure S4A-B). The monoclonal plasma cells present in the 10% spiked sample are easily detected as shown in Figure S4A (left panel). The identified clone was gated and the same gate can be superimposed in the plots of the projected donor-specific cells of the control bone marrow samples (Figure S4A-right panel) and the projected cells from the follow-up sample (Figure S4B). A low number of cells from 3 healthy bone marrow samples was found in the multiple myeloma-specific gate. The cells are mainly at the border of the gate and may be identified as non-plasma cells due to their lower expression of CD38 and CD138. For the follow-up samples, monoclonal cells are found in the multiple myeloma gate. As the percentage of spiked clonal cells diminishes in subsequent samples, less cells were found inside the gate. Even in the samples with the lowest amounts of monoclonal plasma cells, the correct amount of cells was found in the ECLIPSE gate (20 cells in the 0.01% sample; 10 cells in the 0.005% sample and 2 cells for the 0.001% sample). From this we can conclude that using ECLIPSE MRD populations of as low as 0.001% of total cells can be identified. The synthetic samples had a maximum of 200.000 cells, with larger sample sizes the limit could be even lower.

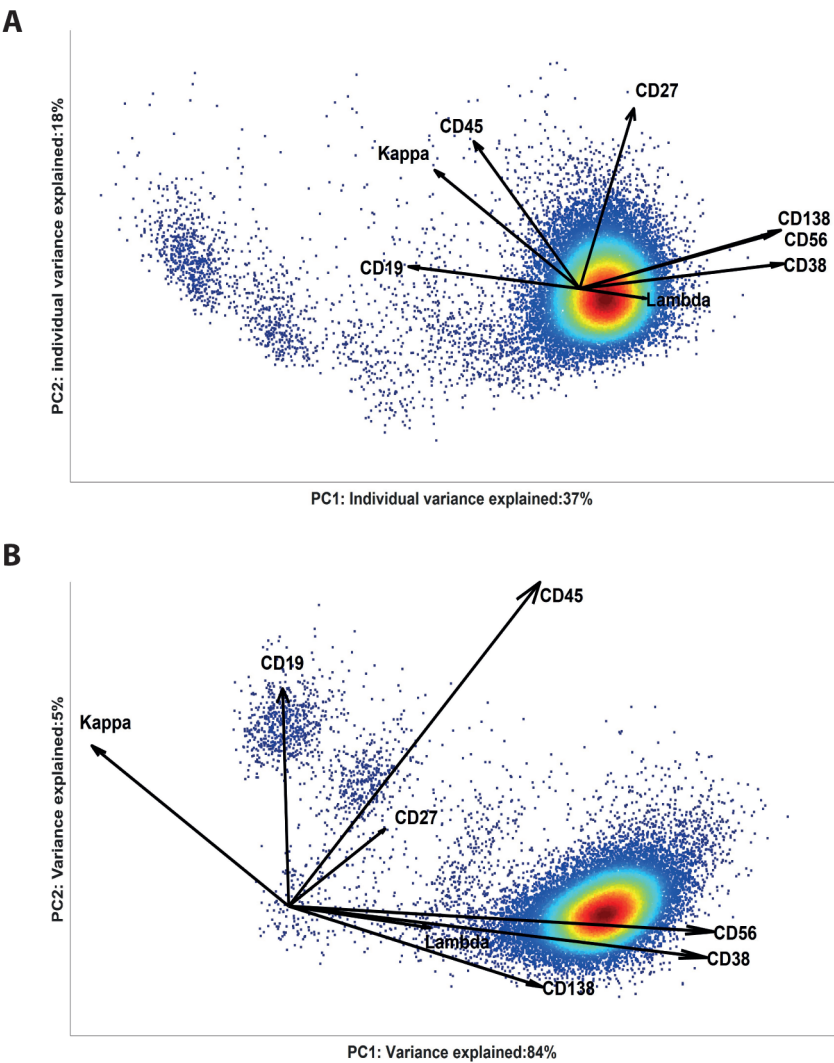




**Figure S4.**

A) Control model built on 12 healthy bone marrow samples, using the first two principal components. The loadings of the Control Model are plotted as vectors and indicate the contribution of each marker to the cell-to-cell variability within the control samples. Left: Density plot of cells of the healthy bone marrow samples. Right: Density plot of the cells from the sample with 10% of clonal plasma cells, used as diagnosis. B) Density plot of the spiked samples projected in the Control Model. Amount of spiked clonal plasma cells diminish going from left to right and from top to bottom.





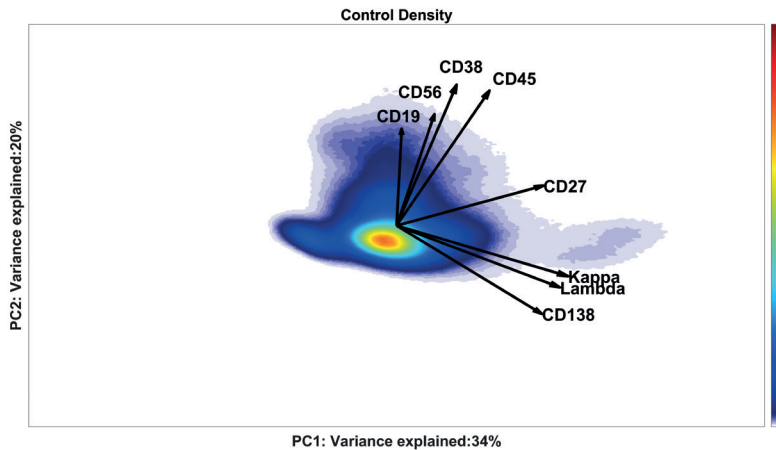
**Figure S5.**

A) ECLIPSE model built on the abnormal cells of the diagnosis sample, having 10% of spiked clonal plasma cells. FMOs pre-processing was applied. Left: Cells from the diagnosis sample in the ECLIPSE diagnosis-specific sample. The cell aggregate associated to the clonal plasma cells is gated. Right: control donor-specific cells are projected in the ECLIPSE diagnosis-specific cells. Few cells are found in the clonal plasma cells gate. B) Spiked samples projected in the ECLIPSE diagnosis-specific model, after removal of normal cells. Multiple myeloma gate is applied to identify the spiked monoclonal cells at different percentages, ranging from 5% to 0.001%.

## Supplementary material III – Control Model Results

### The ECLIPSE Control Model describes healthy bone marrow cells

The ECLIPSE method builds a Control Model, which describes the single-cells marker variability of all the ‘healthy’ bone marrow (HBM) individuals.



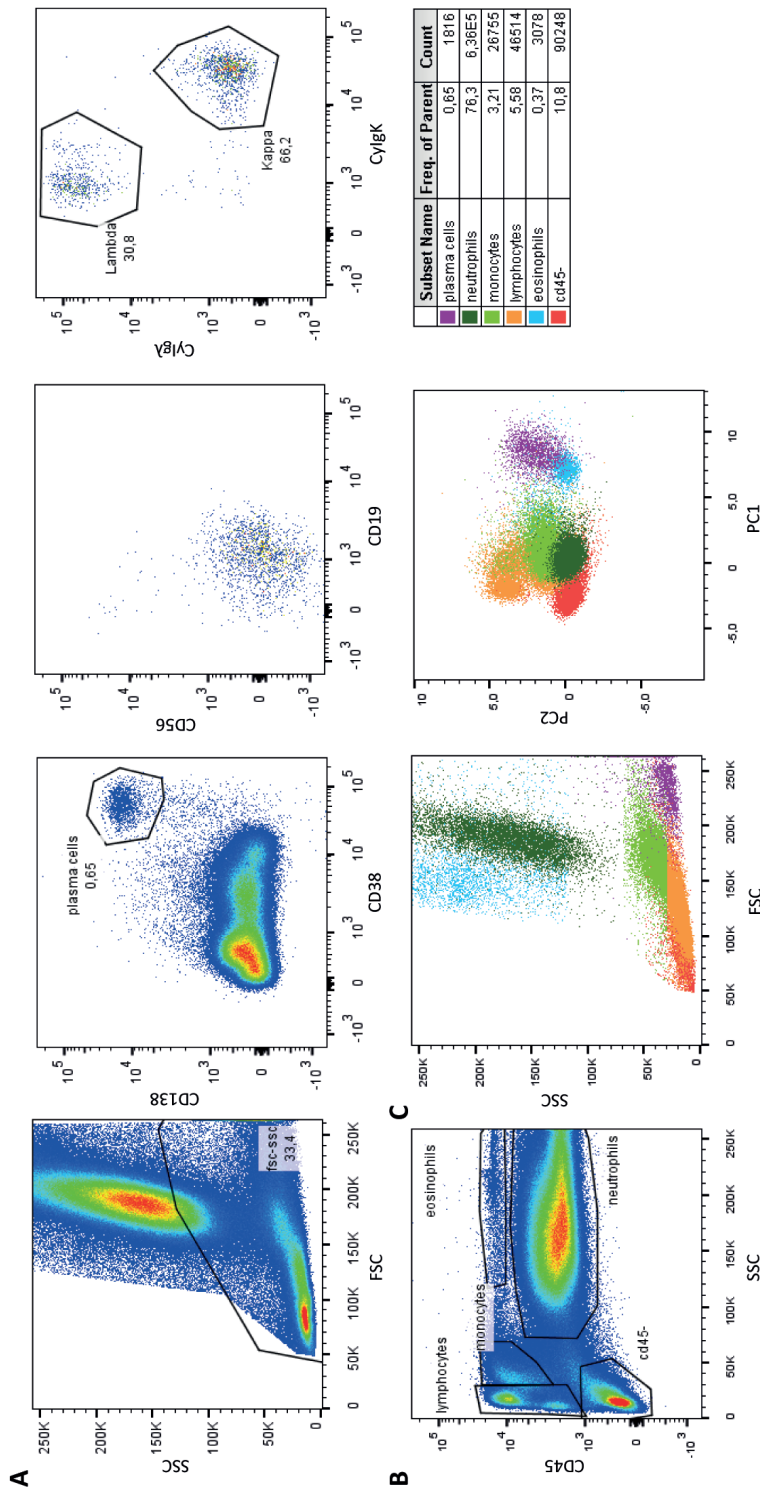
**Figure S6.**

Density plot of cells of the 13 healthy bone marrow samples in the Control Model, built on the first two principal components. The loadings of the Control Model are plotted as vectors and indicate the contribution of each marker to the cell-to-cell variability within the 13 samples.

Figure S6 shows the distribution of cells of the control individuals in the Control Space in a 2-dimensional graph. The loadings (vectors with marker names), superimposed on the Control space, indicate how cellular marker co-expressions differ among the various cell subsets. Cells positioned in a region to which a loading points, have a relatively high expression level for this marker. When cells are located in the opposite direction of the loading, they have a relatively low expression level for this marker.

The markers in this panel are selected to identify plasma cells within the bone marrow compartment and in order to characterize the phenotype of the malignant plasma cells<sup>6,7</sup>. The Control Model was built on the HBM samples, which only contain a maximum of 5% plasma cells.<sup>8</sup> The marker loadings are greatly influenced by the other cell types present in the HBM samples. Various types of HBM cells express CD38,<sup>9</sup> while CD138 was only found expressed by eosinophils (data not shown) and plasma cells (Figure S7). This explains the orthogonal direction of the two marker loadings. Only plasma cells are

CD38+CD138+<sup>10,11,12</sup>. The most abundant cell populations do not express intracellular Kappa (Cylgκ), nor Lambda (Cylgλ), therefore these loadings point into a similar direction. Generally a mix of different CD19 and CD56 plasma cell phenotypes are found in HBM<sup>13</sup>. These markers, together with CD45 are expressed by multiple cell populations in the bone marrow and these cell populations influence the loadings orientation. Since plasma cells are scarce in HBM, they might not be visible in a density representation. By gating the healthy plasma cells present in the HBM samples, we confirmed that they are grouped together on the right side of the control space (Figure S7).

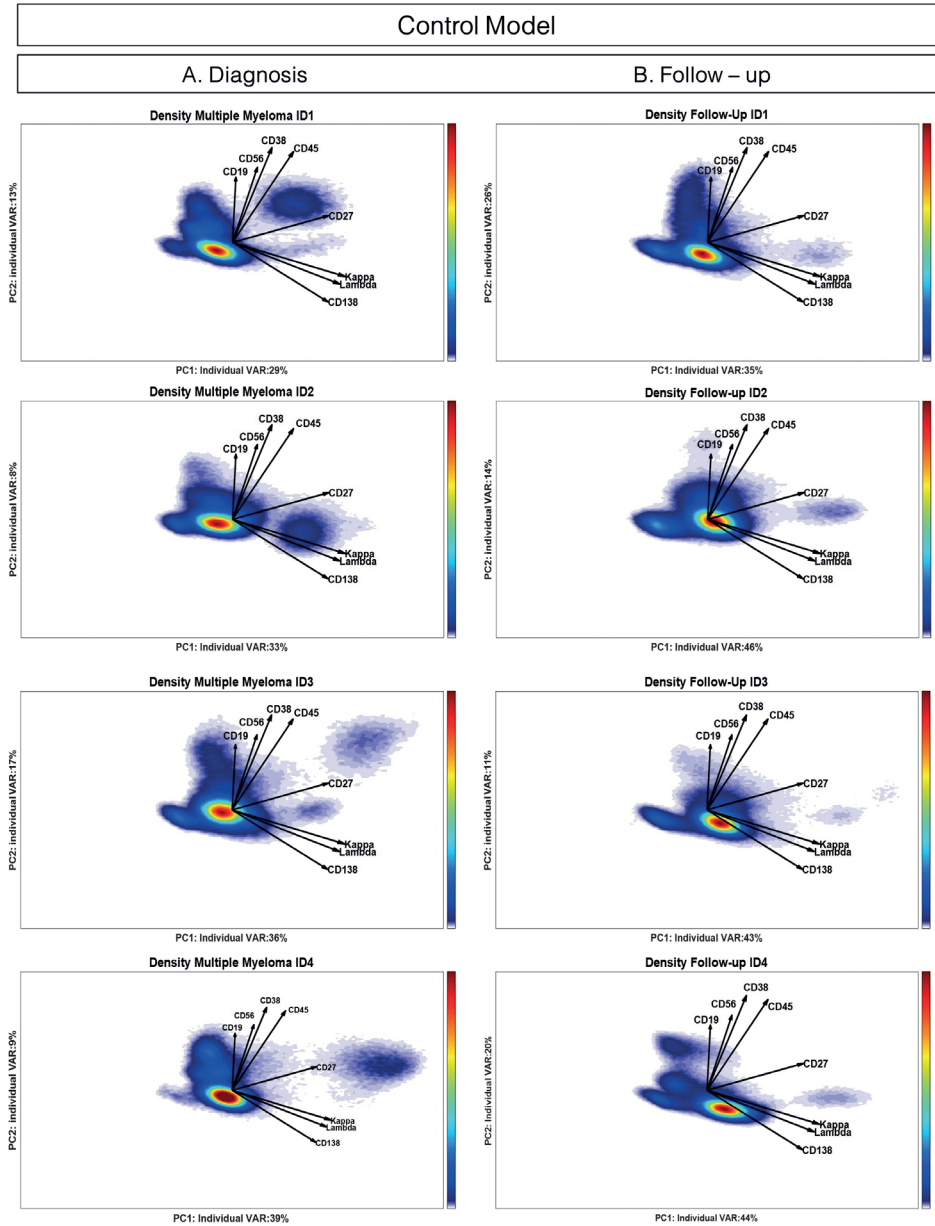


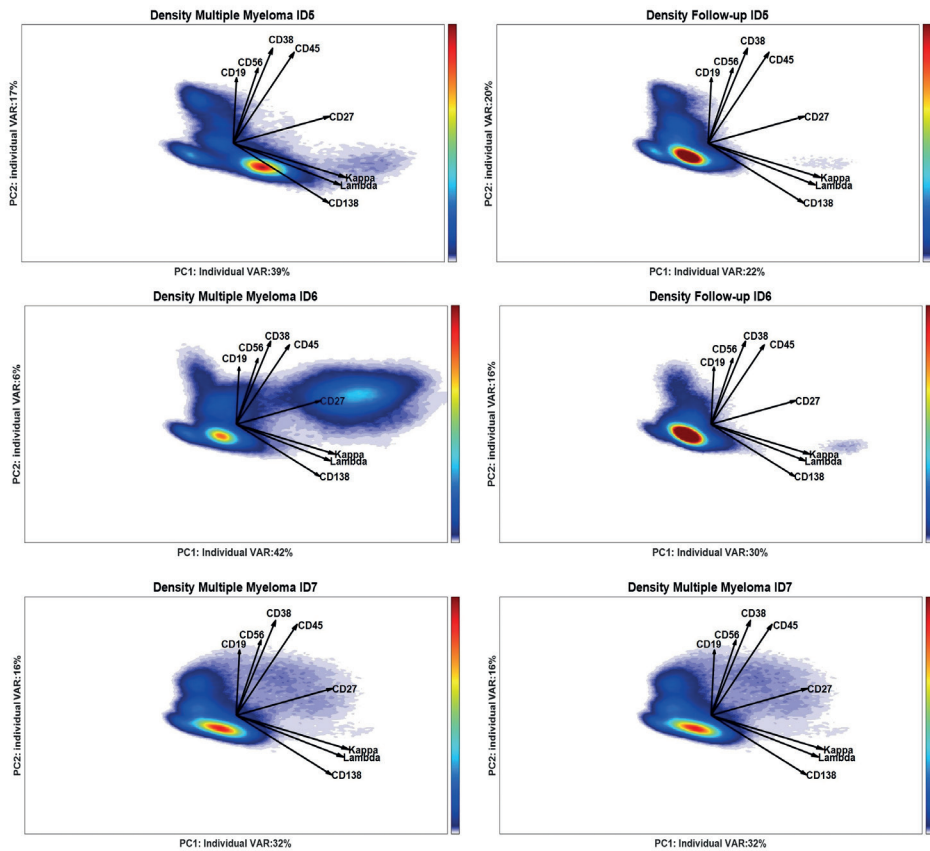
**Figure S7 . Backgating results of a representative control sample in the Control Model.**

To identify the different cell populations present in the control space, per control sample the control PC scores and raw intensities for FSC, SSC and the markers were exported as fcs file. The graphs shown are derived from a representative control sample. A) In flowjo plasma cells were gated using FSC/SSC, CD38+CD138+. Different plasma cell subsets were identified based on CD19/CD56 (co-)expression. Per plasma cell subset the ratio of Kappa/Lambda expressing cells was evaluated. B) Various WBC populations were gated based on SSC and CD45 expression. C) Overlay of gated plasma cells and WBC populations in a FSC/SSC plot (left) and in the Control space (PC1/PC2, right plot).

Healthy bone marrow cells are eliminated by ECLIPSE

Cells from diagnosis and follow-up samples are projected in the Control space. Figure S8 shows the projection of the diagnosis samples (Figure S8A) and the corresponding follow-up samples (Figure S8B). Considerable overlap with the control bone marrow cells can be observed when comparing the position of the cells of the patient samples in Figure S8 to the position of the control bone marrow cells in Figure S6.





**Figure S8.**

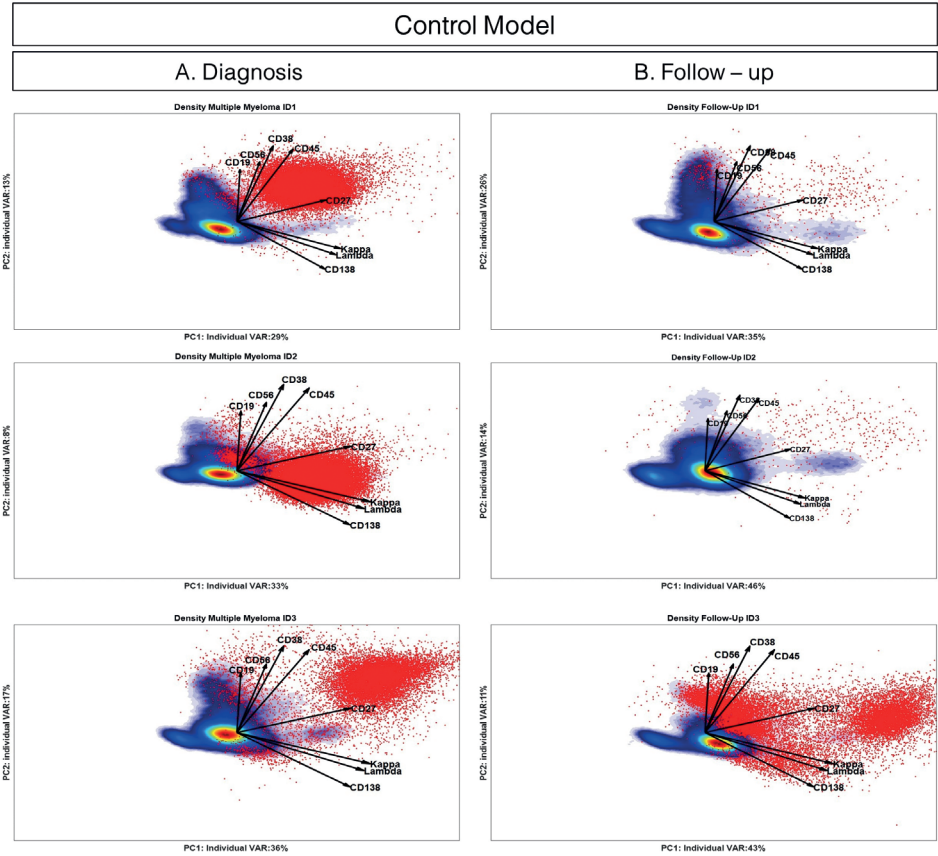
Density plot of diagnosis (A) and the corresponding follow-up (B) samples projected in the Control Model. The loadings specifically describe the contribution of each marker to the cell-to-cell variability typical of the control bone marrow samples.

In all 7 diagnosis samples a relatively small separate cell population is present (Figure S8A), located where CD45, CD27, CylgK, Cylgλ and CD138 loadings point. These cell populations are likely associated with the Multiple Myeloma malignancy. The location of the MM-associated cell population in the Control Model differs per patient. This suggests that the phenotypes of the clonal plasma cells vary between the different patients.

In the follow-up samples (Figure S8B) the overwhelming majority of control cells hampers the detection of MRD populations. The healthy-like cells and the malignant cells in both diagnosis and follow-up samples are identified by estimating the SPE values. Monoclonal plasma cells in diagnosis and follow-up samples are not well explained by the Control Model, as plasma cells in control samples are scarce, and often marker expressions of monoclonal plasma cells can deviate from the expressions of control plasma cells.

Therefore, these cells are expected to exceed the control SPE limit; conversely, healthy-like cells are characterized by lower values of the SPE statistics.

In Figure S7, the cells that exceed the SPE limit are depicted in red in 3 representative diagnosis and follow-up samples (ID1-3). In both diagnosis and follow-up samples (Panel 9A-B) the cells exceeding the SPE limit are located in the right side of the control plot. In all 3 follow-up samples the cells exceeding the SPE limit are partly overlapping with the region of the control bone marrow cells in Figure S4. Removing control bone marrow cells will enhance the focus on the abnormal cells in the patient samples, and the follow-up samples in particular. In samples ID1 and ID2 (Figure 9A-B), the low number of 'abnormal' cells makes the visualization of these cells challenging, while in sample ID3 various intensively populated red regions are present. This might suggest the presence of multiple abnormal cell populations. The position of the abnormal cell population in the Control Model differs per patient sample, which was expected as the phenotypes of the clonal plasma cells can differ per patient. Also the position of the MRD cells of one patient can slightly differ compared to the original MM clone, since treatment might induce changes in marker expressions.<sup>14</sup>

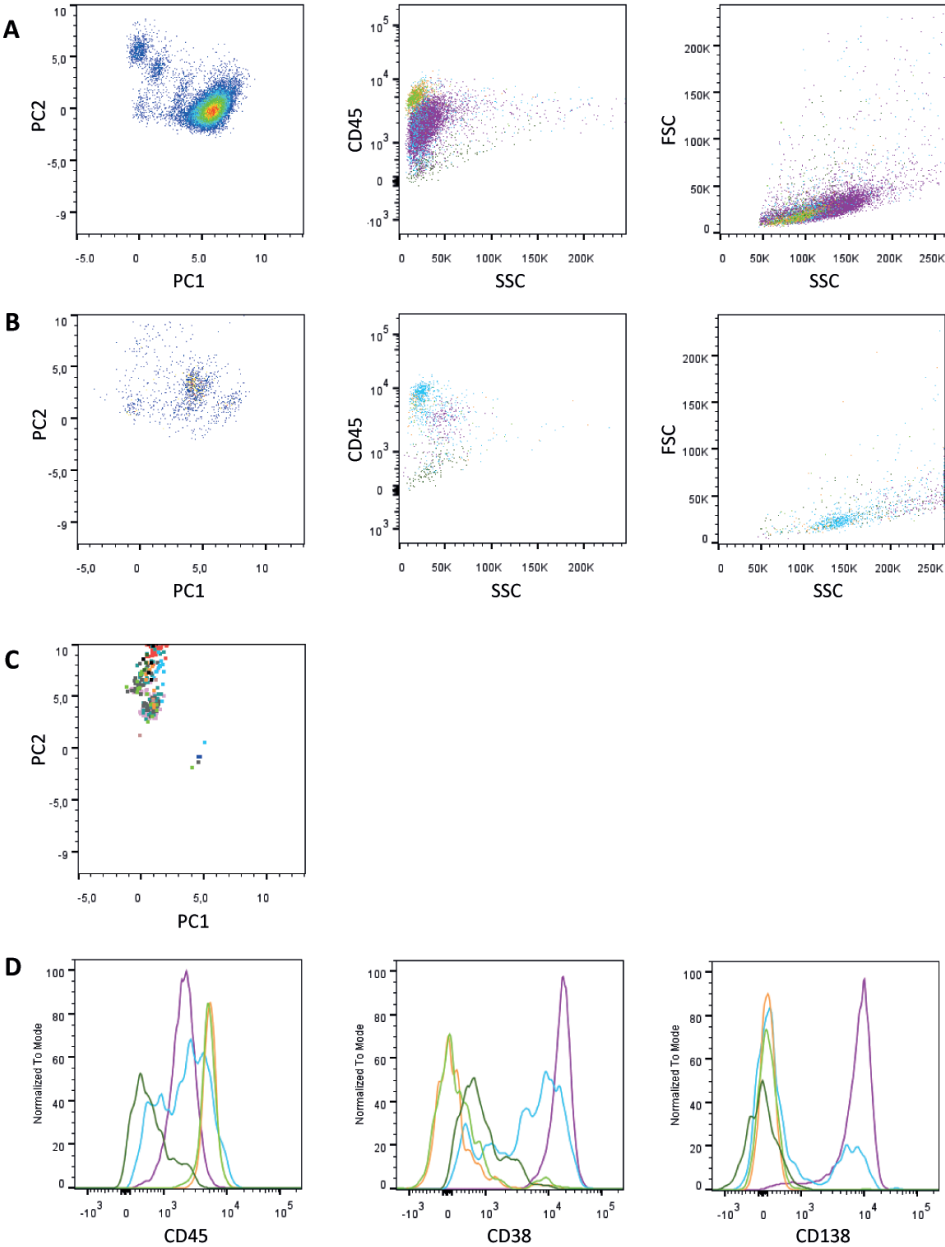


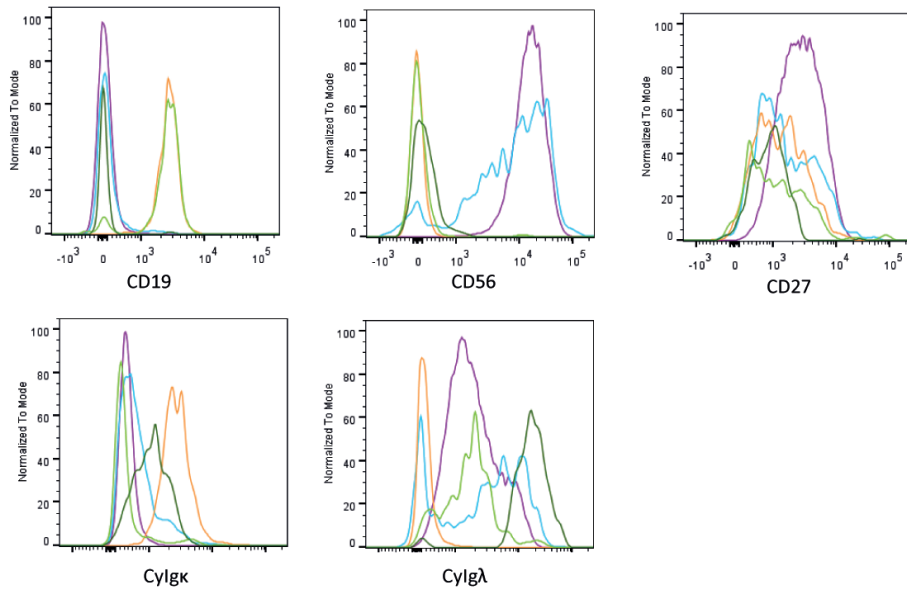
**Figure S9.**

Density plot of 3 diagnosis (Panel A) and the corresponding follow-up (Panel B) samples projected in the Control Model. The cells exceeding the SPE limits, and thus not described by the control bone marrow-specific model, are highlighted in red. Cells that can be identified as multiple myeloma clone in the diagnosis samples are present in slightly different region of the control space. This suggests that deviation in marker expression from the control cells in the malignant clone may vary between the patients.



Supplementary Material IV – ECLIPSE model Results





**Figure S10. Characterization of the cell populations in ECLIPSE space of ID1.**

For the ID1 Diagnosis and follow-up samples as well as the control samples the ECLIPSE PC scores and raw intensities for FSC, SSC and the markers were exported as fcs file. A+B) Left plot shows the gating of the monoclonal plasma cells (purple gate) and the extra 4 populations in the Diagnosis sample (A) and the Follow-up sample (B) of MM patient ID1. The middle plots show the expression of SSC and CD45 of these cell populations. The distribution of the cell populations in a FSC and SSC graph is shown in the right plot. C) Overlay of the donor-specific cells of the control samples in the ECLIPSE space, which remained after ECLIPSE removal. D) Histograms showing the expression of the single markers per gated cell population of the Diagnosis sample. The colors of the lines correspond to the colors of the gates in A.

ECLIPSE model

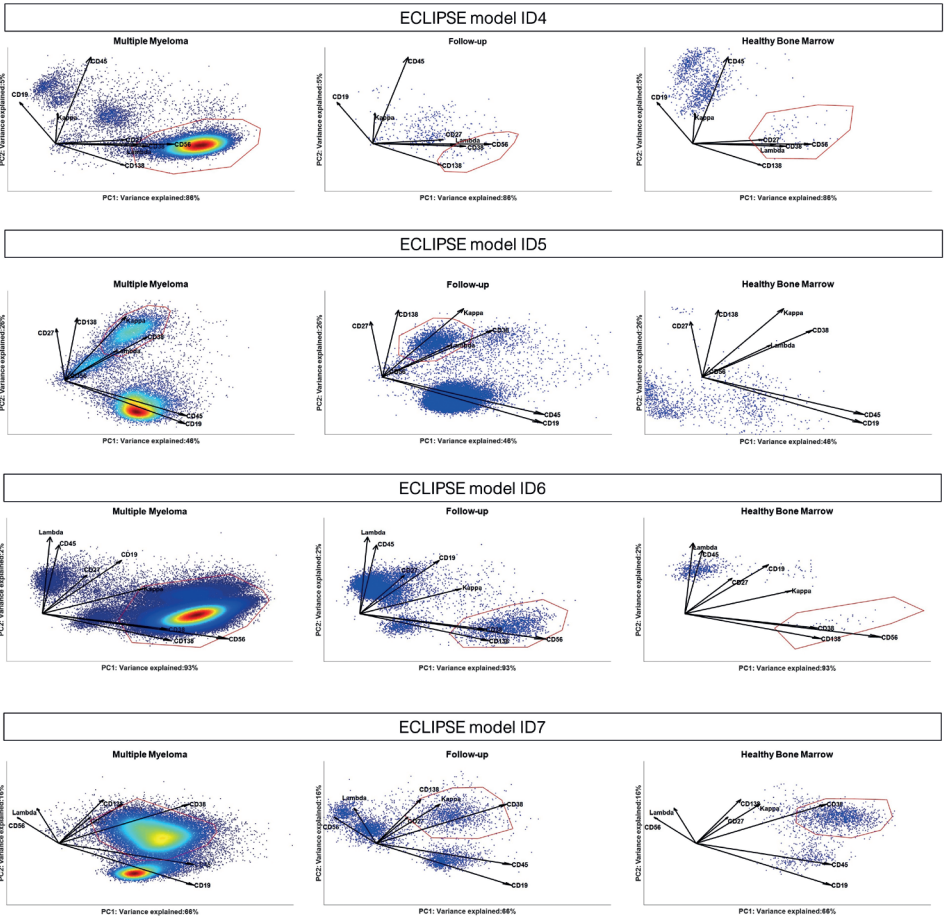
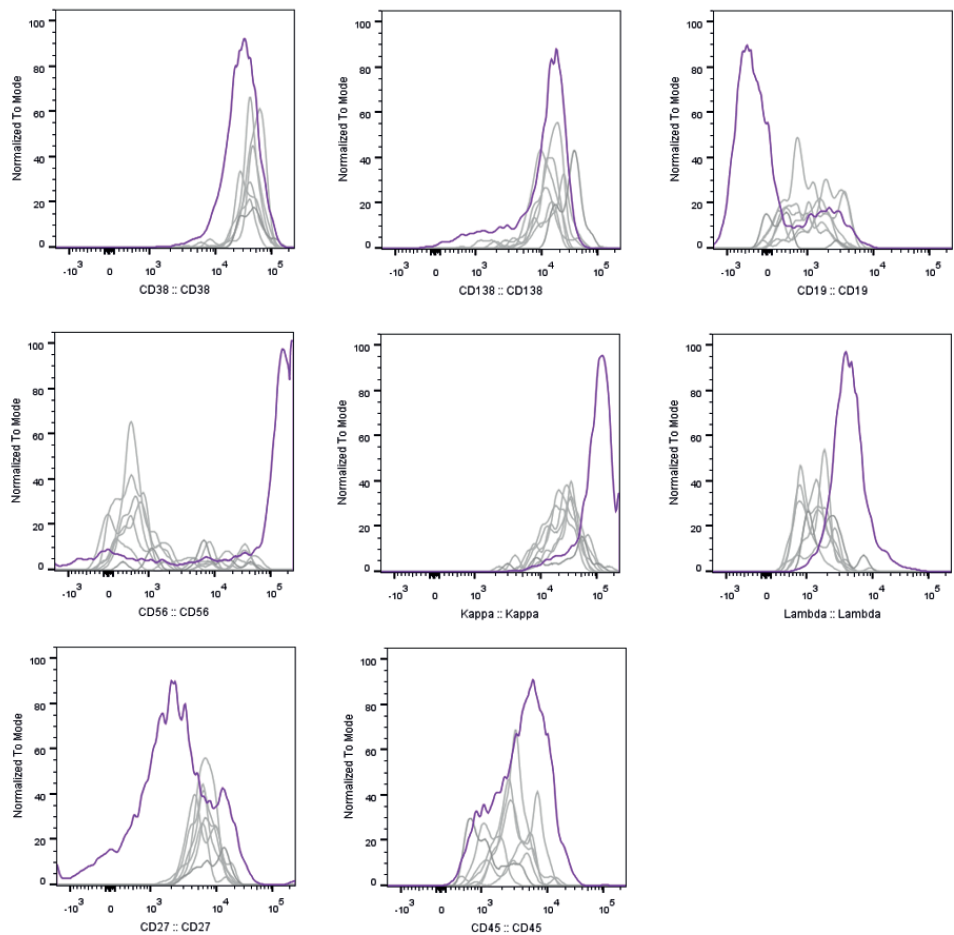


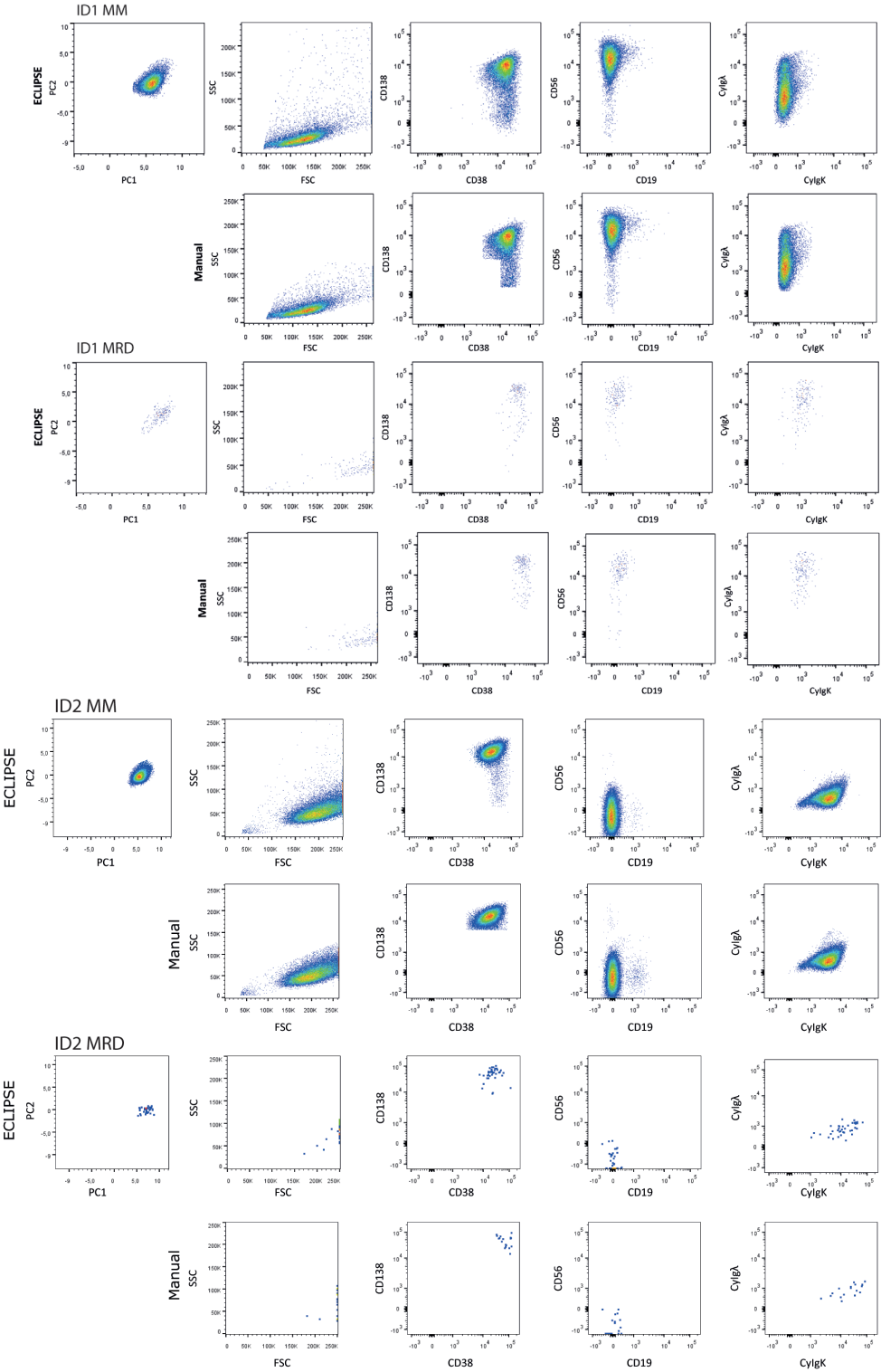
Figure S11.

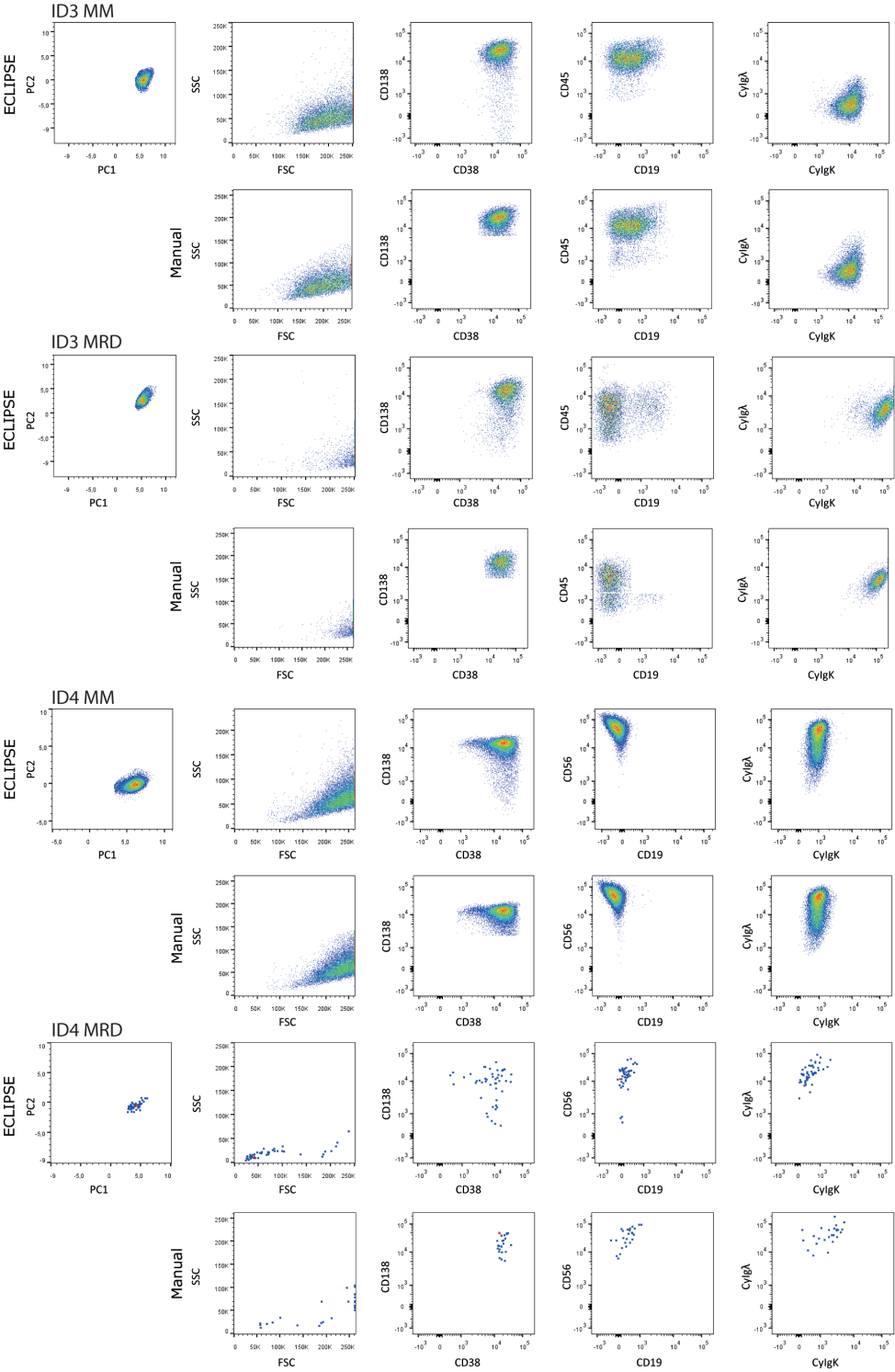
ECLIPSE models of the diagnosis samples ID4, 5, 6 and 7. The left graph shows the ECLIPSE model with the cells of the diagnosis sample, the middle graph shows the ECLIPSE model with the cells of the follow-up sample projected inside. The Right graph shows the donor-specific cells of all control bone marrow samples projected in the ECLIPSE model (the cells are colored per control bone marrow sample).

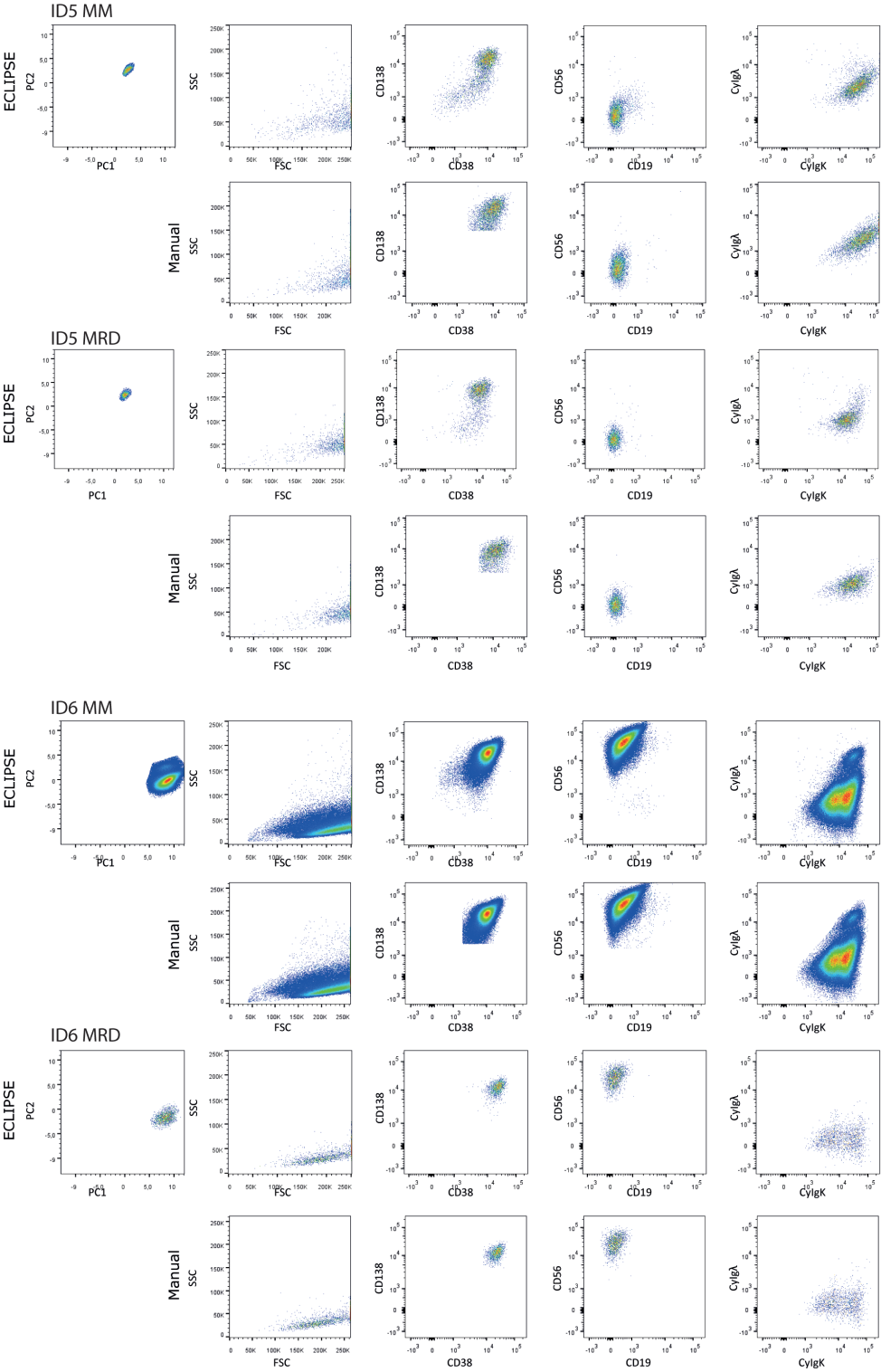


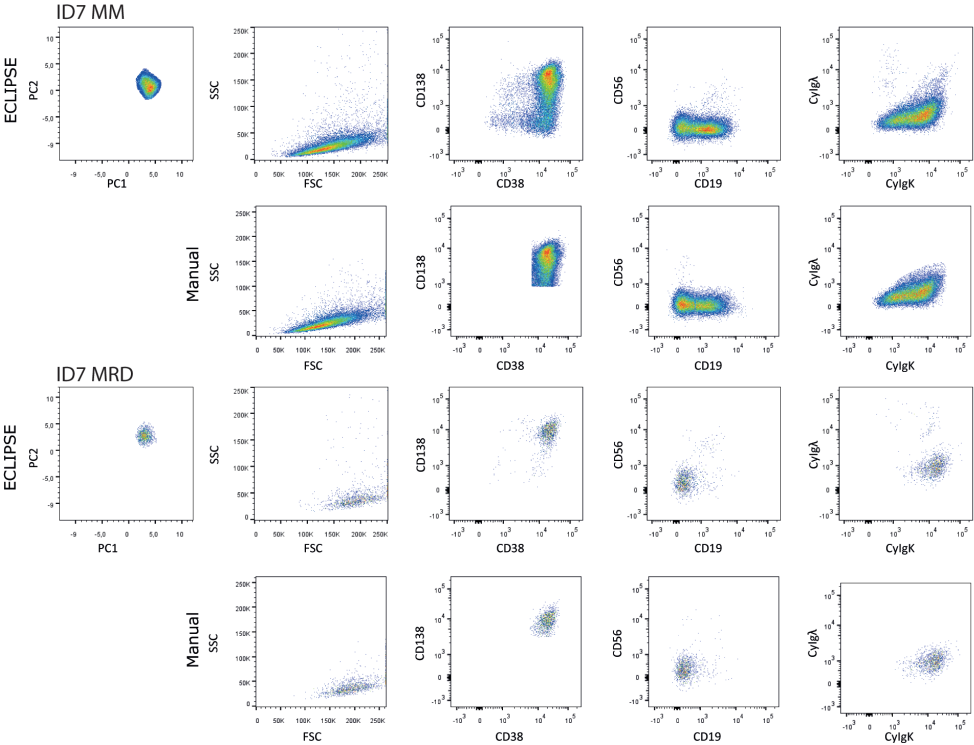
**Figure S12. HBM cells within the MRD gate of ID32 are CylgK+ plasma cells.**

Histogram overlay showing the expressions of the HBM cells within the ECLIPSE MRD gate per control sample (grey lines), and the expression of the cells within the MRD gate of the follow up sample (purple line) per marker.









**Figure S13. comparison of clonal plasma cell phenotypes found by ECLIPSE and by manual gating.**

The sample IDs are depicted in the titles of the panels. For each sample, the top panel shows the ECLIPSE gated phenotype, the bottom panel is the manual gated phenotype. The red gate in the CD19/CD45 plot of ID3 MRD marks cells found by ECLIPSE, which were not identified by manual gating



## Supplementary Material V – HSNE results

### Hierarchical Stochastic Neighbour Embedding (HSNE) comparison

One of the most commonly used multivariate analysis tools to visualize clusters of single cells in a biaxial space is viSNE<sup>15</sup>, which uses non-linear t-Stochastic Neighbour Embedding (t-SNE) algorithm for dimensionality reduction. Although the potential of the method to detect MRD cells has been reported<sup>16</sup>, the viSNE analysis is strongly limited by the total number of cells that can be examined. Computational complexity and crowding effect are the major drawbacks when analyzing massive datasets, consisting of more than 200,000-250,000 events<sup>16</sup>. These limitations may be circumvented by random down-sampling the data. However, decreasing the number of events used for the analysis is detrimental when analyzing rare cells. As an effect of random down-sampling, populations that are more abundant will be retained in the subsampled data at the expenses of rarer cells. The recently developed algorithm Hierarchical Stochastic Neighbor Embedding (HSNE) overcomes the computational complexity.<sup>17</sup> As the name suggests, the algorithm uses a hierarchical approach to cluster the data through different levels. The highest overview level of the hierarchy identifies the various cell populations present in the data, which are shown as separate clusters. The user can decide which cluster(s) to further explore in the following level of hierarchy, based on investigation of the heatmaps showing the relative expression pattern per single marker. This allows to increase the resolution on the chosen cluster(s). In the 2<sup>nd</sup> level, the chosen cluster(s) may be separated into different clusters, as more subtle marker variabilities now define the organization of the clusters. Selection of and zooming-in on interesting cluster(s) can be done until all levels have been investigated, meaning until the data level is reached. At the data level the conventional viSNE map is visualized, showing the single cells.

Because of this hierarchical partitioning approach, HSNE is scalable to large dataset without the need of down-sampling and the risk of losing rare cell populations. However, finding rare cells population of interest may require the exploration of different clusters through different levels.

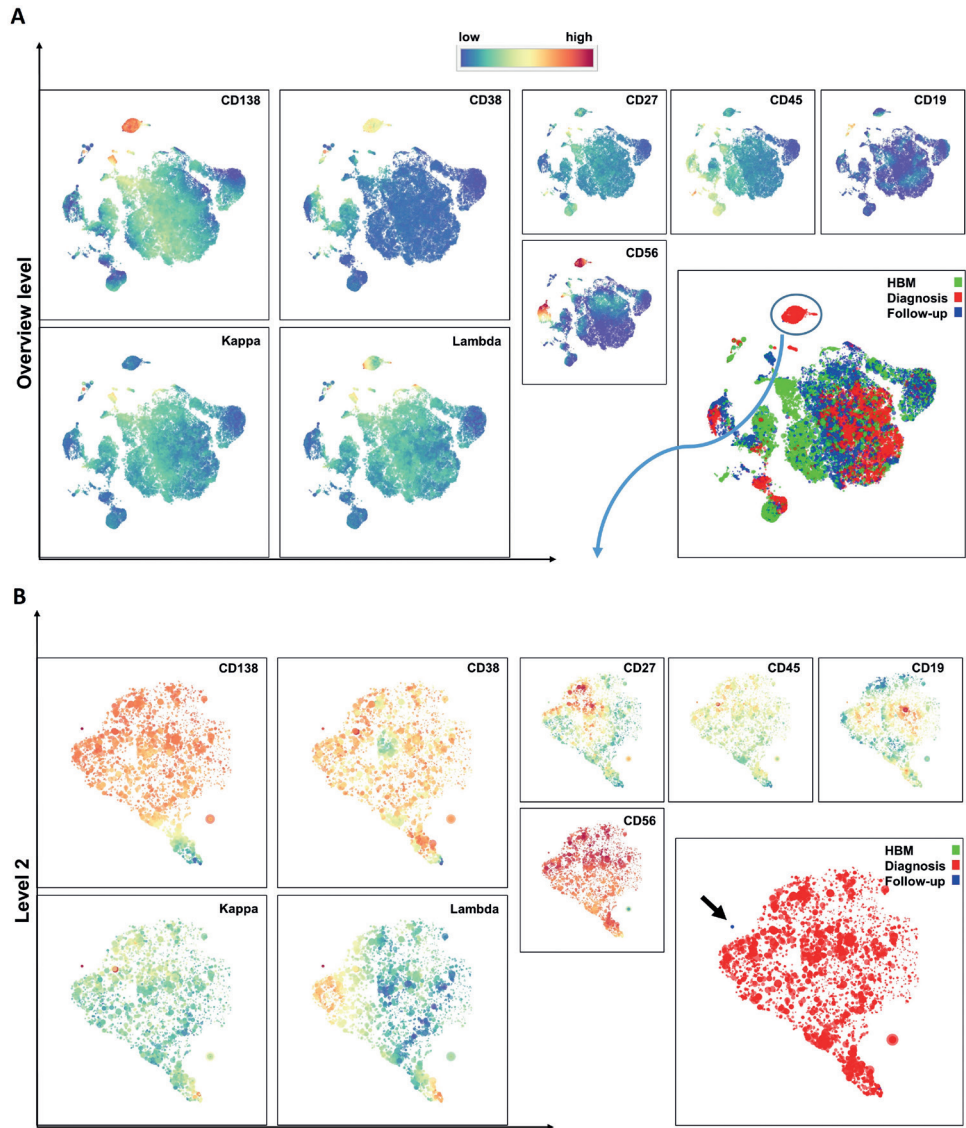
Here representative results are shown for the analysis of 3 Multiple Myeloma patients (ID1-2-3). For each analysis, the complete diagnosis and follow up samples were used, while the 13 control bone marrow samples were subsampled each to 20.000 cells. This subsampling was necessary because the software was not properly functioning with  $8 \times 10^6$  events on our operating system (128 GB of memory, 3.00GHz processor). HSNE was performed using the interactive tool Cytosplore<sup>+HSNE</sup> downloadable from <https://www.cytosplore.org/>.

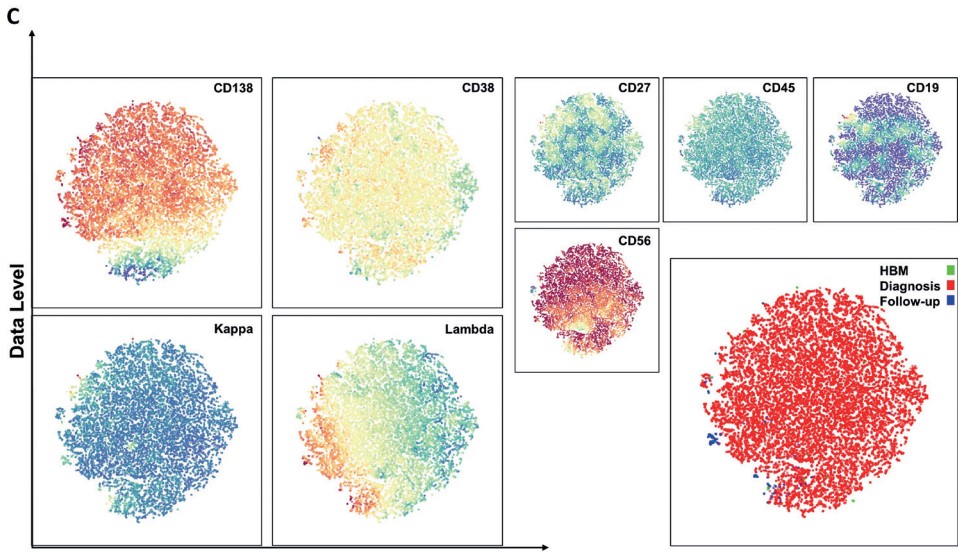
Figure S14 shows the HSNE analysis results for individual ID1. In the overview level (Figure S14a) the most distinctive clusters present in the data were visualised and they were colored based on single marker expressions and sample group (HBM, Diagnosis and Follow-up). Manual selection of interesting clusters was done by investigating patterns in marker expression and absence of control bone marrow cells in the clusters. A distinctive cluster showing CD38+CD138+ expression and Cylg $\lambda$  clonality, which at this overview level consists of cells from the diagnosis sample (red), is visualized in the map and it is highlighted with a circle. We expected that the monoclonal cells from the follow-up sample would overlap with the monoclonal cells at the time point of diagnosis. This cluster was thus selected and it was embedded separately in the second level of hierarchy to achieve increased resolution (Figure S14B). Cell structures within the selected cluster were visualized at a finer scale. Again the resulting clusters were colored according to single marker expressions and sample group. A small cluster with cells mainly belonging to the follow-up sample (blue), indicated by the arrow, is separated from the other cluster mainly characterized by diagnosis sample cells (red). The marker expression, based on the color coding, indicates CD38+CD138+ expression and high expression of both Cylg $\lambda$  and Cylg $\kappa$  for the blue cluster. Based on the high expression of both Cylg $\lambda$  and Cylg $\kappa$ , this cluster was not suspected to consist of monoclonal plasma cells. However, the color coding for Cylg $\lambda$  and Cylg $\kappa$  might be inaccurate in describing the cellular phenotype of such a small cluster. We therefore exported the cells contained in the cluster to deeper investigate their phenotype. The blue cluster comprises 75 cells, of which 72 were identified as clonal plasma cells, based on CD38+CD138+ expression and Cylg $\lambda$  monoclonality (Figure S15A). By comparing the exact expressions per single cell we confirmed that these 72 cells were also found by manual gating (data not shown). However, the majority of the MRD cells found by manual gating which have a similar phenotype were not retrieved from this cluster. This is why we decided to further investigate the whole cell map present in Level 2 at the lower single cell data level, corresponding to the single cell resolution (Figure S14C). A total of 268 cells from the follow-up sample are present at the data level. Some of the follow-up sample cells form small aggregates (left bottom part of the map) and other cells are dispersed through the entire map. Within the big cluster 179 additional monoclonal plasma cells were identified (Figure S15B). We did not find clear differences between the phenotypes of the 72 MRD cells within the small follow up cluster, and the 179 MRD cells within the big aggregate of clusters (data not shown), so it is unclear to us why the MRD cells were not all contained within 1 cluster.

A recall statistic was calculated (Supplementary Table S1) by matching the measured intensities for all markers of all the follow-up sample cells with the manual gating findings. Although the recall is high, the MRD cells were found dispersed over multiple clusters by HSNE and this fragmentation of cells with the same clonal phenotype may hinder the identification of all the MRD cells in a clinical diagnostics setting. Additionally, cells from

the control bone marrow samples were also found in the map corresponding to 0.013% (35 cells) of the total down-sampled events. These cells, overlaying with the MRD cells in the HSNE map, are not plasma cells (Figure S15C).

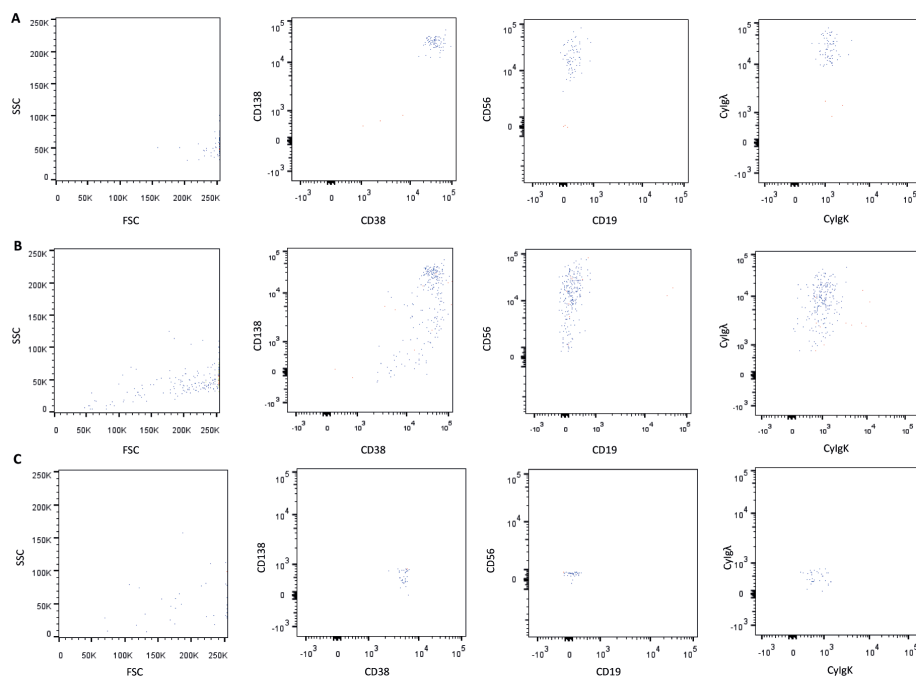
The phenotypes of the cells present in the data level of the HSNE map (Figure S15A+B) can be compared to the phenotypes found by ECLIPSE and manual gating (Figure 5B and C in the main article, respectively).





**Figure S14. HSNE analysis on diagnosis sample cells (ID1), corresponding follow-up sample cells and control bone marrow (HBM) cells from 13 individuals.**

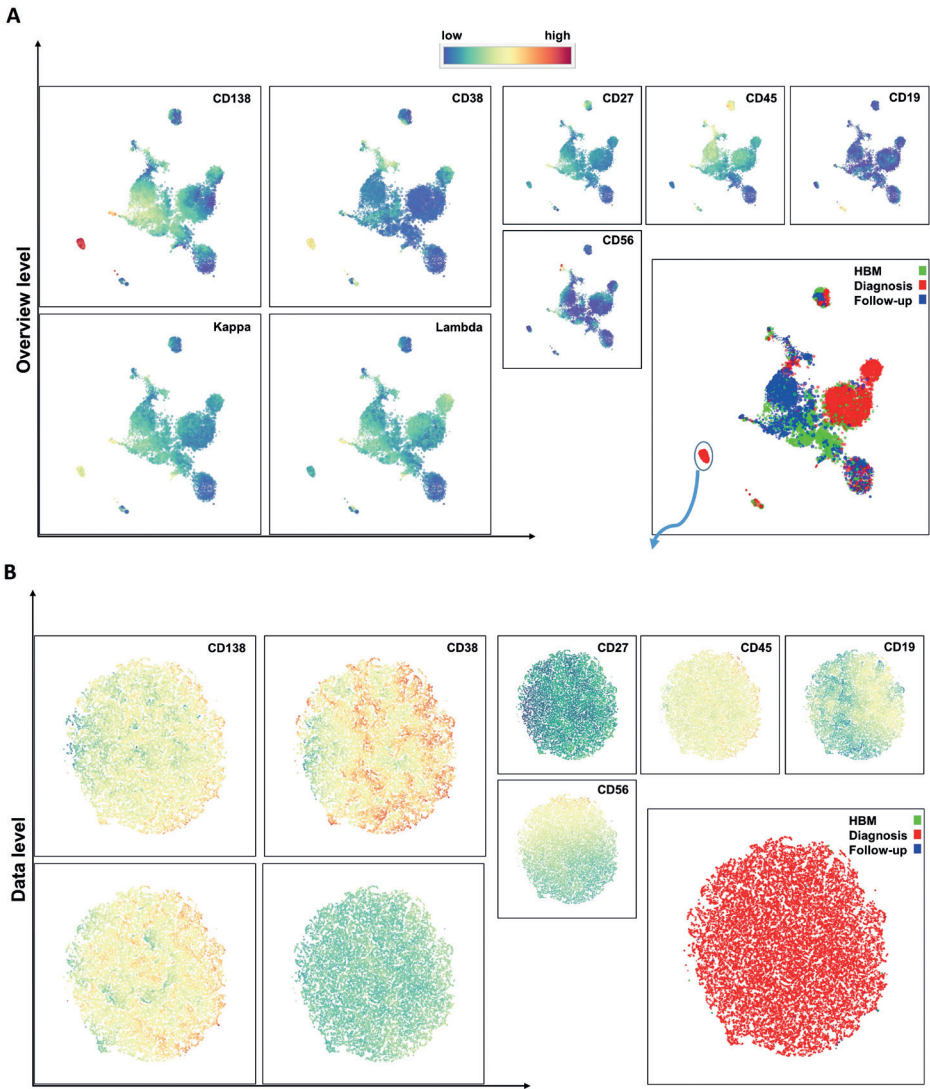
The HSNE maps are colored according to single marker expression levels (see title in the plot) or based on the sample group: HBM (green), diagnosis (red), follow-up (blue). A) Embedding of the cells at the overview level showing the major clusters present in the data. The circle indicates the selection of a cluster mainly constituted of diagnosis cells and CD138+CD38+ and Cylgλ monoclonality. This cluster is explored in the subsequent level, Level 2. B) Level 2 of the hierarchy showing the embedding of cells in the previously selected cluster. A small and distinctive cluster with mainly cells from the follow-up sample is indicated by the arrow. Cells in this cluster were extracted for further analysis while the entire HSNE map was zoomed-in on to be explored at single cell level (data level); C) Single cell resolution level of the clusters present at Level 2.



**Figure S15. 251 MRD cells were identified in the HSNE clusters of hierarchy level 2, ID1.**

A) The cells within the small HSNE cluster in Level 2 (Figure S14b, arrow) were exported and analyzed in FlowJo. This led to the identification of 72 CyIgA<sup>+</sup> monoclonal plasma cells (blue). 3 cells were not (monoclonal) plasma cells as a CD38<sup>dim</sup>CD138<sup>low</sup> and CyIgA<sup>-</sup> phenotype (red) was found. B) Within the big aggregate of multiple clusters 179 additional CyIgA<sup>+</sup> monoclonal plasma cells were identified of the 268 follow up cells in total. C) Control cells of the control bone marrow samples found in the HSNE clusters do not have a monoclonal plasma cell phenotype, with CD38-CD138-CyIgA-CyIgK-expressions.

HSNE analyses were also performed for the diagnosis and follow-up samples of ID2 and ID3. By coloring clusters in the overview level based on sample group a separate cluster with mainly cells from the diagnosis sample was identified for ID2 (encircled in blue, Figure S16A). This cluster has high expression levels of CD138 and CD38 and CyIgK monoclonality. The resolution into the underlying cell populations in this MM candidate cluster was increased through embedding in the next levels. Since no additional clustering was visualized in the more detailed levels, all the cells contained in the initial cluster were zoomed-in until the data level (Figure S16B).

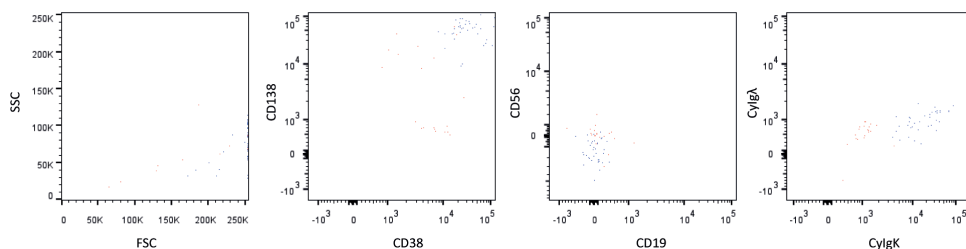


**Figure S16. HSNE analysis on diagnosis sample cells (ID2), corresponding follow-up sample cells and control bone marrow (HBM) cells from 13 individuals.**

The HSNE maps are colored according to single marker expression levels (see title in the plot) or based on the sample group: HBM (green), diagnosis (red), follow-up (blue). A) Embedding of the cells at the overview level showing the major clusters present in the data. The circle indicates the selection of a cluster mainly constituted of diagnosis cells and CD138+CD38+ and CylgK monoclonality. This cluster was zoomed-in on at the single cell resolution level. B) Single cell resolution level of the cluster selected in the overview level.



The majority of the cells present in the cluster are from the diagnosis sample. The monoclonal plasma cell phenotype of the diagnosis cells was confirmed by manual gating (data not shown). Next to cells from the diagnosis sample, also cells from the follow-up sample were present in this cluster, representing a percentage of 0.017% of the cells of the follow-up sample, corresponding to 56 cells (Figure S17). These follow-up cells were compared with the MRD cells identified by manual gating: 16 of these cells were overlapping with the manual gated cells. The remaining cells were partially identified as monoclonal plasma cells, not found by manual gating (20 cells). Strikingly, next to that, 20 cells from the follow-up sample were non-plasma cells, which accounts for 36% of the total follow-up cells present in this cluster. Also a percentage corresponding to 0.012% (32 cells) of the total downsampled control bone marrow cells was found overlapping with the cells from the follow-up sample. These cells were not plasma cells (data not shown).

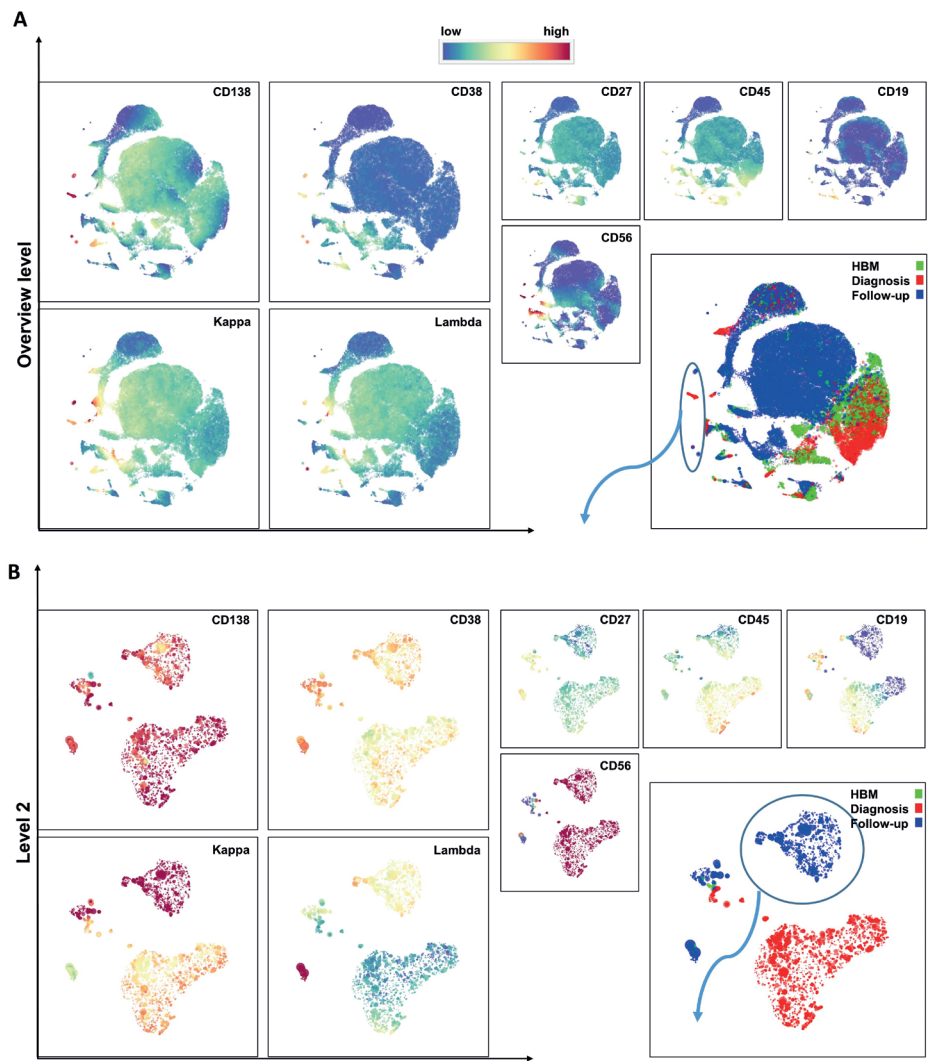


**Figure S17. 36 MRD cells and 20 cells with a non-plasma cell phenotype were identified in the multiple myeloma specific HSNE cluster of ID2.**

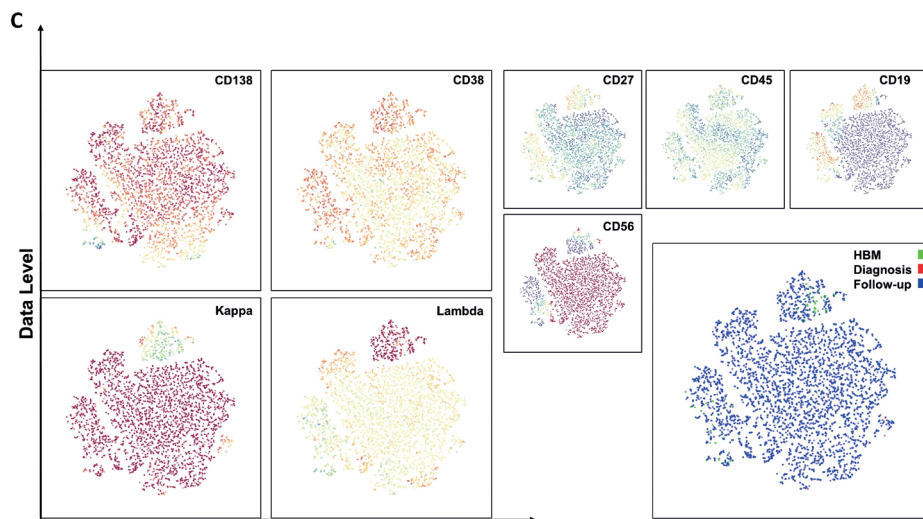
The cells within the HSNE cluster in the overview level (Figure 10a, blue circle) were exported and analyzed in FlowJo. This led to the identification of 36 CyIgK+ monoclonal plasma cells (blue). 20 cells were not plasma cells as a CD38dimCD138low and CyIgK- CyIgλ- phenotype (red) was found.

Figure S18 shows the HSNE analysis results for ID3. Based on marker expression and sample group color-coding several clusters might be selected at the overview level (encircled in blue, Figure S18A). All four cluster had the following characteristics: CD38+CD138+ phenotypes with either CyIgK or CyIgλ expression and not containing cells from control bone marrow samples. The selected clusters showed a more resolved arrangement of four main aggregates at level 2, of which three are mainly made up of follow-up sample cells (Figure S18B, Cluster A, B and C). These three clusters were together selected for further inspection, down to the single-cell level (Figure S18C). The phenotype of the cells in this data level was investigated (Figure S19) and recall statistics were calculated (Table S1). Again, a considerable percentage (~26%) of non-monoclonal plasma cells of the follow-up sample was found overlapping with the monoclonal cells in the HSNE clusters. Like in ECLIPSE, cells with a CD19+CD45+ phenotype were included in the HSNE clusters, which by manual gating were not identified as monoclonal, as a CyIgK;CyIgλ ratio of 2:1 was

found. Also from the HBM samples 0.040% non-plasma cells of the total subsampled control cells, were detected.

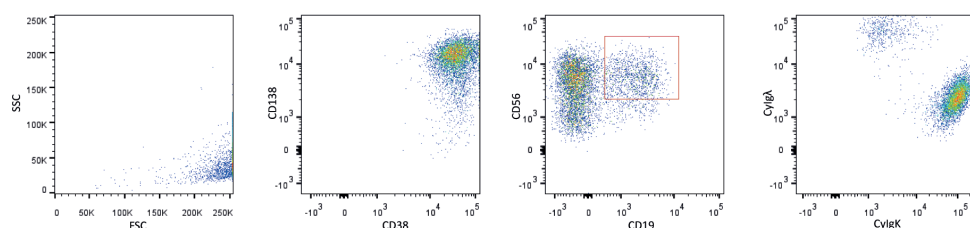






**Figure S18. HSNE analysis on diagnosis sample cells (ID3), corresponding follow-up sample cells and control bone marrow (HBM) cells from 13 individuals.**

The HSNE maps are colored according to single marker expression levels (see title in the plot) or based on the sample group: HBM (green), diagnosis (red), follow-up (blue). A) Embedding of the cells at the overview level showing the major clusters present in the data. The circle indicates the selection of four different clusters mainly constituted of diagnosis and follow up cells and CD138+CD38+ and Cylgk or Cylgl clonality. These clusters were zoomed-in on in Level 2. B) Embedding of cells at Level 2 of the hierarchy. Clusters A-B-C, which are mainly constituted by follow-up sample cells, are separated from diagnosis sample clusters. Based on marker expressions cluster A-B-C might be associated to clonal phenotype and they are zoomed in at the single cell resolution level. C) Data level: Single cell resolution level of clusters A-B-C selected in level 2.



**Figure S19. 4665 MRD cells and 1610 non-monoclonal plasma cells were identified in the multiple myeloma specific HSNE cluster of ID2.**

The cells within the HSNE data level (Figure 12C) were exported and analyzed in FlowJo. This led to the identification of 4665 CylgK+ monoclonal plasma cells. 1610 cells were non-monoclonal plasma cells, of which most cells were characterized with a CD19+CD45+ phenotype (red gate) and a CylgK:CylgA ratio of 2:1.

Table S1 summarizes the number of MRD cells found in the HSNE analysis compared to the manual gating findings, assumed as conventional standard. High recall was obtained for all the three samples analyzed. However, HSNE is not suitable as a diagnostics tool to find MRD.

Preferably we would have found 1 cluster containing all the monoclonal plasma cells from the follow-up sample. In contrast, in 2 out of 3 cases the MRD cells were dispersed in the resulting HSNE map through multiple clusters (ID1 and ID3). For example, in the analysis of ID1 we checked the cells in the small cluster and found 72 MRD cells. Out of curiosity we checked the large adjacent cluster (Figure S14), although the CD38 and CD138 expressions seemed lower than for the cells in the small cluster and no clear CylgK<sup>high</sup> or Cylgλ<sup>high</sup> phenotype was found. Surprisingly, there were 179 monoclonal plasma cells present in this cluster, meaning that easily 71% of the monoclonal plasma cells in the HSNE analysis could have been missed.

The distribution of the MRD cells over multiple clusters might be due to the fact that with zooming into deeper levels, HSNE focuses on more subtle differences within the clonal cells. Plasma cells can have different phenotypes regarding CD19, CD45 and CD56 expression. For ID3 we checked whether the cells from the various clusters differed based on the expression levels of CD45 and CD19. That was not the case. In some clusters 1 phenotype was more represented than the other, but the cell phenotypes were overlapping between the clusters. However, cluster C (Figure S18,S19) mainly contained Cylgλ positive cells, while the CylgK positive cells of this phenotype were located in another cluster. The resulting over-fragmentation of the MRD cells is suboptimal. In fact, when only a part of the clusters is analyzed and the rest are overlooked or hidden in the HSNE map, this could lead to wrong diagnosis: either falsely MRD positive, or falsely MRD negative.

Next to the dispersed character of the MRD cells in the HSNE results, for 2 out of 3 follow-up samples (ID 2 and ID3) a considerable amount of non-(monoclonal)plasma cells (25-30%) were found overlapping with MRD cells. This is not directly visible in the HSNE plots. In the higher HSNE cluster levels the cluster colors of single marker expressions represent the expression of the majority of the cells within this cluster, so expressions of less common non-plasma cells within that cluster are not well represented by the color of the cluster.

Also, at the data level the crowding effect remains a problem, even when smaller clusters are selected. This was the case for 2 out of 3 analyses (ID1 and ID2): at the data level the cells formed a homogeneous cloud, one cell overlaying the other. Therefore, single cells can be missed. This stresses the fact that phenotyping the cells found within the clusters in a conventional gating program remains necessary, to prevent overestimation of the number of MRD cells found. If in the future the percentage of MRD cells in the bone

marrow will be leading for the decision-making to change the treatment regime, it is of great importance that the percentage of monoclonal plasma cells found by HSNE is accurate.

In addition, in all the HSNE analyses, MRD cells were overlapping with a percentage (0.01-0.04%) of non-plasma cells from healthy bone marrow samples. This makes the HSNE results less reliable from a diagnostic point of view and an extensive analysis to check and separate malignant cells from normal phenotype is needed.

Regarding analysing time, running the Cytosplore<sup>+HSNE</sup> program, performing both the initial and the subsequent embedding of the cells into clusters in the different hierarchal levels, took quite a short time span (order of 5-10 minutes with a sample size  $\sim 1 \times 10^6$  events). However, the analysis costs at least one hour. Identification of relevant clusters may be challenging, because HSNE, just as viSNE, does not provide a direct view of the co-expression of the marker measured associated to a specific cells subset or cluster. Information about the correlations between the markers can only be obtained by investigating multiple single marker heat map plots. Characterizing multiple clusters from the first overview level through subsequent levels to the data level, exporting the cells and verifying the monoclonality by backgating in conventional analysis software is very time consuming. In that case it would be faster to analyse the data by a skilled technician using conventional gating software.

An additional (indirect) disadvantage of the HSNE analysis by using Cytosplore<sup>+HSNE</sup> is the impossibility to save the analysis performed. This prevents the researcher to explore additional landmarks and visualize different embedding than the ones observed in previous (one-run) analysis. In this case, a new hierarchy needs to be recreated leading to slightly different representation of the various maps per level.

In conclusion, the HSNE algorithm works better with large datasets than viSNE. The hierarchical levels are a smart way to downsample data. However, there are important drawbacks regarding the identification of rare cells which, in the case of follow-up samples from multiple myeloma patients, could lead to falsely positive or falsely negative diagnoses.

Table S1.

	# monoclonal plasma cells HSNE analysis	# cells manual gate	Recall	Fold-increase HSNE
ID 1	230	163	136/163= 0.83	1.41
ID 2	36	19	16/19=0.79	1.89
ID 3	4664	4327	4146/4327=0.96	1.08

Comparison of the number of cells found in the HSNE analysis with the manual gate for 3 patients. Recall statistics are calculated per sample, representing the amount of manual gated cells found also by the HSNE algorithm. Fold-increase summarizes the number of cells in the HSNE analysis compared to the manual gate.

## References

- 1 Bro, R. & Smilde, A. K. Principal component analysis. *Analytical Methods* 6, 2812-2831, doi:10.1039/C3AY41907J (2014).
- 2 Folcarelli, R. *et al.* Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm. *Scientific Reports* 8, 10907, doi:10.1038/s41598-018-29367-w (2018).
- 3 Finak, G., Perez, J. M., Weng, A. & Gottardo, R. Optimizing transformations for automated, high throughput analysis of flow cytometry data. *BMC bioinformatics* 11, 546, doi:10.1186/1471-2105-11-546 (2010).
- 4 Roederer, M. Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats. *Cytometry* 45, 194-205, doi:doi:10.1002/1097-0320(20011101)45:3<194::AID-CYTO1163>3.0.CO;2-C (2001).
- 5 Bro, R. & Smilde, A. K. Centering and scaling in component analysis. *Journal of Chemometrics* 17, 16-33, doi:10.1002/cem.773 (2003).
- 6 van Velzen, J. F., van den Blink, D. & Bloem, A. C. Inability of a monoclonal anti-light chain antibody to detect clonal plasma cells in a patient with multiple myeloma by multicolor flow cytometry. *Cytometry Part B: Clinical Cytometry* 84B, 30-32, doi:doi:10.1002/cyto.b.21044 (2013).
- 7 Krejci, J. *et al.* Monocytes and granulocytes reduce CD38 expression levels on myeloma cells in patients treated with daratumumab. *Clinical cancer research : an official journal of the American Association for Cancer Research* 23, 7498-7511, doi:10.1158/1078-0432.CCR-17-2027 (2017).
- 8 Durie, B. G. M. *et al.* International uniform response criteria for multiple myeloma. *Leukemia* 20, 1467, doi:10.1038/sj.leu.2404284 (2006).
- 9 Malavasi, F. *et al.* Evolution and Function of the ADP Ribosyl Cyclase/CD38 Gene Family in Physiology and Pathology. *Physiological Reviews* 88, 841-886, doi:10.1152/physrev.00035.2007 (2008).
- 10 Rawstron, A. C. Immunophenotyping of Plasma Cells. *Current protocols in cytometry* 36, 6.23.21-26.23.14, doi:doi:10.1002/0471142956.cy0623s36 (2006).
- 11 Terstappen, L., Johnsen, S., Segers-Nolten, I. & Loken, M. Identification and characterization of plasma cells in normal human bone marrow by high-resolution flow cytometry. *Blood* 76, 1739-1747 (1990).
- 12 Wijdenes, J. *et al.* A plasmocyte selective monoclonal antibody (B-B4) recognizes syndecan-1. *British Journal of Haematology* 94, 318-323, doi:doi:10.1046/j.1365-2141.1996.d01-1811.x (1996).
- 13 Peceliunas, V., Janiulioniene, A., Matuzeviciene, R. & Griskevicius, L. Six color flow cytometry detects plasma cells expressing aberrant immunophenotype in bone marrow of healthy donors. *Cytometry Part B: Clinical Cytometry* 80B, 318-323, doi:doi:10.1002/cyto.b.20601 (2011).
- 14 Paiva, B. *et al.* Minimal residual disease monitoring and immune profiling in multiple myeloma in elderly patients. *Blood* 127, 3165-3174, doi:10.1182/blood-2016-03-705319 (2016).

- 15 Amir el, A. D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology* 31, 545-552, doi:10.1038/nbt.2594 (2013).
- 16 DiGiuseppe, J. A., Tadmor, M. D. & Pe'er, D. Detection of minimal residual disease in B lymphoblastic leukemia using viSNE. *Cytometry. Part B, Clinical cytometry* 88, 294-304, doi:10.1002/cyto.b.21252 (2015).
- 17 van Unen, V. *et al.* Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. *Nature Communications* 8, 1740, doi:10.1038/s41467-017-01689-9 (2017).



# 7

## CHAPTER 7



# Phagocytosis of microplastics leads to accelerated cell death in human neutrophils.

Selma van Staveren<sup>1,2</sup>, Erinke van Grinsven<sup>1</sup>, Tom Vos<sup>1</sup>, Fabienne Derks<sup>1</sup>, Josse Depla<sup>1</sup>, Pim de Voogt<sup>3,4</sup>, Leo Koenderman<sup>1</sup>, Nienke Vrisekoop<sup>1</sup>

1 Department of Respiratory Medicine, Laboratory of Translational Immunology, UMC Utrecht, The Netherlands

2 TI-COAST, Amsterdam, The Netherlands

3 Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands.

4 KWR Watercycle Research Institute, Nieuwegein, The Netherlands.

Microplastics are found in many consumables such as cosmetics, toothpaste, water, beer and seafood. It is still unknown what the influence of microplastics is on human health, but ingestion of these particles has been shown to be harmful for sea organisms. Immune cells are likely the first cells to detect and respond to these particles when microplastics translocate from the gut or the lung to interior of the body. Neutrophils are the most abundant white blood cell in our peripheral blood and they have a high capacity to phagocytose (internalize) foreign particles. In this report the behaviour of neutrophils was studied upon interaction with different microplastics. Since it is unlikely that neutrophils are able to degrade plastic particles, we hypothesized that such interaction with microplastics would affect neutrophil function.

We studied the consequences of interaction of microplastics with healthy donor neutrophils *in vitro* using various fluorescent polystyrene beads with a size range of 49 nm to 7.3  $\mu\text{m}$ . Neutrophil survival and phagocytosis of fluorescent microplastics was determined by using (imaging) flow cytometry. Furthermore, we studied neutrophils phagocytosing 1  $\mu\text{m}$  sized microplastics in living mice using intravital microscopy.

Using imaging flow cytometry we found that neutrophils phagocytosed 7.3  $\mu\text{m}$  and 880 nm particles more readily than the smaller 382 and 49 nm microplastics after 45 min incubation. As less than 1% of the neutrophils were found to contain microplastics at neutrophil to microplastic ratio of 1, phagocytosis seemed relatively rare. In line, neutrophils in the mouse ear mostly seemed to ignore the non-opsonized 1  $\mu\text{m}$  microplastics. On the other hand, *in vitro* microscopic analysis indicated that the 7.3  $\mu\text{m}$  particles were readily phagocytosed by human neutrophils and were released again upon cell death. When analyzing neutrophil survival, co-incubation of neutrophils with smaller particles (49-880 nm) led to fast death of 49-53% of the neutrophils within 45 minutes, whereas almost all neutrophils (97%) died during 45 min incubation with 7.3  $\mu\text{m}$  particles. Long term survival of neutrophils which outlasted the initial 45 minutes was unaffected.

Phagocytosis of 880 nm microplastics by surviving neutrophils did not impair their subsequent phagocytosis of bacteria. In fact, surviving neutrophils which phagocytosed microplastics preferentially phagocytosed bacteria. Co-phagocytosis of bacteria and microplastics did not decrease the killing capacity of the surviving neutrophils.

In conclusion, neutrophils can associate with microplastics *in vitro* and *in vivo* and this leads to death upon ingestion with large particles *in vitro*. Interaction of small particles also induces cell death, however to a lesser extent than the large particles.

About half the neutrophils survive the initial 45 minute incubation with the smaller particles and the phagocytosis, survival and bacterial containment of surviving neutrophils appears unaffected. Future *in vivo* mouse experiments should be conducted to fully understand the possible health hazards of these particles.

## General Introduction

The production of plastics has been steadily increasing for more than half a century. In 1950 the total plastic production was approximately 1.9 million tons, while in 2013 the total was approximately 330 million tons.<sup>1</sup> A portion of these plastics ends up in the marine environment, and enter our food chain. This may be dangerous to human health. In this general introduction the sources and exposure of microplastics in our environment and food is shortly reviewed. Next, we will focus on the risks of microplastics on the human immune system, as we examined the effects of microplastics on neutrophil function in particular.

Plastics in the marine environment are broken down by the mechanical action of waves and currents, the uptake by birds and fish, chemical processes such as oxidation and hydrolysis, and exposure to UV light.<sup>2,3</sup> Therefore most plastics deposited in the Ocean will exist for a time as 'microplastics', a term coined by Professor Richard Thompson in 2004.<sup>4</sup> It is important to emphasize that not all microplastic pollution originates from the breakdown of larger plastic particles. 'Microbeads' are added as mild abrasives to cosmetics such as whitening toothpastes, facial scrubs and shower gels. They are too small to be filtered out by most water treatment facilities, and therefore potentially pose a contamination risk to drinking water. This caused enough concern in Illinois where a new State Law banned the sale of cosmetics containing such microbeads in 2014.<sup>5</sup> Another direct vector through which microplastics can penetrate marine and other environment, is through small plastic pellets which are used as precursors for macroplastic construction, and may enter environments through spillage or deliberate 'dumping'. For instance, in 2012 a cargo ship lost in excess of 165 tons of these pellets along Hong Kong's coastline, during Typhoon Vicente.<sup>6</sup>

Although it is very challenging to quantify the amount of plastic in the Oceans, a team from The University of Connecticut attempted to quantify the amount of plastics in all oceans in 2014. They estimated the Figure of plastic particles in the ocean to be in excess of 5 trillion, with over a quarter of a million tons floating on the surface.<sup>7</sup> They also estimated the percentages of each size class of plastics, using the slightly atypical definition of microplastics particles of less than 4.75 mm. The most surprising find of this study, even to those familiar with microplastics, was the estimate that 92.4% of all plastic particles in the Ocean exist as microplastics.<sup>7</sup>

Plastic particle detection remains a challenge, especially in the lower size ranges. Preferably, particles should be analysed directly in the environment, without the need for water sampling and particle isolation. Unfortunately, detection of particles of <300 nm in size in the aquatic environment is still impossible.<sup>8</sup> Floating flow cytometers have

been developed which can be integrated in a buoy.<sup>9</sup> These flow cytometers are mainly used for monitoring algae populations in the aquatic environment, but other particles within similar size range, like plastics, are also measured. Unfortunately, the differentiation between microplastics and other particles in surface waters is challenging, as microplastics do not contain fluorescent properties. Recently stimulated Raman scattering was used for the identification of plastic particles from a water sample. Also this technique needs improvement to be used for small sized particles in a high throughput mode.<sup>10</sup>

The above mentioned detection techniques deal with instrumental noise, which may interfere with the signal of the particle. As said before, microplastics might not be very easily identified by a flow cytometer as plastics may overlap with other particles based on fluorescence and scatter properties. (van Staveren et al, unpublished) A promising approach to solve these problems, would be the integration of multidimensional data analysis methods which use the multidimensional properties of the measured events to discern these signals from the noise signal, or from other small non-plastic particles. Various multi-dimensional analysis techniques are developed which could be applied to the data generated from flow cytometry or stimulated Raman scattering.<sup>10, 11, 12</sup>

There is no consensus on the exact size of 'microplastics' in literature with a range of particles being less than 5 mm in diameter, and some less than 1 mm.<sup>13</sup> This issue is further complicated by the term 'nanoplastics' or 'plastic nanoparticles', often used for plastics with sizes less than 1  $\mu\text{m}$ <sup>14</sup> or 100 nm.<sup>8</sup> For the sake of simplicity all plastic particles in this report ( $< 8 \mu\text{m}$ ) will be referred to as 'microplastics' or microplastic particles.

Regardless of the exact sizes and compositions of these particles, they are of great concern as they would not be visible in food, and hence they could be consumed without knowing. There is increasing evidence that zooplanktons consume the smaller microplastic particles, and therefore the particles can enter the food chain at early trophic stages.<sup>15</sup> Plastic pollution is serious and ubiquitous throughout the many oceanic ecosystems, and this is shown clearly by the presence of microplastics in diverse and multiple marine organisms. For instance a 2015 study showed the presence of microplastics in all specimens sampled in a six site survey of the French-Belgian-Dutch coast, the sample focused on mussels and lugworms, which are filter feeders. Filter feeders are predisposed to take up particulate contaminants, and therefore are a large concern for transfer to humans.<sup>16</sup> Karlsson et al showed that mussels had much higher levels, approximately 1000-fold, of MPs compared to surrounding water and sediment.<sup>17</sup> Another study found that around 20% of 26 fish species studied contained microplastics (or at least one microplastic particle).<sup>18</sup>

Freshwater rivers are a key vector by which microplastics move towards the oceans, with the microplastics entering the rivers through waste water treatment plants, and mishandled plastic waste.<sup>19</sup>

Besides marine microplastics, there is increasing evidence that microplastics are prevalent amongst terrestrial environments, and are common contaminants in certain foods. For instance, a study in Southeastern Mexico found that chicken gizzards contained microplastic particles.<sup>20</sup> Additionally, a study of 24 popular German beers in 2014 revealed the presence of plastic fibres, fragments and granules in varying sizes, concentrations, and compositions.<sup>21</sup> The list continues to grow when it comes to finding further sources of microplastic contamination, with even table salt being a culprit; confirmed by several studies, including by Iñiguez et al which demonstrated that all twenty-one samples that were tested contained microplastic particles, with the range being in the order of 50-280 microplastic particles per kilogram.<sup>22</sup> Also, it is very common for people to drink from plastic bottles, be they single-use, or reusable. These bottles may be another prominent source of microplastics.

A study from February 2018 used Raman spectroscopy to detect microplastic particles of down to a size of 1  $\mu\text{m}$  (experimental limitations restricted the detection size to 1  $\mu\text{m}$ ), next to polyethylene terephthalate (PET) additives leaching from the plastic itself directly into the water. The study looked at 32 samples from 21 brands of mineral water, 22 of which used PET bottles, and 10 used glass. The study found that reusable PET bottles contained more microplastic particles per litre compared to single use PET bottle. A surprising find of the study was that the glass bottles contained comparable concentrations of microplastics as single-use PET bottles.<sup>23</sup> In these samples, about 90% of plastic particles were less than 5  $\mu\text{m}$ , and 40% were less than 1.5  $\mu\text{m}$ . The PET bottles largely contained microplastics comprised of PET, however the glass bottles had a variety of polymers, such as polyethylene, polypropylene, and a styrene-butadiene copolymer.<sup>23</sup> Therefore the likelihood that one has not been exposed to microplastics from any source is very low.

## Introduction to microplastics and the risk to human health

The rising prevalence of microplastics in the environment has led to *in vitro* and *in vivo* testing of the particles in an effort to see the effects on different cell lines and organs. For instance a mouse study by Deng et al in 2017 was a key mammalian study that specifically observed the effects of microplastics in an *in vivo* model. The mice were fed with polystyrene microplastic particles of 5 and 20  $\mu\text{m}$ . It was found that the 20  $\mu\text{m}$  microplastic particles preferentially accumulated in the liver, and the 5  $\mu\text{m}$  particles were found in much higher concentrations versus the 20  $\mu\text{m}$  particles in the gut. The kidneys showed a higher concentration for the 5  $\mu\text{m}$  particles, but not as substantial a difference compared to the other two aforementioned tissues. The microplastic mice showed a notable drop in ATP concentration in liver mitochondria and disturbed lipid metabolism.<sup>24</sup> A similar study by Lu et al in 2018 fed mice 1000  $\mu\text{g/L}$  of 0.5  $\mu\text{m}$  and 50  $\mu\text{m}$  polystyrene beads, and found a disturbed lipid metabolism, and altered triglyceride levels. They went further and claimed that the relative bodyweight of the mouse decreased after a certain period of oral microplastic exposure.<sup>25</sup> The paper also indicated that the polystyrene microplastic particles altered the microbiome of the mouse.<sup>25</sup> Another significant observation was the overall decrease of mucus production observed in the gut for all mice fed with microplastics. This observation is relevant for the immune system, as mucus secretions in the gut act as an immunological barrier to microbes, and are linked to the health of the microbiome and gut flora diversity.<sup>26</sup>

Microplastics that penetrate the gut epithelial layer might be encountered by immune cells. Our immune system defends our body against foreign particles, like pathogens. Microplastics that enter our body via our gastro-intestinal tract might also be considered foreign. Many early phagocytosis studies used small latex or polystyrene beads to study phagocytosis.<sup>26</sup> This provides some idea of what can happen when certain immune cells encounter such particles. For instance a study in 1986 using rat peritoneal macrophages demonstrated a clear 'preference' for those of 1100 nm, compared to those of 100 nm.<sup>28</sup> This provides the first evidence that phagocytes are 'searching' for particles in size more comparable to that of a bacterium. There is a very early study from 1965 by Movat et al that is likely the first to observe the fact that small latex spheres (in this case of 300 nm) cause platelet aggregation, and that the chelator EDTA is capable of inhibiting this.<sup>29</sup>

Neutrophils are the most abundant white blood cell type present in the human blood and the first to arrive at a site of infection to phagocytose and kill bacteria. We hypothesized that neutrophils will be one of the first immune cells to interact with microplastics in the human body. Here we studied the interaction between human neutrophils and microplastics, and the subsequent effects on neutrophil function. We co-cultured the neutrophils with microplastics for a short time period and evaluated whether neutrophils

phagocytose microplastics. Also we evaluated neutrophil function after incubation with microplastics by studying the phagocytosis *in vitro* and *in vivo* and containment of bacteria and the effect on the neutrophil life span *in vitro*.

## Methods

### Whole blood assays

Human blood samples were collected from anonymous, healthy volunteers between the age of 18–65 years, male and female, using sodium heparin as an anticoagulant. All donors gave informed consent under protocols approved by the Medical Ethical Committee of the University Medical Center Utrecht. Whole blood was counted on a CellDyn Emerald cell counter (Abbot Diagnostics). The granulocyte count was used to calculate relevant neutrophil : microplastic ratios of each type of microplastic.

### Neutrophil isolation

To isolate platelet-rich-plasma for incubation with microplastic particles, whole blood was centrifuged at 155 RCF for 20 minutes, and platelet-rich-plasma was aspirated. Experiments were carried out with autologous plasma.

Neutrophils were isolated using Ficoll-Paque™ PLUS (GE Healthcare). Whole blood (minus PRP) was diluted with Phosphate-buffered Saline + 0.32% sodium citrate + albumin (4 g/l) (PBS2+) in equal volume to original donation sample (most often 9mL). The PBS2+ and whole blood (minus PRP) mix was slowly layered on top of Ficoll Paque™ PLUS, so the layers remained separate. Hereafter, the tubes were centrifuged at 880 RCF for 20 minutes. After centrifugation, plasma and the ring fraction were removed and erythrocytes were lysed using a lysis buffer that consisted of 150mM  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$  and 0.1mM  $\text{Na}_2$  ethylenediaminetetraacetic acid (EDTA) dissolved in  $\text{ddH}_2\text{O}$  and adjusted to a pH of 7.4. The remaining neutrophils were washed twice and resuspended in 60:40 HEPES 3+:PRP. HEPES3+ consisting of Hepes buffer (20 mM Hepes, 132 mM NaCl, 6 mM KCl, 1.2 mM  $\text{KH}_2\text{PO}_4$ , 1 mM  $\text{MgSO}_4$ ) supplemented with 5 mM glucose, 1 mM  $\text{CaCl}_2$ , and 0.5% (w/v) human serum albumin. The PRP was used to facilitate phagocytosis, and given at 40% to represent whole blood concentrations (approximately).



### **Phagocytosis Assays of Microplastics and *S. aureus* on the FACS Canto (Becton Dickinson) and the ImageStream®X Mark II Imaging Flow Cytometer (Merck)**

Whole blood was aliquoted into 500 µl samples, and microplastics (Table 1) were added in a low volume (1-5 µl); which was calculated from the granulocyte count given by the Cell-Dyn Emerald. If necessary microplastics were diluted using PBS for ImageStream Experiments, and HEPES3+ for flow cytometry experiments. Microplastics were sonicated for 5 minutes at room temperature in an ultrasonic cleaner (VWR) prior to addition to blood. Incubation of whole blood from healthy donors was performed at 37°C for 45 minutes for ImageStream assays, and 40 minutes for Canto assays. Both were shaken at 180 RPM in an Innova 44 Incubator.

For *S. aureus* phagocytosis assays, an additional incubation of 40 minutes was performed as above, but in the dark in the presence of a GFP expressing MW2 *S. aureus* strain<sup>30</sup> at indicated MOI. In this assay samples were fixed with 1% paraformaldehyde for 20 minutes, then samples were stained with LIVE/DEAD™ Fixable Dead Cell Stain kit (L34955; Invitrogen). Samples were then stained with anti-CD16 (Fc RIII) antibody (APC, 17-0168-42; eBioscience) and anti-CD45 antibody (PE-Cy7, 557748; BD Bioscience). Next flow cytometry was performed on the BD FACS Canto™ II Flow Cytometer (BD Bioscience), with further data analysis using FlowJo® software. In these assays only PE fluorescent microplastic beads of 880 nm diameter (PFP-0558; Kisker Biotech, Steinfurt, Germany) were used.

For the ImageStream assays, erythrocytes were lysed, and remaining cells were fixed using BD FACS™ Lysing Solution. Prior to the experiment, the BD FACS™ Lysing Solution was diluted 10 times in sterile water, and 5% of the volume was added to each sample, followed by incubation for 15 minutes at room temperature. Cells were washed two times by centrifugation at 494 RCF, room temperature, for 5 minutes. Cells were taken up in 50 µL PBS. The samples were then stained with CD16-APC-H7 antibody (clone B73.1, 561306; BD Biosciences) for 20 minutes at room temperature. Then cells were washed and resuspended in 50 µL in PBS. Flow cytometry was performed on the ImageStream®X Mark II Imaging Flow Cytometer (Merck), and data was analyzed on IDEAS® software (Merck).

Samples were analyzed on the ImageStream visual flow cytometer with at least 100,000 events. The excitation laser for the beads had a wavelength of 488nm, and the emission was recorded in channel 2, which detects emission spectra between 505 and 560nm. The CD16 antibody linked to APC-H7 was excited using the 642nm laser, and its emission spectra was recorded in channel 6, between 740 and 800nm. Note that due to the use of channel 6, side-scatter was not recorded. The bright field was recorded in channel 1, and is illuminated using 6 spectrally filtered LEDs. The magnification for all assays was 60x, and

the voltage of the 642nm laser was constant. The speed was always set to slow, because of the superior image quality.

*Determining internalization by IDEAs software*

First, cells which were out of focus were excluded, using gradient RMS exclusion criteria present in the IDEAs software, and then singlets were selected using a gating strategy based on a plot of aspect ratio and area in the bright field channel. To determine the percentage of neutrophils which phagocytosed microplastics, and exclude the microplastics attached to the cell membrane, an adaptive erode gate was crated on the IDEAs software. In this case it was set to 80% cell area; the ‘adaptive’ parts of this means it adapts to the cell shape, and erodes 20% of the cell area evenly from the membrane.

**Table 1. Fluorescent microplastics used in the assays**

Microplastics Kisker Biotech GmbH & Co. KG.	fluorescence	product no.	size (µm)	w/v
Polystyrene fluorescent particles 0.04-0.09 µm	yellow	PFP-00552	0.049	1%
Polystyrene fluorescent particles 0.4-0.6 µm	yellow	PFP-0552	0.38	1%
Polystyrene fluorescent particles 0.7-0.9 µm	yellow	PFP-0852	0.79	1%
Polystyrene fluorescent particles 6.0-7.9 µm	yellow	PFP-6052	7.3	1%

**Short-Term Survival Assay in suspension- Net Cell Count and flow cytometry analysis**

Microplastics were added to neutrophils at ratios equivalent to those used for the long-term assays. They were incubated as mentioned for 45 minutes at 37°C, whilst shaking at 180 RPM, and then manual counting was performed. After counting, cells were fixed with 2%PFA in equal volume, and then examined using flow cytometry.

**Long-term Survival Assay in fibrin- thrombin gel**

200,000 neutrophils were aliquoted into Eppendorfs tubes, to which non-fluorescent microplastics of 49nm, 382 nm, 880 nm or 7.3 µm at specific microplastic to neutrophil ratios (MNR) were added. Microplastics were sonicated for 5 minutes prior to incubation. Incubation was performed at 37°C, for 45 minutes, whilst shaking at 494 RCF in an New Brunswick™ Innova 44 Incubator (Eppendorf).

Cells were centrifuged at room temperature, for 5 minutes at 1500 RPM. The supernatant was removed, and the pellet was resuspended in 80 µL (40:40 PRP:HEPES3+ mix). This mix promotes survival, and so differences in survival should be clearer. The gel was made in

96 Well Corning Incorporated Costar® 3603 Assay plate, with black base; samples were added, followed by 10 µL Thrombin and 2.5 µL Propidium Iodide. Lastly 8uL of fibrin was added with rapid resuspension on ice, as the gel forms quickly. Samples were incubated for 140 hours in FLUOStar Optima microplate reader (BMG Labtech). After 140 hours, the plate was removed, 0.5% Triton X solution was added, and incubated for another 10 hours.

It is important to emphasize that the microplastic particles used for the gel survival assay had no internal fluorescence. The sizes were the same, as was the material, polystyrene.

**Table 2. Non-fluorescent microplastics used in the survival assays**

Microplastics Kisker Biotech GmbH & Co. KG.	fluorescence	product no.	size (µm)	w/v
Polystyrene beads 0.05 um in size	plain	PPs-0.05	0.047	2.50%
Polystyrene beads 0.4 um in size	plain	PPs-0.4	0.382	2.50%
Polystyrene beads 0.8 um in size	plain	PPs-0.8	0.792	2.50%

**Confocal Phagocytosis Assays using pHrodo-labelled Microplastics**

**Labelling of microplastics with pHrodo**

Microplastics beads were labelled with pHrodo™ Red, succinimidyl ester (pHrodo™ Red, SE, Thermofisher Scientific) Catalog number: P36600. This pH sensitive dye binds generic proteins, such as those in human serum. Its fluorescence increases as pH decreases, and is sensitive in the range of pH7 down to pH 4. Sonicated beads were incubated with human serum for 1 hour at room temperature, and then 1mL H3+ was added. Beads were then spun down at 2000 RCF for 7 µm beads, and 5000 RCF for 880 nm beads. Resuspension was performed in 140uL 0.1M NaHCO<sub>3</sub> at pH 8.4.

One aliquot of pHrodo red dye was suspended in 10µL dimethyl sulfoxide, and then added to the bead solution above. This mix was then incubated in the dark for 1 hour at room temperature. Four washes were performed in H3+, using spin speeds as above (for each size). All spins were 10 minutes. The final suspension was also in H3+.

**Confocal phagocytosis assay**

Neutrophils were isolated from the whole blood of one healthy donor using the aforementioned Ficoll method. The concentration was then set at 6.67 million cells per mL, in HEPES3+. This concentration is optimal for use in µ-Slide 8 Well Chamber Slide (IBIDI 80826), as the 120uL added gives 800,000 cells per well. The pHrodo beads were added, using 50uL of the phrodo-labelled mix. The fluorophore is excited by the 561nm laser on the confocal, and peak emission is approximately 590nm.

Confocal images were then taken every thirty seconds for 135 minutes. After 135 minutes, 1  $\mu$ L PI (Sigma) was added to a final concentration of 100ng/mL in order to observe cell death.

### Two-photon intravital microscopy of the ear and image analysis

C57BL/6 LysM-EGFP knock-in animals<sup>31</sup> were obtained through the National Institute of Allergy and Infectious Diseases/Taconic exchange program. Animals were maintained in specific pathogen-free facilities and accredited by the Association for Assessment and Accreditation of Laboratory Animal Care. All animal procedures were carried out under approved animal study protocols in accordance with National Institute of Allergy and Infectious Diseases Animal Care and Use Committee guidelines. Mice were anesthetized using isoflurane vaporized in an 80:20% mixture of O<sub>2</sub> and air (2.5% for induction and 1-1.5% for maintenance). Microplastics (1.1  $\mu$ m yellow-green or red fluorescent polystyrene microspheres from Invitrogen Molecular Probes) were injected intradermally in the ear pinna using a 31G insulin syringe. Next the ear pinna was prepared for imaging on a custom made imaging platform as described in detail elsewhere.<sup>32,33</sup> The images were acquired on an inverted LSM 510 NLO system (Carl Zeiss Microimaging) equipped with a Chameleon XR Ti:Sapphire laser (Coherent), and either a 20x/0.8 NA or 25x/0.8 NA air objective. An incubation chamber was used to maintain a temperature of 32°C. In order to avoid hair follicles and melanin autofluorescence, images were captured towards the anterior half of the ear pinna. Z stacks with 3- $\mu$ m step size were acquired every 30 s. External non-descanned PMT detectors were used to collect the fluorescence. Data were processed with Image J software. Twelve videos from 5 different mice were analyzed.

### Bacterial containment assays

The procedure of the killing assay has been described in detail by van Grinsven et al.<sup>34</sup> The killing capacity of neutrophils was tested upon pre-incubation of plain 792 nm polystyrene microplastics (Table 2, PPS-0.8) as described above. Thereafter neutrophils were isolated according to the method described in *neutrophil isolation*. The culture was performed in a 100  $\mu$ L fibrin gel in a 96-well plate (black, clear bottom; Corning Life Sciences) at 37°C, the concentration of neutrophils in each well was 5\*10<sup>6</sup> cells/ml. The fibrin gel contained fibrinogen (2mg/ml, FIB3 free of Plg, vWF & Fibronectin; Stago), thrombin (10U/ml; Sigma-Aldrich), 40% human AB serum (H4522; Sigma-Aldrich). Control neutrophils without microplastics were incubated in the same manner. The bacteria were added in the gel with a MOI of 1. A control without bacteria was also added. There was also a control without any neutrophils to measure background signal of the gel. The fluorescent signal of MW2-GFP outgrowth was measured in a FLUOStar Optima microplate reader (BMG Labtech) every 20 minutes for 83 hours. From the raw data the first derivative of all time points was

calculated with Graphpad Prism, in arbitrary fluorescence units/20 min. The growth curve without neutrophils was used to determine what derivative correlated with unrestricted outgrowth. When the derivative was above this threshold for three time points in a row, the first of the three time points was defined as the end of the lag time.

## Statistics

For statistics a paired one-way repeated-measures ANOVA test was applied (GraphPad Prism 6.0). Results were regarded as significant when  $p < 0.05$ . P values were depicted in the legends of the corresponding graphs.

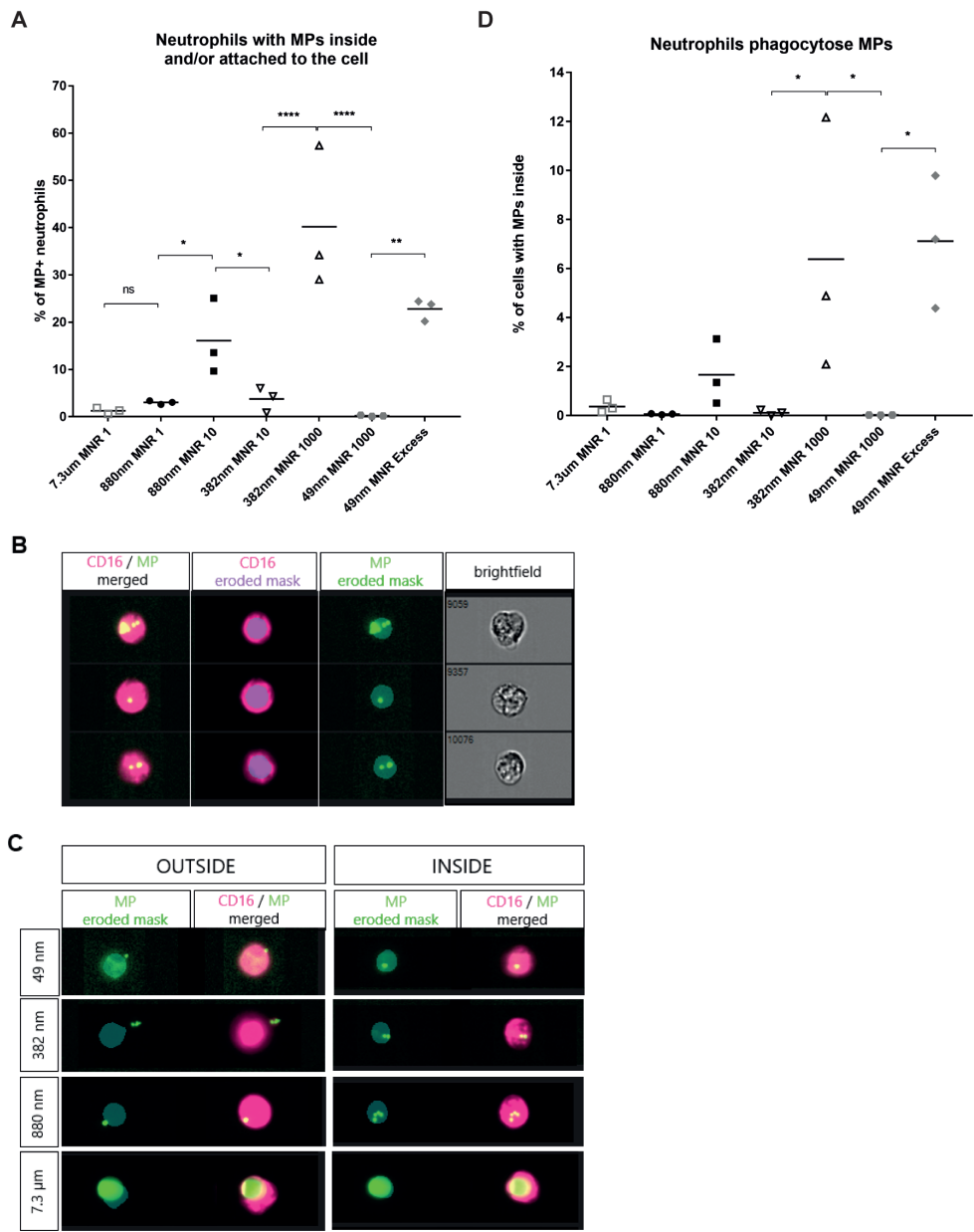
## Results

### Human peripheral blood neutrophils phagocytose microplastics in vitro

First we tested whether human neutrophils could interact with and phagocytose microplastics (MPs). Fluorescent microplastics of various sizes were added to whole blood in different microplastic to neutrophil ratios (MNRs) and shaken at 37°C for 45 minutes. Neutrophils were identified by staining with a CD16 antibody. With the ImageStream flow cytometer the fluorescence of every cell was determined combined with the generation of an image. First the percentage of microplastic positive neutrophils (Figure 1A) was measured. This percentage consisted of neutrophils which were bound to and/or phagocytosed microplastics. The higher the concentration of microplastics, the higher percentage of microplastic positive neutrophils was found. Interestingly, only a small percentage (3%) of neutrophils were interacting with the 7.3  $\mu\text{m}$  particles at an MNR of 1 after 45 min of co-incubation.

Next, we determined the number of neutrophils that phagocytosed microplastics. To objectively quantify the percentage of cells with MPs inside, a cellular mask was generated in IDEAS software which was subsequently eroded by 20% (Figure 1 B+C). Only MP+ events within this conservative mask were considered cells that phagocytosed particles. Next, the percentage of neutrophils which phagocytosed microplastics was calculated (Figure 1D). Unfortunately, the 20% eroded mask was too conservative for the 7.3  $\mu\text{m}$  MP which were, therefore, manually inspected and calculated, a mean of 0.4% was found (ranging from 0.4-0.7%). Comparing MP sizes with similar MNRs, a trend was found that higher percentages of neutrophils phagocytosed the 880 nm MPs (with a mean of 1.7%, ranging from 0.5-3.1%) than the 382 nm MPs (with a mean of 0.1%, ranging from 0.003-0.2%) (MNR 10). Due to the wide range of percentages between the samples, these results were not significantly different. The 382 nm were significantly more efficiently phagocytosed (with

a mean of 6.4%, ranging from 4.9-12.2%) than the 49 nm particles (with a mean of 0.025%, ranging from 0.023-0.029%,  $p = 0.01$ ) (MNR 1000). Overall, the phagocytosis of the MPs at lower concentrations seemed rather low.

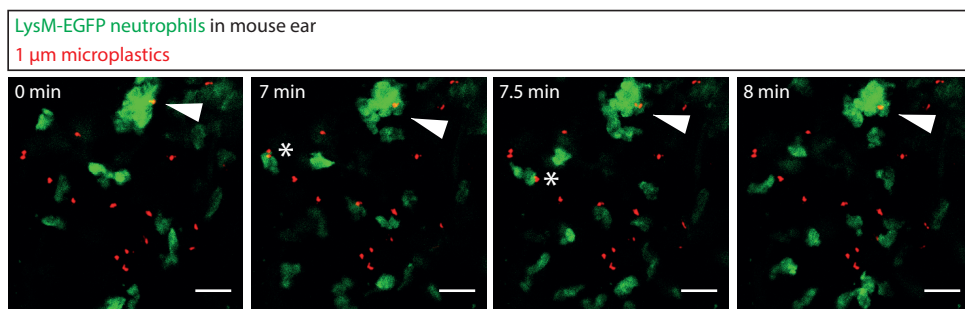


**Figure 1. Human peripheral blood neutrophils phagocytose microplastics in vitro.**

A) CD16-stained neutrophils (pink) were measured on the ImageStream after 45 min incubation with green-fluorescent microplastics. The percentage of MP positive neutrophils per MP size (either attached to the cell surface or inside the cell) was measured. B) ImageStream images of MP+ neutrophils. The first column shows the ImageStream image with both CD16-stain and green fluorescent MPs; 2<sup>nd</sup> column: To determine whether the MPs were attached to or inside the cell, an eroded cellular mask was generated using the IDEAS software (in purple); 3<sup>rd</sup> column: image of the eroded mask with microplastics inside; 4<sup>th</sup> column: brightfield image. C) Representative Figures of MPs attached to or inside the cell for every MP size. CD16 stain in pink; MPs in green. D) The percentage of neutrophils which phagocytosed particles was calculated using the eroded cellular mask and depicted in this graph. Only the phagocytosis of the 7.3  $\mu$ m particles was manually quantified. MNR = microplastic to neutrophil ratio; ns = non-significant; \*  $p < 0.05$ ; \*\*  $p = 0.0025$ ; \*\*\*\*  $p < 0.0001$ .

**Neutrophil phagocytosis of microplastics in the mouse in vivo**

We made an initial attempt to monitor microplastic phagocytosis in living mice using 2-photon intravital microscopy. To this end 1  $\mu$ m polystyrene particles were intradermally injected in the ear of LysMEGFP mice which have green fluorescent neutrophils. Neutrophils which were recruited to the site of injection were often found in close proximity to the beads. In line with our in vitro results, which showed that on average only 0.4% of the neutrophils phagocytosed the 7.3  $\mu$ m MPs after 45 min (as described above and shown in Figure 1D), occasional phagocytosis could be observed, but microplastics were predominantly ignored by neutrophils (Figure 2 and Supplemental Video 1).

**Figure 2. Neutrophils can phagocytose microplastics in living mice.**

Fluorescent 1  $\mu$ m microplastics (red) were intradermally injected in the ears of LysM-EGFP mice expressing GFP in neutrophils (green). Neutrophils are recruited to the site of injection. Interactions between neutrophils and microplastics occasionally lead to phagocytosis (arrow) but often microplastics were ignored (\*). Representative example of 12 movies from 5 mice. Scale bar represents 20  $\mu$ m.

### Interaction of human neutrophils with microplastics leads to cell death

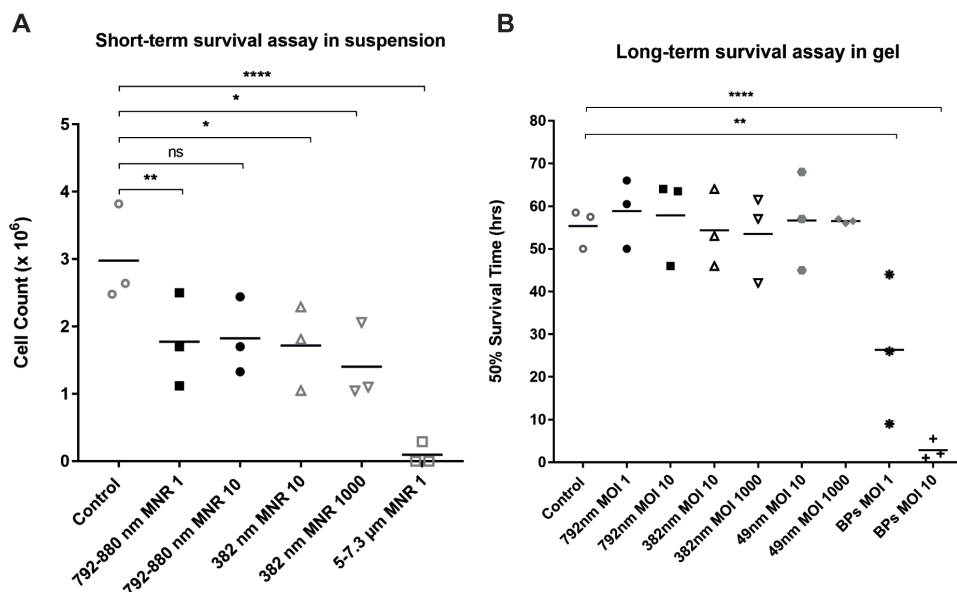
Next we aimed to investigate the effect of interactions between neutrophils and microplastics on neutrophil viability. Initially  $3 \times 10^6$  cells were incubated with different sizes of microplastics. After the 45 min incubation period we manually counted the cells which were left (Figure 3A). When compared to the microplastic free control sample, incubation with the smaller sized particles lead to a decrease in cell number. From the control samples a mean of  $2.64 \times 10^6$  cells were found after incubation, while a mean of  $1.8 \times 10^6$  cells for the 880 nm beads at an MNR of 1 and a mean of  $1.7 \times 10^6$  cells for the 382 nm beads MNR of 10 were recovered. Interestingly, a 10-fold or even a 100-fold increase in microplastics did not result in increased cell death compared to the lower concentrations of microplastics, as incubation of neutrophils with 880 nm MPs at an MNR of 10 resulted in a mean cell count of  $1.8 \times 10^6$  cells and incubation with 382 nm beads at an MNR of 1000 of  $1.4 \times 10^6$  cells. When neutrophils were incubated with the relatively big  $7.3 \mu\text{m}$  particles only a mean of  $0.097 \times 10^6$  cells (ranging from 0 to  $0.29 \times 10^6$ ) was recovered (Figure 3A).

The low percentage of phagocytosed  $7.3 \mu\text{m}$  MPs in Figure 1D might be an underestimation considering the fast death of neutrophils found in Figure 3A. In order to investigate whether or not  $7.3 \mu\text{m}$  beads were phagocytosed before neutrophil death, live time lapse microscopy was performed. The green fluorescent beads were additionally labelled with a pH sensitive pHrodo dye which increases red fluorescence in low pH in the phagolysosome. Phagocytosis of the  $7.3 \mu\text{m}$  beads was readily observed and even multiple beads could be phagocytosed by a single neutrophil (Figure 4A and Suppl video 2). An increase in pHrodo signal could be detected upon phagocytosis (Figure 4B and Suppl video 2). Later in the movie neutrophils appeared to die releasing the bead and losing the pHrodo signal (Figure 4B and Suppl video 2). Adding propidium iodide at the end of the experiment indeed confirmed many neutrophils died (Figure 4C).

Next we aimed to investigate the impact of the smaller particles on long-term neutrophil survival. After the initial cell death after incubation with MPs (as described above and depicted in Figure 3A), the surviving pre-incubated fluorescent neutrophils were transferred into a three-dimensional fibrin gel mimicking a more physiological environment and improving cell survival.<sup>34</sup> The time point at which 50% of the neutrophils which were transferred to the gel died was measured. We excluded the  $7.3 \mu\text{m}$  MPs from analysis, as there were only a few cells left after the initial 45 min pre-incubation, making the results unreliable. In this extended survival assay we did not find a difference in the 50% survival time between the microplastic free control (mean of 55 hr) and the 792 (MNR1 59 hr; MNR 10 58 hr), 382 (MNR1 and 10 both 54 hr) and 49 nm MPs (MNR 10 and MNR 1000 57 hr) (Figure 3B, Table 3). Similar as found for the short-term survival assay, also for the long-term survival assay a 10-fold or even 100-fold increase in the MP concentration

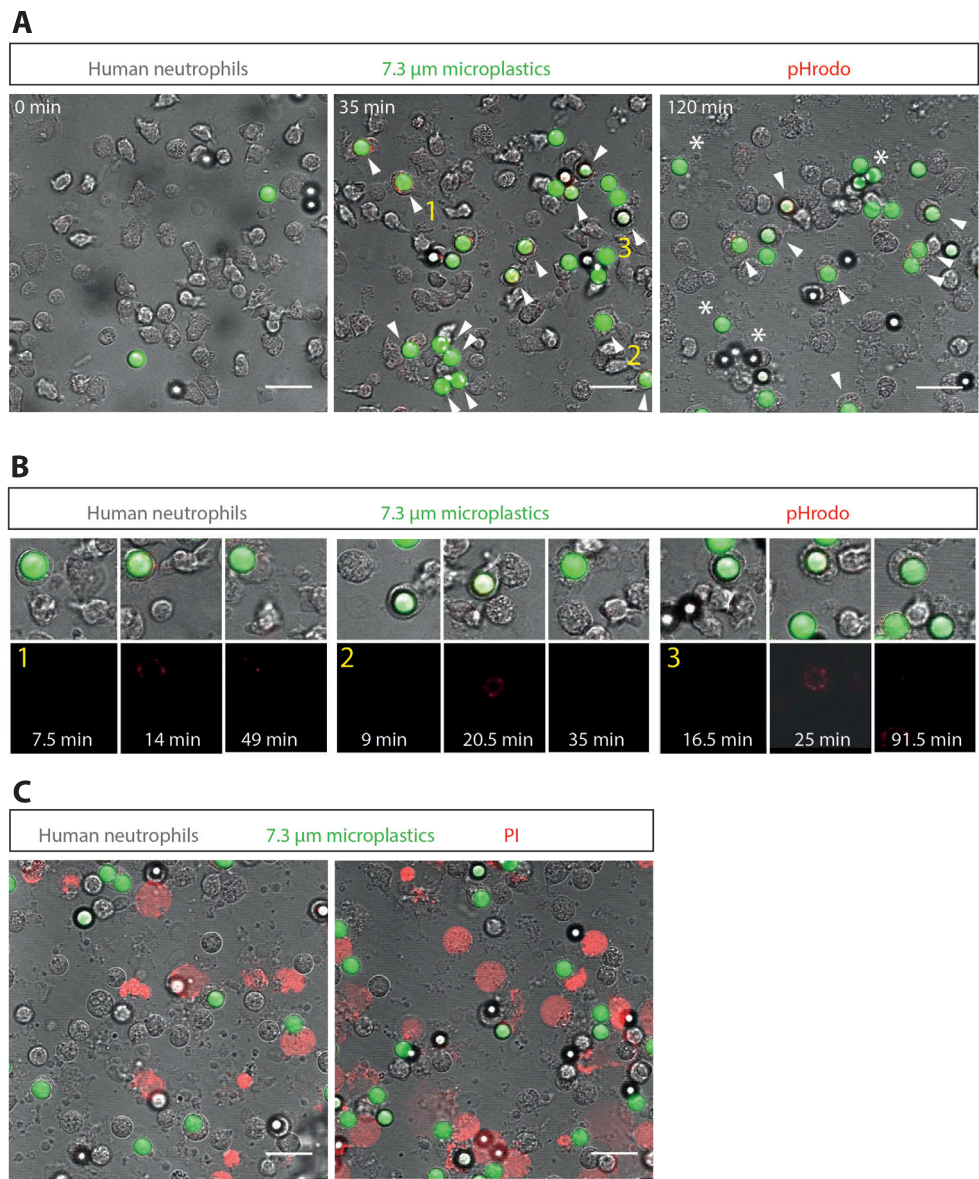


did not change the 50% survival time (Figure 3B, Table 3). For the smaller particles this might mean that more delicate neutrophils die quickly, leaving a sample of relatively resilient neutrophils. Incubation with killed bacteria, also termed bioparticles (BPs), led to a significantly shorter 50% survival time (26 hr for a multiplicity of infection (MOI) of 1; and 3 hr for an MOI of 10) than control neutrophils which were not pre-incubated with BPs/MPs and neutrophils that were pre-incubated with 49-792 nm MPs. This demonstrates the assay was capable of measuring long term survival differences.



**Figure 3. Incubation of neutrophils with microplastics leads to rapid cell death.**

A) Fluorescent neutrophils were incubated with various sizes of non-fluorescent microplastics (MPs), or without MPs (Control) at 37°C whilst shaking. After 45 minutes, the total number of neutrophils was counted manually. B) Fluorescent neutrophils which survived the co-culture with MPs for 45 min at 37°C were transferred to a fibrin-thrombin gel (which mimics tissue structure) and the 50% survival time was estimated using a fluorescence plate reader. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*\*  $p < 0.0001$ ; ns = non-significant; MNR = microplastic to neutrophil ratio; BP = bioparticles; MOI = multiplicity of infection



**Figure 4. Human neutrophils phagocytose 7.3  $\mu\text{m}$  microplastics.**

A) Fluorescent 7.3  $\mu\text{m}$  beads were labelled with pHrodo and added to isolated human neutrophils. Timelapse microscopy images are shown. Microplastics phagocytosed by neutrophils are shown by arrows and \* depicts neutrophil death. B) The signal of pHrodo is shown in detail for 3 cells marked 1,2,3 in A. Images right after phagocytosis, at the first timepoint of acidification and at the moment acidification disappears are shown. C) After the time lapse PI was added to demonstrate neutrophil death in two positions of the same sample. Scale bar represents 20  $\mu\text{m}$ .

**Table 3. Neutrophil survival assay**

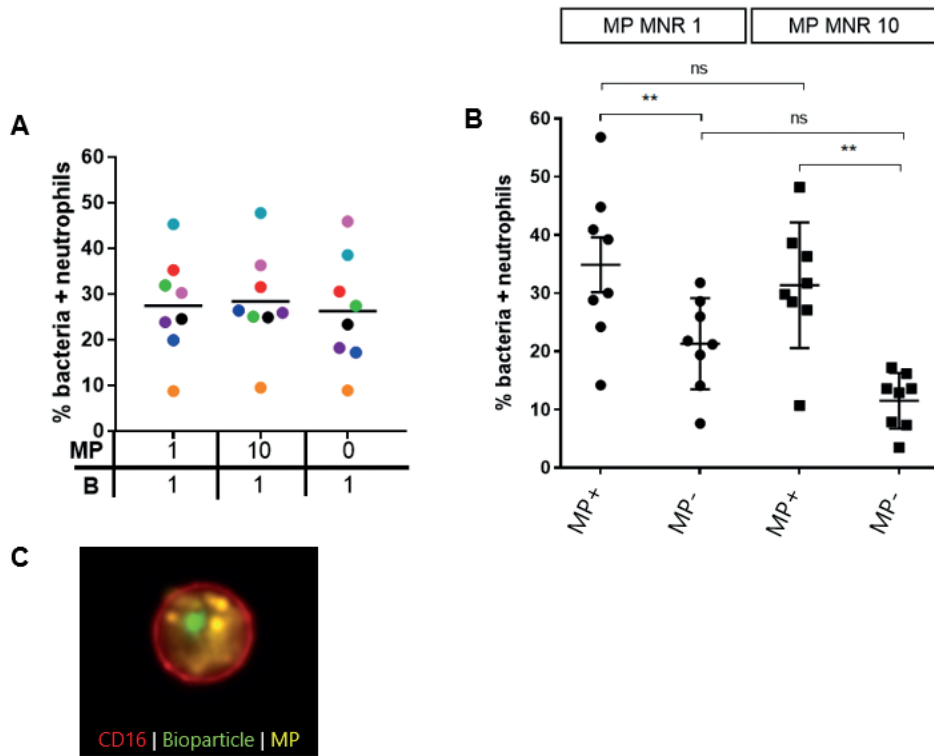
	Control	792nm MNR 1	792nm MNR 10	382nm MNR 10	382nm MNR 1000	49nm MNR 10	49nm MNR 1000	BPs MNR 1	BPs MNR 10
Mean 50% survival time (hr)	55	59	58	54	54	57	57	26	3
range (hr)	50-59	50-66	46-64	46-53	42-62	45-68	56-57	9-44	1-6
p value (ctrl)	-	ns	ns	ns	ns	ns	ns	0.0031	0.0001

*Mean 50% survival time, the range between the survival times (n=3) and the significance of the difference found between a particular group of samples and the control measurements.*

**Microplastic phagocytosis by neutrophils does not impair the subsequent phagocytosis of bacteria**

Next, we hypothesized that the neutrophils which phagocytosed microplastics might not be very capable of phagocytosing bacteria because neutrophils are not able to degrade plastic particles,. For this experiment we used the 880 nm particles, as those were best phagocytosed at a low MNR of 10 (as was shown in Figure 1A). Neutrophils were incubated without (MNR 0) or with different concentrations (MNR 1 or 10) of fluorescent 880 nm microplastics for 45 min. Hereafter, fluorescent bacteria were added to the neutrophils with an MOI of 1. Bacteria positive neutrophils were measured on a flow cytometer. The percentage of bacteria positive neutrophils was similar between the control and samples with microplastics, thus the presence of microplastics did not influence the association of neutrophils to bacteria (Figure 5A). Notably, this result only shows the effect of MPs on bacterial phagocytosis of the surviving neutrophils, while the initial significant cells death was not taken into account in this assay.

We wondered whether the bacteria were preferably phagocytosed by neutrophils which did not interact with MPs. Interestingly, by comparing MP+ and MP- neutrophils we found that of the MP+ neutrophils a higher percentage was also bacteria positive, compared to the MP- neutrophils (Figure 5B+C). This means that neutrophils that interact with microplastics are more prone to associate with bacteria than MP-negative neutrophils. This is in line with the findings of Hellebrekers et al<sup>35</sup> and another indication that the neutrophil compartment partly consists of good ‘phagocytosers’ and of cells which seem to be less programmed to phagocytose.



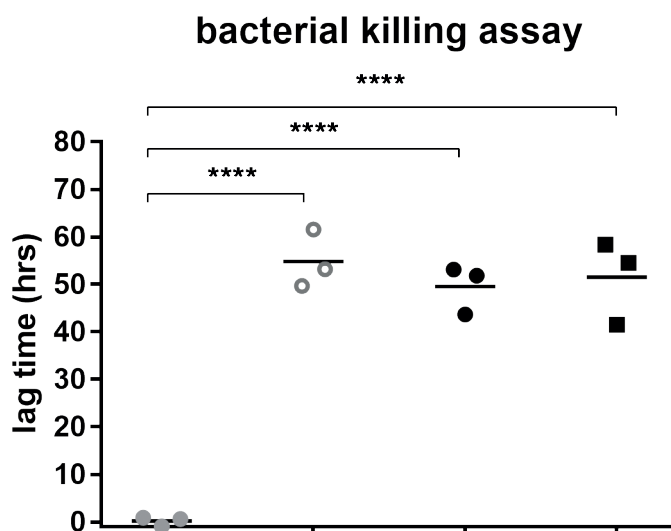
**Figure 5. Microplastic phagocytosis by neutrophils does not impair the subsequent phagocytosis of bacteria.**

A) Neutrophils were incubated without (MNR 0) or with different concentrations (MNR 1 or 10) of fluorescent 880 nm microplastics for 45 min at 37°C. After that, differently stained fluorescent bacteria were added to the neutrophils with an MOI of 1, as shown on the x axis. The percentage of bacteria positive neutrophils were measured on a flow cytometer and depicted here, no significant differences were found; B) Graph showing the same data as in A), but now the neutrophils incubated with an MP MNR of 1 and 10 were separated into MP+ and MP- neutrophils on the x-axis, and the % of bacteria positive neutrophils is shown on the y-axis. C) Microscopic image of a neutrophil (red fluorescent cell membrane) which has phagocytosed yellow fluorescent microplastic beads and green fluorescent bacteria, measured on the ImageStream. \*\* < 0.01; ns = not significant.

### Neutrophil bacterial containment was not influenced by the pre-incubation with microplastics

To study the bacterial containment by neutrophils, we performed a bacterial killing assay<sup>34</sup> (Figure 6). First, whole blood was incubated with (MNR 1; MNR 10) or without microplastics. Subsequently, the pre-incubated neutrophils were recounted and transferred to a fibrin-thrombin gel and fluorescent bacteria with an MOI of 1 were added to the neutrophils. Bacterial outgrowth was measured by measuring the fluorescence signal. The lag time, a

quantitative measure of relative antibacterial capacity of the neutrophils, was calculated. This value defines the time period until the fluorescent signal starts to show an exponential increase, i.e. the time until bacterial outgrowth.<sup>34</sup> Without neutrophils, bacteria grew out immediately. When neutrophils were incubated without microplastics a lag time of 50-62 hrs was measured. When neutrophils were pre-incubated with microplastics (880 nm, MOI 1 or MOI 10) a similar containment was shown by demonstrating that the cells were able to prevent the bacteria from growing out for 42-62 hrs.



**Figure 6. Pre-incubation with microplastics does not affect the bacterial containment capacity.**

A) whole blood was either incubated with (MOI 1; MOI 10) or without microplastics. Subsequently, the pre-incubated neutrophils were transferred to a fibrin-thrombin gel and fluorescent bacteria with an MOI of 1 were added to the neutrophils. The 'lag time' served as a quantitative measure of relative antibacterial capacity of the neutrophils. This value defines the time period until the fluorescent signal starts to show an exponential increase, i.e. the time until bacterial outgrowth.<sup>34</sup> Different conditions were tested: lag time without neutrophils (grey filled circles), after 45 min pre-incubation without plastic (grey circles, no MP) and after pre-incubation with microplastics (880 nm, MOI 1 or MOI 10, black filled circles and black filled squares, respectively). Every data point represents the mean of a duplo measurement of a healthy donor.

## Discussion

Recently increased attention is given to plastic pollution of our oceans and microplastics in the aquatic environment. Although numerous studies have focused on the effects of microplastics on aquatic invertebrates and fish, less is known about the effects on human health. In this study we tested the hypothesis that human neutrophils can interact with MPs and this interaction can affect neutrophil function. Our study came to the conclusion that exposure to microplastics can lead to neutrophil cell death possibly in a subpopulation more sensitive to these beads. This microplastic-induced death was a fast process (within 45 min 40-53% of neutrophils died when co-incubated with 49-880 nm MPs). Especially the larger particles of 7.3  $\mu\text{m}$  had a fatal effect on neutrophils when co-incubated (approximately 97% cell death).

### Neutrophil death after phagocytosis of microplastics – differences between MP sizes

Many neutrophils died during incubation with 7.3  $\mu\text{m}$  MPs and this caused a lower percentage of neutrophil phagocytosis. This could either mean that microplastic phagocytosis leads to cell death or that the beads were too big to be phagocytosed and caused death by “frustrated phagocytosis”.<sup>36,37</sup> In order to distinguish between these two mechanisms, time lapse confocal microscopy was used to visualize the behaviour of neutrophils in the presence of 7.3  $\mu\text{m}$  MPs. Interestingly, many neutrophils interacted with and phagocytosed the 7.3  $\mu\text{m}$  MPs, but died shortly after. This was further supported by the pHrodo label present on the surface of the MP's, which demonstrated acidification of the microplastic surface after phagocytosis. Before drawing solid conclusions, however, we planned on repeating the confocal experiment with different sizes of MPs and with neutrophils from different donors, in order to substantiate and quantify our findings. The fact that we found neutrophils with 2 or more 7.3  $\mu\text{m}$  beads inside the cell argued against frustrated phagocytosis, as apparently these cells are quite capable to ingest the whole particle. Nonetheless, neutrophil death might be caused by the fact that intracellularly the cells cannot deal with (such large) plastic particles and thus die, i.e. plastic-induced cell death. It remains unclear why the cells die. Are the plastics directly toxic to neutrophils upon internalisation? Or is the cell death evolutionary programmed as a last resort for neutrophils to destroy more resistant microbes, cells, organic materials or any other type of particles which are not degraded in the phagolysosome? An internal ‘suicide attack’ targeted on the ingested particle, which, unfortunately, leaves microplastics intact. The recorded neutrophil death in the time lapse movie (Suppl Video 2) and the images of the PI results (Figure 4C) reminded us of NETosis, which has been reported as a process in which neutrophils release neutrophil extracellular traps (NETs) to kill a variety of microbes.<sup>39</sup> Increased NETosis was found for microbes which are too big for neutrophils

to phagocytose, compared to smaller microbes.<sup>40</sup> In contrast to NETosis, here neutrophils succeeded in phagocytosing the large MPs, and the plastics-induced cell death was visible within 1 hour, while NETosis has been reported as a slower process, which takes approximately 4 hours to be completed.<sup>40</sup>

Interaction with the smaller MPs resulted in much less neutrophil death in the first 45 min, even at higher MNRs, although it was still significantly increased compared to the control. For the smaller MPs higher percentages of phagocytosis were found at 45 min, however higher microplastic to neutrophil ratios were used here. The question remains whether also during incubation with the smaller beads the cells die upon phagocytosis of the particles. However, phagocytosis of smaller particles is a slower process than for the large particles, leading to less cell death at 45 min. Alternatively, a different mechanism might be causing cell death. Therefore, it is important to also study the phagocytosis of the smaller particles real-time using confocal microscopy.

### **Neutrophil death after phagocytosis of microplastics – differences between experimental conditions**

Comparing the *in vitro* results obtained with confocal microscopy of the 7.3  $\mu\text{m}$  particles to the *in vivo* results of phagocytosis of 1  $\mu\text{m}$  MPs in the mouse ear, clear differences were found. While the *in vitro* experiment showed massive phagocytosis of the 7.3  $\mu\text{m}$  MPs followed by cell death, in the *in vivo* experiment the 1  $\mu\text{m}$  MPs were mostly ignored by the neutrophils and cell death was difficult to objectify. These discrepancies could be explained by differences in the experimental procedure.

Firstly, the MPs had different sizes, which might explain the difference in results. Interestingly, regarding size we would have expected the 1  $\mu\text{m}$  particles to be more efficiently phagocytosed, as this size resembles the size of bacteria, which are the natural neutrophil target. However, size was not the only difference here.

Secondly, the timeframe of the experiments differed considerably. The time lapse of the confocal experiment lasted 2.25 hours, while the time lapse of the intra-vital experiment were recorded 10–45 min. And although we do not think a many more neutrophils will start phagocytosis after 45 min, we might have been unable to record plastic-induced cell death in the short time frame of the intra-vital experiment.

Thirdly, apart from difference in size of the particles and timing of the experiment, the environment is also important, as proteins in the environment bind to microplastics. Prior to incubation of MPs with neutrophils *in vitro*, the MPs were opsonized in platelet rich plasma which contains proteins and platelets. The 1  $\mu\text{m}$  particles were not opsonized prior to the *in vivo* experiment in the mouse ear. Proteins form a corona around the surface of



the particles. The composition of the corona is dependent on the size and surface charge of the particle.<sup>41</sup> The corona consists of biomolecules which rapidly exchange between the medium and the particle and of macromolecules which are attached to the particle surface with a high affinity. After incubation in serum, the corona of microplastics consists of immunoglobulins (IgG), apolipoproteins, complement proteins, acute-phase proteins and coagulation proteins.<sup>41</sup> Receptors on cells can bind to proteins of the corona. This means that the environment and thus the corona affects the attachment and uptake of the microplastics by cells. We performed a confocal experiment with the 7.3  $\mu\text{m}$  particles which were not incubated in PRP prior to the experiment. This led to much less phagocytosis (data not shown). From these results we can conclude that without a corona, microplastics are not readily phagocytosed. For the *in vivo* experiments we aimed to monitor phagocytosis under physiological corona formation. However, it would be interesting to compare natural corona formation with pre-opsonized particles using the same conditions as our *in vitro* experiments. In addition, it would be of interest to study what type of corona is formed on environmental microplastics, whether toxic molecules bind to plastics, whether the corona from the aquatic environment is beneficial for uptake in the intestines, and also whether the corona changes in the human/mouse organs or blood, and how long this takes.

### **Larger MPs are more fatal to human neutrophils than smaller MPs and an increase in microplastic concentration does not affect neutrophil survival for the smaller MPs**

Counterintuitively, we found that larger plastic particles are more problematic for neutrophils than small plastics, as more cells died during incubation with the 7.3  $\mu\text{m}$  MPs. The potential increased danger which has been assigned to the smaller particles was probably based on the reports showing that these smaller particles were readily taken up by our intestine wall.<sup>42,43</sup> In addition, small particles of 20 nm in size were passively taken up by phagocytes<sup>44</sup> and were found to pass the blood-brain-barrier. A last important argument relates to the fact that the smaller particles are even more prevalent in our environment than larger particles. However, from our experiments we can conclude that for phagocytosis of smaller MPs a higher concentration is required compared to the 7.3  $\mu\text{m}$  MPs. And surprisingly, higher concentrations of the smaller MPs did not lead to more cell death, as measured after 45 min of incubation by flow cytometry. Even a 10- to a 100-fold increase in MP concentration did not increase neutrophil death. This is quite striking, as we did find a higher phagocytosis rate when increasing the MP dose. There are two possible explanations for these findings. First, it might indicate that neutrophil cell death as measured by flow cytometry after 45 min of incubation with the smaller sized particles is caused by MP phagocytosis. A subpopulation of neutrophils (+/- 50% of the total population) is more sensitive and dies quickly, already with low MP concentrations.



Adding higher concentrations of microplastics does not make a difference, as all the delicate cells died already. On the other hand it could be explained by the fact that for the flow cytometry experiments we centrifuged the sample of cells and microplastics after co-incubation. Maybe under these high g forces the microplastics spin down faster than cells and thereby damage part of the cells in suspension, leading to an exaggerated death rate.

The surviving neutrophils, of which a small proportion (<8%) has internalised MPs, remain unaffected regarding phagocytosis of bacteria, bacterial containment and neutrophil life span *in vitro*. It is worth noting that these conclusions cannot be drawn about the total neutrophil population, as in the aforementioned experiments the initial cell death was not taken into account. Especially for the 7.3 µm neutrophils the massive cells death must affect the neutrophil killing capacity. Therefore, it is hard to translate these results to *in vivo* human susceptibility to infections, as the initial cell death might play an important role here.

Surprisingly, our experiments indicated that the surviving neutrophils which interacted with microplastics were more likely to phagocytose bacteria than the microplastic-free neutrophils. Recently Hellebrekers et al obtained similar results upon sequential phagocytosis with bacteria. This has led to the conclusion that certain neutrophils are more potent “phagocytosers” than others.<sup>35</sup> Importantly, after the initial cell death of the more delicate neutrophils, the surviving neutrophils were as potent in containing bacteria as the control samples were (comprising of a mix of delicate neutrophils and potent “phagocytosers”). This means that the loss of delicate cells does not affect the bacterial containment, as the potent “phagocytosers” contain the bacteria very well.

## Limitations of the study and future experiments

This study mainly serves as a proof of principle that neutrophils phagocytose microplastics, and this leads to neutrophil death. However the study has some important limitations, which we should take into account when designing future experiments.

- i. Most experiments were performed with low number of donors ( $n$ ). To be able to draw solid conclusions, and increase the power of the study, the experiments have to be repeated.
- ii. Spinning down cells with microplastics might lead to cell death, as the microplastics may destroy neutrophils under such high  $g$  forces. Therefore, future *in vitro* experiments should not include the centrifugation of cells and microplastics together in one sample.
- iii. For the smaller particles (49-782 nm) it remains unknown what happens in the first 45 min regarding phagocytosis and whether the phagocytosis leads to acute cell death, as was found for the 7.3  $\mu\text{m}$  particles. Therefore, future experiments should include time lapse confocal microscopy of the smaller sized particles.
- iv. In our *in vitro* bacterial killing assay we corrected for the initial cell death (as a certain MOI of bacteria was added to the surviving neutrophils per experimental condition, and not a fixed absolute number for all experimental conditions). Although it is good to know that the surviving neutrophils remain unaffected regarding their bacterial killing capacity and their survival, it would also be interesting to find out what the effect of the initial cell death will be on the bacterial killing capacity. Are the surviving neutrophils able to compensate for the initial cell death and capable to delay bacterial outgrowth? Therefore future killing assays should be conducted with the surviving neutrophils, and bacteria should be added at an MOI of 1 relatively to the neutrophils count we originally started with.
- v. As mentioned earlier, the *in vitro* experiments are hard to translate to the real life situation. First, we plan on repeating the intravital imaging experiment. Microplastics of different sizes will be injected in the mouse ear, with and without opsonisation and the phagocytosis of MPs will be examined.
- vi. Another *in vivo* experiment which might give more insight in the health effects of orally ingested microplastics, could comprise of feeding mice MPs. To guarantee microplastic uptake by the mice, preferably the mice food pellets should be spiked with different sizes of various fluorescent microplastics. The mice will be fed high doses of microplastics on a daily basis and a variety of "clinical" parameters will be studied, for instance:
  - a. Challenge the mice with bacteria at different time points to examine their susceptibility to infections
  - b. Monitor the bodyweight of mice which indicates whether they develop a failure to thrive
  - c. Study whether mice which were fed microplastics die earlier than mice with a microplastic-free diet.

- d. Stop feeding the mice microplastics after a long period of MP feeding and monitor the presence of microplastics in stool and urine, to find out whether and how microplastics are cleared from the body
- e. At the end of the experiment (when mice die or are sacrificed), cryosections of organs can be analysed using microscopy and single cell suspensions of organs, blood and bone marrow can be analysed for the presence of fluorescent particles using flow cytometry.

Comparing studies is difficult, as next to the wide size range of particles studied, also many different types of particles have been used ranging from carboxyl polystyrene and  $\text{TiO}_2$  to PEGylated gold particles. The combination of size, surface charge and hydrophilicity all contribute to uptake affinity (as discussed in <sup>45</sup>), and are of key importance to define the risk of different types of microplastics for (human) tissues.

In conclusion, it remains difficult to translate our findings to human health. However, our results indicate that microplastics might form a potential health hazard. A recent report calculated that a top shell fish consumer from Europe may consume 11,000 microplastic particles per year,<sup>46</sup> and this is only one of the many sources of microplastics entering our gastro-intestinal tract. Provided that these microplastics pass the intestinal wall and neutrophils die upon phagocytosis, the question remains whether the effects will be clinically relevant, as our blood contains 4,300 – 10,000 neutrophils per microliter. After subsequent release from a dying neutrophil, it would be interesting to know whether microplastics will be cleared from our body through bile fluid (as has been reported for nanoparticles in nanomedicine,<sup>47</sup> or whether ingested microplastic will continue to “infect” and kill our neutrophils. Therefore, future experiments should compare beads of different sizes, materials, shapes and corona’s in the *in vitro* and *in vivo* microscopy experiments. Also it would be of great interest to perform long exposure *in vivo* animal experiments, to be able to address the question whether and to which extent microplastics affect our health.

## References

1. PlasticsEurope. China Leads Global Plastics Production while Europe Ranks Second [press release]. Brussels, Belgium:PlasticsEurope AISBL (12 June 2014).
2. Seltnerich N. New Link in the Food Chain? Marine Plastic Pollution and Seafood Safety. *Environmental Health Perspectives*. 2015;123(2):A34-A41. doi:10.1289/ehp.123-A34.
3. Alimi, O.S., Farner Budarz, J., Hernandez, L.M., Tufenkji, N. and Nanoplastics in Aquatic Environments: Aggregation, Deposition, and Enhanced Contaminant Transport (2018) *Environmental Science and Technology*, 52 (4), pp. 1704-1724.
4. Thompson R.C.; Olsen Y.; Mitchell R.P.; Davis A.; Rowland S.J.; John A.W.G.; et al. Lost at sea: where is all the plastic? 2004, . 304: 838–840. doi: 10.1126/science.1094559
5. Schroeck, N. J. Microplastic pollution in the Great lakes: State, federal, and common law solutions. *University of Detroit Mercy Law Review* 2016, 273–291
6. Chan K. Hong Kong plastic pellets spill: Sinopec vows to clean up beaches. *Huffington Post, Green section* (9 August 2012)
7. Eriksen, M.; Lebreton, L. C.; Carson, H. S.; Thiel, M.; Moore, C. J.; Borerro, J. C.; Galgani, F.; Ryan, P. G.; Reisser, J. Plastic pollution in the world's oceans: More than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea. *PLoS One* 2014, 9 (12), e111913
8. Koelmans, A.A., Besseling E., Shim W.J., (2015) Nanoplastics in the Aquatic Environment. *Critical Review*. In: Bergmann M., Gutow L., Klages M. (eds) *Marine Anthropogenic Litter*. Springer, Cham. DOI: [https://doi.org/10.1007/978-3-319-16510-3\\_12](https://doi.org/10.1007/978-3-319-16510-3_12) Publisher Name Springer, Cham Print ISBN: 978-3-319-16509-7 Online ISBN: 978-3-319-16510-3
9. Duforêt-Gaurier L., Moutier W., Guiselin N., Thyssen M., Dubelaar G., Mériaux X., Courcot L., Dessailly D., Loisel H. (2015) Determination of backscattering cross section of individual particles from cytometric measurements: a new methodology, *Opt. Express*. 23, 31510-31533
10. Zada L. Fast microplastics identification with stimulated Raman scattering microscopy. 2018; 1136–1144. doi:10.1002/jrs.5367
11. Tinnevelt GH, Kokla M, Hilvering B, Van Staveren S, Folcarelli R, Xue L, et al. Novel data analysis method for multicolour flow cytometry links variability of multiple markers on single cells to a clinical phenotype. *Sci Rep*. 2017;7. doi:10.1038/s41598-017-05714-1
12. M.K. Thomas, S. Fontana, M. Reyes, F. Pomati, Quantifying cell densities and biovolumes of phytoplankton communities and functional groups using scanning flow cytometry, machine learning and unsupervised clustering, *PLoS ONE* (<https://doi.org/10.1371/journal.pone.0196225>)
13. Costa, M.F., Ivar do Sul, J.A., Silva-Cavalcanti, J.S. et al. *Environ Monit Assess* (2010) 168: 299. <https://doi.org/10.1007/s10661-009-1113-4>
14. da Costa, J.P., Santos, P.S.M., Duarte, A.C., Rocha-Santos, T. (Nano)plastics in the environment - Sources, fates and effects (2016) *Science of the Total Environment*, 566-567, pp. 15-26. DOI: 10.1016/j.scitotenv.2016.05.041

15. Cole M., Lindeque P., Fileman E., Halsband C., Goodhead R., Moger J., Galloway T.S. Microplastic ingestion by zooplankton. *Environmental Science and Technology*. 2013;47:6646–6655. doi: 10.1021/es400663f.
16. Van Cauwenburghe L., Claessens M., Vandeghechuchte M.B., Janssen C.R. Microplastics are taken up by mussels (*Mytilus edulis*) and lugworms (*Arenicola marina*) living in natural habitats. *Environmental Pollution*. 2015;199:10–7. doi: 10.1016/j.envpol.2015.01.008
17. Karlsson Therese M., Vethaak A. Dick, Almroth Bethanie Carney, Ariese Freek, van Velzen Martin, Hassellöv Martin, Leslie Heather A. (2017) Screening for microplastics in sediment, water, marine invertebrates and fish: Method development and microplastic accumulation. *Mar. Pollut. Bull.*, 122, 1–2
18. Neves, D., Sobral, P., Ferreira, J.L., Pereira, T. Ingestion of microplastics by commercial fish off the Portuguese coast(2015) *Marine Pollution Bulletin*, 101 (1), pp. 119–126.DOI: 10.1016/j.marpolbul.2015.11.008
19. Horton, A. A.; Walton, A.; Spurgeon, D. J.; Lahive, E.; Svendsen, C. Microplastics in freshwater and terrestrial environments: Evaluating the current understanding to identify the knowledge gaps and future research priorities. *Sci. Total Environ*. 2017, 586, 127–141
20. Lwanga, E.H., Vega, J.M., Quej, V.K. Chi., J.d.l.A, Cid., L.S.d., Chi., C., Segura., G.E., Gersten, H., Salánki., T., Ploeg M.v.d., Koelmans., A.A., Geissen, V. Field Evidence for Transfer of Plastic Debris Along a Terrestrial Food Chain Scientific Reports Volume 7, Article number: 14071 (2017)
21. Liebezeit G., Liebezeit E. Synthetic particles as contaminants in German beers. *Food Additives & Contaminants* 2014; 1574–1578
22. Iñiguez, M.E., Conesa, J.A., & Fullana, A. (2017). Microplastics in Spanish Table Salt. *Scientific Reports* volume 7, Article number: 8620 (2017) DOI: 10.1038/s41598-017-09128-x
23. Oßmann, B.E., Sarau G., Holtmannspötter H., Pischetsrieder M., Christiansen S.H., Dicke W., Small-sized microplastics and pigmented particles in bottled mineral water. *Water Res*. 2018 Sep 15;141:307–316. DOI: 10.1016/j.watres.2018.05.027. Epub 2018 May 16. PubMed PMID: 29803096.
24. Deng, Y., Zhang, Y., Lemos, B., & Ren, H. (2017). Tissue accumulation of microplastics in mice and biomarker responses suggest widespread health risks of exposure. *Front. Environ. Sci.*. Scientific Reports volume 7, Article number: 46687 (2017) doi: 10.3389/fenvs.2017.00063
25. Lu, L., Wan, Z., Luo, T., Fu, Z., Jin, Y. Polystyrene microplastics induce gut microbiota dysbiosis and hepatic lipid metabolism disorder in mice(2018) *Science of the Total Environment*, 631–632, pp. 449–458. Cited 2 times. DOI: 10.1016/j.scitotenv.2018.03.051
26. Cornick S., Tawiah A., Chadee K. (2015). Roles and regulation of the mucus barrier in the gut. *Tissue Barriers* 3:e982426. 10.4161/21688370.2014.982426
27. Gallin JI. Human neutrophil heterogeneity exists, but is it meaningful? *Blood*. 1984.
28. Pratten M.K., Lloyd J.B. Pinocytosis and Phagocytosis—The Effect of Size of a Particulate Substrate on Its Mode of Capture by Rat Peritoneal-Macrophages Cultured Invitro. *Biochim. Biophys. Acta*. 1986;881:307–313. doi: 10.1016/0304-4165(86)90020-6.

29. Movat H.Z., Weiser W.J., Glynn M.F., Mustard J.F. Platelet phagocytosis and aggregation. *J. Cell Biol.* 1965;27(3):531–54
30. Surewaard, B.G., de Haas, C.J., Vervoort, F., Rigby, K.M., DeLeo, F.R., Otto, M., van Strijp, J.A., Nijland, R. (2013) Staphylococcal alpha-phenol soluble modulins contribute to neutrophil lysis after phagocytosis. *Cell. Microbiol.* 15, 1427–1437.
31. Faust N., F. Varas, L. M. Kelly, S. Heck, T. Graf. 2000. Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. *Blood* 96: 719–726
32. Peters Nathan C; Egen Jackson G, Secundino Nagila, Debrabant Alain, Kimblin Nicola, Kamhawi Shaden, Lawyer Phillip, Fay Michael P, Germain Ronald N, Sacks David. (2008) In Vivo Imaging Reveals an Essential Role for Neutrophils in Leishmaniasis Transmitted by Sand Flies. *Science* 15 Aug 2008 : 970-974
33. Grinsven E. van, Prunier C., Vrisekoop N., Ritsma L. (2017) Two-Photon Intravital Microscopy Animal Preparation Protocol to Study Cellular Dynamics in Pathogenesis. In: Markaki Y., Harz H. (eds) *Light Microscopy. Methods in Molecular Biology*, vol 1563. Humana Press, New York, NY
34. Grinsven E. van, Leliefeld P.H.C., Pillay J., Aalst C.W. van, Vrisekoop N., Koenderman L. (2018) A comprehensive three-dimensional assay to assess neutrophil defense against bacteria. *Journal of Immunological Methods.* 462 : 83-90. doi: 10.1016/j.jim.2018.09.001
35. Hellebrekers P, Hietbrink F, Vrisekoop N, Leenen LPH and Koenderman L (2017) Neutrophil Functional Heterogeneity: Identification of Competitive Phagocytosis. *Front. Immunol.* 8:1498. doi: 10.3389/fimmu.2017.01498
36. Rosenkranz, A. R., Kormoczi, G. F., Thalhammer, F., Menzel, E. J., Horl, W. H., Mayer, G., & Zlabinger, G. J. (1999). Novel C5-dependent mechanism of neutrophil stimulation by bioincompatible dialyzer membranes. *J Am Soc Nephrol*, 10(1), 128–135.
37. Hochegger, K., Perco, P., Enrich, J., Mayer, B., Mayer, G., Rosenkranz, A. R., & Rudnicki, M. (2007). In vitro - Transcriptional response of polymorphonuclear leukocytes following contact with different antigens. *European Journal of Clinical Investigation*, 37(11), 860–869. <https://doi.org/10.1111/j.1365-2362.2007.01872.x>
38. Brinkmann, V. et al. (2004) Neutrophil extracellular traps kill bacteria. *Science* 303, 1532–1535.
39. Amulic, B., Cazalet, C., Hayes, G.L., Metzler, K.D. & Zychlinsky, A. (2012) Neutrophil function: from mechanisms to disease. *Annu. Rev. Immunol.* 30, 459–489.
40. Branzk N., Lubojemska A., Hardison S.E., Wang Q., Gutierrez M.G., rown GD & Papayannopoulos V. (2014) Neutrophils sense microbe size and selectively release neutrophil extracellular traps in response to large pathogens. *Nat Immunology*, 15 (11), pp. 1017-1024.
41. Lundqvist M, Stigler J, Elia G, Lynch I, Cedervall T, Dawson KA. Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. *Proc Natl Acad Sci U S A.* 2008;105:14265–70.
42. Jani, P. U., Nomura, T., Yamashita, F., Takakura, A., Florence, A. T., & Hashida, M. (1996). Biliary excretion of polystyrene microspheres with covalently linked FITC fluorescence after oral and parenteral administration to male wistar rats. *Journal of Drug Targeting*, 4, 87–93.

43. Florence, A. T., & Hussain, N. (2001). Transcytosis of nanoparticle and dendrimer delivery systems, evolving vistas. *Advanced Drug Delivery Reviews*, 50, S69–S89.
44. Prietl B, Meindl C, Roblegg E, Pieber TR, Lanzer G, Fröhlich E. Nano-sized and micro-sized polystyrene particles affect phagocyte function. *Cell Biol Toxicol*. 2014;30(1):1-16. doi:10.1007/s10565-013-9265-y
45. Awaad, A., Nakamura, M., & Ishimura, K. (2012). Imaging of size-dependent uptake and identification of novel pathways in mouse Peyer's patches using fluorescent organosilica particles. *Nanomedicine: Nanotechnology Biology and Medicine*, 8, 627–636.
46. Cauwenberghe, L van; Janssen, Colin R (2014) Microplastics in bivalves cultured for human consumption. *Environmental Pollution* 193, pp 65-70.
47. Garrett, N. L., Lalatsa, A., Uchegbu, I., Schätzlein, A., & Moger, J. (2012). Exploring uptake mechanisms of oral nanomedicines using multimodal nonlinear optical microscopy. *Journal of Biophotonics*, 5(5–6), 458–468. <https://doi.org/10.1002/jbio.201200006>

## Supplemental Material



### Supplemental Video 1. Neutrophils can phagocytose microplastics in living mice.

Fluorescent 1  $\mu\text{m}$  microplastics (red) were intradermally injected in the ears of LysM-EGFP mice expressing GFP in neutrophils (green). Neutrophils were recruited to the site of injection. Interactions between neutrophils and microplastics occasionally lead to phagocytosis (arrow), but often microplastics were ignored. Representative example of 12 movies from 5 mice. Scale bar represents 20  $\mu\text{m}$ .



### Supplemental Video 2. Human neutrophils phagocytose 7.3 $\mu\text{m}$ MPs

Fluorescent 7.3  $\mu\text{m}$  beads were labelled with pHrodo and added to isolated human neutrophils. Timelapse microscopy images are shown. The left panel shows neutrophils in brightfield, green fluorescent 7.3  $\mu\text{m}$  beads and red fluorescent pHrodo signal upon acidification in the phagolysosome. The right panel shows only the fluorescent pHrodo signal. Multiple neutrophils phagocytose microplastic, one neutrophil phagocytosing a microplastic is tracked, as indicated by the white arrow. The blue arrow indicates neutrophil cell death, the microplastic is released and the pHrodo signal disappears.





## CHAPTER 8



# General Discussion



With increasing computer power the variables being measured not only in medicine or biology, but also in social sciences, businesses, education and journalism have massively expanded. For the flow cytometry field the increase in number of markers that can be simultaneously measured leads to the situation that more different cell types can be more accurately identified within the same sample. Additionally, individual cell types can be characterized in more detail. Thus, measuring more markers in the same sample offers the prospect of knowing more with less material. This will likely aid our understanding on health and disease and thereby could potentially help the diagnosis or even elucidate the mechanism of the pathophysiology of the disease. This is the consensus in the field,<sup>1-3</sup> but it is yet to be established that this promise holds true.

**A common saying in Dutch is “To measure is to know”. But do larger flow cytometry datasets lead to more knowledge or better understanding?**

The challenge from measuring more markers is to extract all the *relevant* information from the large multicolour flow cytometry (MFC) datasets. With an increasing number of markers manual analysis quickly becomes overwhelming and subtle correlations can get lost. Researchers tend to use certain markers solely for identification of cell types or to dichotomize cells as either positive or negative, ignoring possibly interesting correlations between cells present within one population. Furthermore, manual analysis is very time consuming for data that contain more than three dimensions. To tackle these problems and gain most knowledge from MFC data, multidimensional analysis methods are being developed. However, these analysis methods have generally not been developed by researchers using MFC but by statisticians and mathematical modellers. Frequently, the use of the methods and the interpretation of the results obtained from these multidimensional analysis techniques remains difficult for immunologists or researchers from other types of medical or biological fields. Firstly the use is not widely implemented as many of the multidimensional analysis methods are initially published as complex algorithms in bioinformatics journals. Also, as the development of the multidimensional analysis methods are ongoing processes, many do not have user friendly interfaces available. However, most of the programming codes are available online (on Bioconductor or Cytobank websites) for free and can be run in R, Matlab, Java or Python. Basic knowledge of the programming software is needed, which can be an important barrier for many laboratory scientists.

## Most multidimensional analysis methods are not suited for the detection of rare cells

During my research for this thesis it became clear that good multidimensional analysis methods can only be developed and properly used for analysis if the two research fields invest in tight collaborations and are willing to thoroughly understand each other's field. At the start of the studies described in this thesis many multivariate analysis (MVA) methods had great drawbacks. For instance, some methods that were recently developed, such as the t-SNE based method viSNE,<sup>4</sup> needed down sampling of datasets due to limits in computational power and to prevent overcrowding of the viSNE map. However recently various improvements have been made for the viSNE method. For instance, the Barnes–Hut algorithm<sup>5</sup> was implemented, which uses a tree-like structure to reduce the number of pairwise similarities that actually have to be calculated. This now allows for a higher number of cells to be analyzed and shorter computation time. However, despite this improvement viSNE running time takes considerably longer than PCA-based analysis. And although more than the initial 30,000 cells can be analysed,<sup>4</sup> still large datasets need to be down sampled to approximately 1,000,000 cells. This impedes identification of rare disease-specific cell subsets in a dataset containing multiple samples, such as found in clinical samples from patients with minimal residual disease (MRD). Another disadvantage lies in the fact that newly measured samples cannot be projected into an existing viSNE model, meaning that again a new model has to be created containing all samples. Also, repeating the analysis on the same dataset results in a different multidimensional space. After creating a low-dimensional t-SNE map, an optimization step is performed. The difference in distance between 2 points in the high dimensional space vs. the low dimensional t-SNE map are measured the position of the data points is adjusted until the difference between the high-dimensional and low-dimensional distances is lowest. This optimization step is initiated randomly, so the difference between distances is analysed in a random order, and as a result, it is possible that different runs give you different solutions.<sup>4</sup>

In 2017 another t-SNE based method, HSNE,<sup>6</sup> was launched as a solution to the down sampling problem. In HSNE similar cells are clustered together. Clusters of interest can be selected and zoomed in on at higher levels of the cluster hierarchy. The highest level contains the single cells. The cluster hierarchy makes exploration of the results more time-consuming, as it is necessary to scan through multiple levels of hierarchy to find disease-specific cells and characterize them. In contrast to manual gating, similar cells are clustered automatically and objectively. However, the clusters (and thus cells) which are not selected by the researcher to zoom in on are again neglected by the researcher, most probably based on his/ her assumptions and interests. Overall, selection of and zooming in on clusters gives a more fragmented overview of the dataset, compared to a single cell overview. Other disadvantages from the viSNE method, are the variance in results when

analysing the same dataset multiple times, the impossibility to project additional cells in an existing HSNE model and the inability to save the HSNE analysis in the Cytosplore software for further analysis. Redoing the analysis leads to a different cluster map and cellular map. Meaning that the resulting figures have to be (photo)copied and stored before closing the analysis. After closing it is impossible to explore additional findings at a later time point.

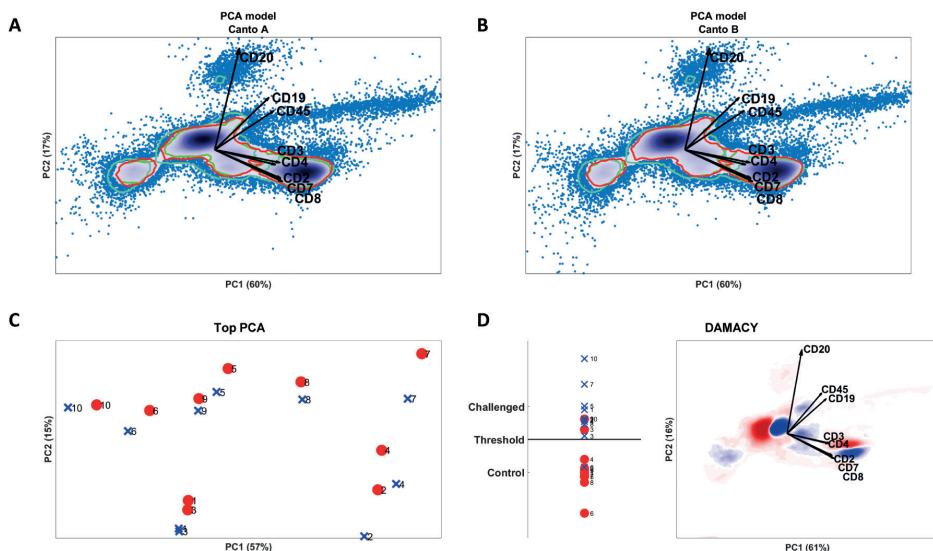
### **ECLIPSE automatically removes normal cells, detects disease-specific cells and rare cell populations are emphasized**

In an attempt to solve the challenges regarding multidimensional identification of rare cells I have, in collaboration with the Jansen group, developed the MVA method ECLIPSE (**Chapter 5**).<sup>7</sup> ECLIPSE is a Simultaneous Component Analysis (SCA)-based<sup>8</sup> method in which control-like cells are removed from the dataset, leaving only the disease-specific cells for further analysis. Thereby, ECLIPSE increases the focus on disease-specific cell populations. When additional follow-up samples are measured for MRD detection, these can be projected into the ECLIPSE space. Also ECLIPSE does not require down sampling of datasets and repetitive analyses of the same dataset result into an identical multidimensional space. Additionally, in **Chapter 5** we have shown that when first performing the smart down sampling of ECLIPSE (elimination of control-like cells) and subsequently run tSNE on the down sampled dataset, also improves the viSNE results (Chapter 5, Supplementary figure S14).

### **For multidimensional analysis standardization of cytometer settings and data pre-processing is required**

Unsupervised multidimensional analysis is difficult to be performed on samples measured with different experimental settings. As a proof of principle 10 peripheral blood samples were stained with the same antibody panel and measured on 2 different BD Canto machines with the same settings including excitation wavelengths of lasers and similar detectors. Even though cytometer settings were optimized by using a mix of multi-color beads of varying brightness levels, adjusting PMT voltages in order to obtain similar MFI's per bead and per channel between the 2 machines (data not shown), upon multidimensional analysis differences were found. Regular PCA results only show minimal differences between the same samples measured on the different Canto's (Figure 1A+B). To analyse the samples further a Top PCA was performed, which describes the multi-dimensional variance between the samples. The corresponding samples were not plotted at the same location of the PCA space (Figure 1C). The difference between the two

Canto's was mostly described in PC2, as the distance between similar samples was largest in this direction. Furthermore, when performing DAMACY analysis (Figure 1D), samples were quite accurately classified into two groups, defined by the 2 different machines, with an accuracy of 85%, sensitivity of 90% and specificity of 80% (Figure 1D). This implicates that when samples from different patient groups or time points are compared with each other, and multiple cytometers or cytometer settings have been used, these technical differences might lead to multidimensional variances in the results. Even if the samples within the two groups are identical to each other, multidimensional differences may be attributed to the fact that the two groups are different, while actually this is not the case.



**Figure 1. Flow cytometry measurements on different BD Canto instruments lead to differences in multidimensional analyses.**

10 Peripheral blood samples were stained with an 8-color MFC panel. The 10 samples were measured on 2 BD Canto flow cytometers (named Canto A and B). A PCA analysis was performed based on the 20 fcs files (excl doublets and cell debris). A+B) representative plots of one of the samples measured on Canto A (A) and on Canto B (B). C) A Top PCA was performed. The red filled circles represent the samples measured on Canto A; while the blue crosses represent the samples measured on Canto B. D) DAMACY analysis of the 20 samples. Samples were classified in two groups with an accuracy of 85%, sensitivity of 90% and specificity of 80% for Canto A and B measurements.

In general, the quality of the dataset is very important. When analysing data manually, and checking gates per sample, variances caused by technical problems will most probably be noticed. When analysing data using an automated algorithm, it is important to be aware of technical issues such as compensation errors. Also variations within the dataset over time, for example due to the use of a new antibody lot which was not titrated appropriately, will influence the results and may lead to wrong conclusions. As statisticians say “crap in is crap



out”, which is also true for multidimensional flow cytometry analysis. Multiple automated pipelines to clean the data prior to multidimensional analysis are available online, in Cytobank. This is an online platform ([www.cytobank.org](http://www.cytobank.org)) where datasets can be imported and these cleaning algorithms can be applied to the dataset. The different options for pre-processing have been elaborated upon by Saeys et al.<sup>9</sup> As listed in this review, data cleaning consists of the following steps: proper compensation and transformation of the data; filter out any marker value regions with abnormal behaviour over time; compare samples to detect batch effects and normalize the data if necessary; correct for internal or external confounding effects; remove debris, dead cells and doublets; check boundary effects (outlier events); and consider pre-gating on a cell population of interest.<sup>9</sup> In common practice, however, these data cleaning codes are not standard used prior to flow cytometry analysis. This again shows that there is a gap between the mathematical modellers who build the codes and researchers who would benefit from using the codes.

### **OTflow defines the correct parameter for transformation of flow cytometry data**

Another important problem I encountered after analysing many different datasets with various MVA methods, was the fact that improper parameters chosen for biexponential (arcsinh or logicle) transformations resulted in a “peak-splitting artefact”. When for a marker with homogeneous expressions (one peak) the cofactor for arcsinh, or W for logicle transformation is too low, the linear range of the biexponential transformation becomes too wide. This will result into a dip in the expression levels around 0. As has been shown in **Chapter 4**, prior to multidimensional analysis the choice of improper transformation parameters can lead to the separation of single homogeneous cell populations into multiple cell populations. The artificial cell populations typically consist of cells with negative intensities for the improper transformed marker, and the other cell population only contains positive intensities for the same marker. This can lead to misleading results showing extra non-existent populations in multidimensional cell maps or cluster maps. Together with the Jansen group I developed OTflow, (**Chapter 4**) a data-driven method which calculates the optimal parameter for transformation per channel measured. The algorithm performs a wide range of possible transformations and scans through the transformed data in search of artefacts like skewed peaks, very sharp or very broad peaks, negative mean intensities of peaks and/or the heterogeneity between peaks of different intensities which is inherent to flow cytometry measurements, but does not have a biological meaning. Per explored transformation parameter a score is generated and the parameter which creates the lowest score is identified as the optimal parameter value. OTflow is the only transformation method which prevents peak splitting. Thereby

misleading conclusions originating from the peak-splitting artefact due to improper transformation are prevented.

### Gating prior to multidimensional analysis

In order to increase focus on biologically relevant information, cell debris and dead cells are preferably gated out prior to multidimensional analysis. If not, the variance explained in the multidimensional model might focus on aspecific fluorescence properties of dead cells.

Depending on the panel used, it might be useful to gate a certain population prior to multidimensional analysis, as we did for the neutrophils from the peripheral blood samples measured in **Chapter 3**. On the other hand, with extensive manual gating prior to the MVA, again unexpected biologically relevant heterogeneity in the cells that were gated out, is ignored. It is worth noting that prior to ECLIPSE it is not necessary to gate one specific cell population, as in ECLIPSE all control-like cells are automatically removed. Thereby, interesting unexpected disease-specific cells are retained in the ECLIPSE model, which may have been gated out manually, and the focus on the more subtle variance within one cell population will be more enhanced. This will be explained further in the next paragraph.

### Increasing MFC panels and challenges for multidimensional analysis

#### In PCA, information might remain hidden in higher PCs

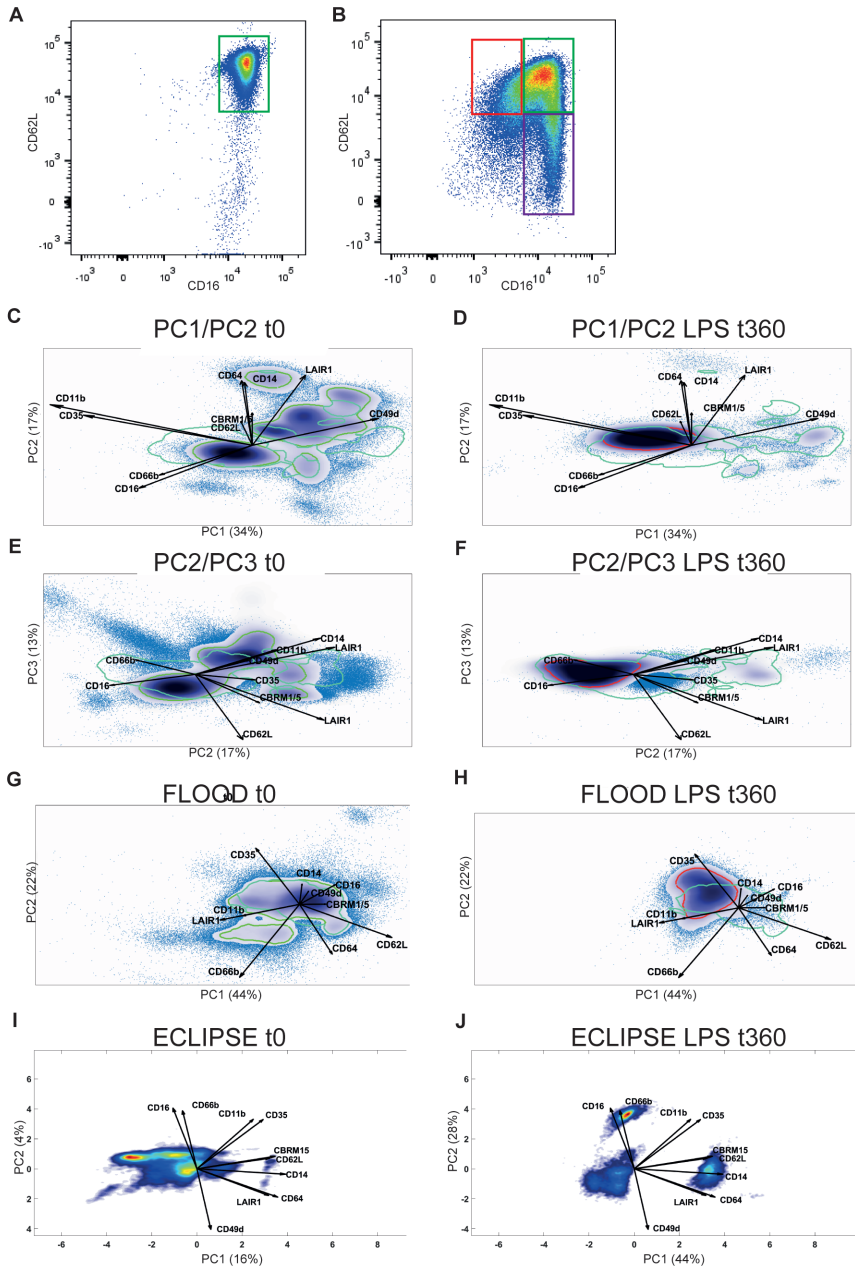
Although we have managed to improve many aspects of the MVA methods, other concerns and challenges remain. Using PCA-based analysis methods like FLOOD,<sup>10</sup> DAMACY<sup>11</sup> and ECLIPSE<sup>7</sup> we mainly analysed PC1 and PC2 or the first three PCs, as cross-validation showed that these PCs were sufficient to explain the most important variance between the different cell subsets (**Chapter 3, 5 and 6**). However, the more markers measured, the more variance and possibly relevant information will be present in higher PCs. An example is shown in Figure 2. A PCA model of 6 LPS challenged individuals was built containing samples at t=0 and t=180 minutes (Figure 2A and B, respectively). In this case, the markers which are positive or negative for specific leukocyte populations are predominantly represented in PC1 and PC2, while variances within cell types, such as within the neutrophil compartment in this example, (Figure 2C+D) are not visualised in the first 2 PCs. The characteristics of the population of neutrophils in the multidimensional analysis shown in Figure 2B support the presence of a homogenous population while the analysis in 2D (CD16/CD62L, Figure 2D) clearly identifies 3 overlapping populations,

as indicated with the gates. In this case, the neutrophil heterogeneity for CD16/CD62L variance was described in PC2/PC3 (Figure 2D+E).

Therefore, for large panels it may be needed to analyse more PCs to derive the different correlations between the markers and to be able to characterise the various (possibly subtle) changes in phenotypes of cells. Otherwise, important information might remain hidden in higher PCs. Which PCs are important to explain the relevant information, is dependent on the percentage of variance explained per PC. The number of PCs needed to analyse should be validated using a cross-validation method. In the example of Figure 2A+B, showing PC1 and PC2, the variations in expression within the neutrophil compartment were overshadowed by population differences.

Fortunately, FLOOD and ECLIPSE are not “simple” PCA methods, they use additional calculations to emphasize disease-specific information. Both methods start with creating a normal PCA model based on the control group, called the control model. FLOOD continues projecting the response group in the control model. The control model does not describe disease-specific information, as the model was solely built based on control samples. As a subsequent step, the information explained in the control model is removed from the dataset. So the variance of control model PC1 and PC2 is removed. Then a new PCA model is generated, based on the remaining variance within the response group. The response model can be visualised using the new response PC1 and PC2. Analysing the above mentioned dataset using FLOOD, the response model does show the neutrophil heterogeneity at t=180 min (Figure 2F+G). In ECLIPSE response samples are projected in the control model and the control-like cells are defined. As a smart down sampling step the control-like cells are subsequently removed from the dataset. The response model is generated based on the disease specific or response specific cells of all response samples. Individual control or response samples can be projected in the ECLIPSE space. ECLIPSE was performed on the same dataset and results are shown in Figure I+J. In the t=180 min sample 2 additional cell populations were found (Figure 2I) which were not present at t=0 (Figure 2J). The marker expressions of these cells could be easily characterized based on the marker loadings.

In conclusion, FLOOD and ECLIPSE describe the control group in the control model; while the FLOOD response model and the ECLIPSE model describe the disease-specific variance or cells best. Therefore, more relevant information is described in the first 2 PC's of the response and ECLIPSE models. Nonetheless, when MFC panels will continue to increase, also for FLOOD and ECLIPSE a limit might be reached and relevant information might end up in higher PC's.



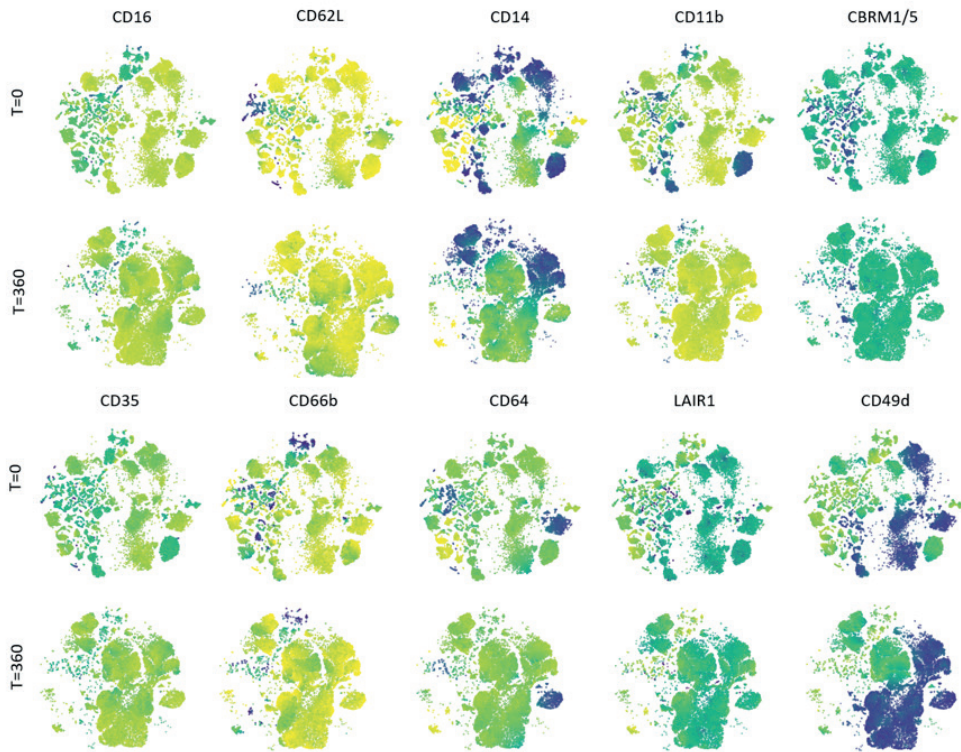
**Figure 2. PCA-based analyses of peripheral blood leukocytes showing differences between cell populations, but within-population differences remain hidden.**

Peripheral blood leukocytes were stained with a 10-color panel at  $t=0$  and  $t=180$  min after LPS administration. A) In homeostasis human peripheral blood contains one main homogeneous population of neutrophils with a CD16<sup>bright</sup>CD62L<sup>bright</sup> phenotype. At  $t=180$  min after LPS administration 3 main neutrophil subtypes were found based on CD16 and CD62L expressions,

CD16<sup>bright</sup>CD62L<sup>bright</sup> (green gate), CD16<sup>dim</sup>CD62L<sup>bright</sup> (red gate), CD16<sup>bright</sup>CD62L<sup>dim</sup> (purple gate). C-F) PCA model of the LPS dataset. At  $t=0$  (C), the main variance explained in PC1 comprises of cells that express CD66b, CD16, CD11b and CD35 (neutrophils, main cell population in the center of the plot) and cells which do not express these markers, of which some populations do express CD49d and/or LAIR1 (lymphocytes and eosinophils) or CD14 and CD64 (monocytes, at 12 o'clock). At  $t=180$  min (D) all neutrophils have higher expressions levels of CD11b and CD35, as they shift towards the direction of these markers. The neutrophil heterogeneity based on CD16/CD62L is not clearly visualised PC1 vs. PC2. In the PCA model, CD62L is represented by a very short loading, meaning the variance is not well explained by this model. CD16 is a relatively long loading, however in the  $t=180$  PCA plot no additional CD16 low neutrophils are visualised when compared to  $t=0$ . Analysing PC2/PC3 plots of the same samples shows neutrophil heterogeneity at  $t=180$  with a part of the neutrophils having low CD16 or low CD62L expression levels. G+H) FLOOD response space at  $t=0$  and  $t=180$  min after LPS administration. The main cell types were described in the Control model, so this information is not represented here. At  $t=180$  min additional CD16<sup>dim</sup> and also CD62L<sup>dim</sup> neutrophils were found. In C-H the cyan outline, or benchmark, encompasses 80% of the control cells of all 8 individuals at  $t=0$ . The green benchmark represents the region in which 80% of the cells of this specific individual are located at  $t=0$ . The red benchmark encompasses the region in which 80% of the cells of the same individual at  $t=180$  min. I+J) ECLIPSE model of the same individual at  $t=0$  and  $t=180$  min. After removal of control-like cells, donor-specific cells at  $t=0$  are visualised when projected in the ECLIPSE model (I). The cells of the same individual at  $t=180$  min was projected in the ECLIPSE model in J. Two extra cell populations were identified which were nearly not present at  $t=0$ . One cell population was characterized as CD16 and CD66b high and with relatively low expression levels of CD62L, CBRM1/5, CD64 and CD14. The other cell population exhibits opposite expression levels.

## In viSNE analyzing marker expression becomes inefficient for large panels

Surely, the challenges with bigger datasets are not specific for PCA-based analysis methods. For tSNE-based analysis methods like viSNE or HSNE also increasing the number of markers makes the analysis more complex. In contrast to PCA-based analysis, in viSNE there are not multiple axes to choose from. tSNE-x and tSNE-y are the axes to plot the cellular map. To illustrate this we performed tSNE analysis on the same LPS dataset as described above, containing peripheral blood leukocytes at  $t=0$  and  $t=180$  min after LPS challenge (Figure 3). Here the biggest difference between the two time points was the appearance of extra cells at  $t=180$  and the disappearance of the small cell subsets. Part of the increased cells were CD16<sup>bright</sup>CD62L<sup>dim</sup> and CD16<sup>dim</sup>CD62L<sup>dim</sup> neutrophils. However, these are not visualised in the tSNE space. It is worth to notice that with increasing parameters correlations between marker expressions become difficult to discover. Marker expressions are visualised by imposing a colour heatmap as a third dimension over the tSNE cell map (**chapter 5 and 6** and Figure 3). Checking 10 heatmaps and searching for correlations between markers, (as was done for the LPS analysis (Figure 3)), was inefficient and laborious. Imagine checking for example 27 heatmaps separately and searching for possible correlations between markers will be very time-consuming, if at all possible.



**Figure 3. t-SNE analysis of peripheral blood leukocytes showing differences between cell populations, but within-population differences remain hidden.**

tSNE analysis of the LPS dataset, showing the peripheral blood cellular map of 6 healthy volunteers at  $t=0$  and 3 hours later, at  $t=180$  min after LPS administration. The expression levels per marker are shown as heatmaps (blue: relatively low expression; yellow: relatively high expression). At  $t=0$  cells are clustered in various cell populations. Clearly additional cells appear after LPS administration. However, no clear CD16dimCD62Lbright or CD16brightCD62Ldim neutrophil subsets were identified.

### **Multidimensional analysis methods give a more complete overview of a cellular compartment compared to manual gating**

Apart from the challenges which remain when analysing large datasets, also when applying MVA, multidimensional analysis has greatly helped in understanding differences in health and disease or variances over time in the studies described in this thesis. As shown in **Chapter 3**, the neutrophil compartment of amateur cyclists who participated in an intensive exercise challenge for 8 days, changes greatly over time.<sup>12</sup> Using FLOOD analysis for deep phenotyping of the peripheral blood neutrophils over 6 different time



points gave us insights in the exercise-induced changes and the inability to fully recover overnight. The results showed that extra CD16<sup>bright</sup>CD62L<sup>dim</sup> cells were recruited to the peripheral blood and also more CD16<sup>bright</sup>CD62L<sup>dim</sup> were found. Next to the rise of these subsets, the whole peripheral blood neutrophil compartment changed over time and the same markers were up- and downregulated on the different neutrophil subsets. Apart from the fact that the FLOOD analysis saved a lot of time compared to the manual gating of 180 fcs files, which is a key advantage of MVA, the FLOOD analysis also resulted in a more 'holistic' overview of the neutrophil compartment. This would have been impossible for manual gating, as gating subsets and comparing MFI's gives a more fragmented overview. Also by projecting LPS challenged peripheral blood neutrophils in the FLOOD exercise model, I could analyse heterogeneity in the changes of the neutrophil compartment between these 2 challenges. This demonstrated that the LPS-induced and exercise-challenged neutrophils did not overlap in the FLOOD space, meaning that, even though in both challenges the two-dimensional CD16/CD62L changes in expression levels were similar, the 10-dimensional phenotypes of the cells were completely different from each other. We could conclude that after multiple days of intensified endurance exercise, the neutrophil compartment did not recover overnight. Importantly, the aberrant neutrophil subtypes increased during the tour. These neutrophils did not upregulate classical activation markers arguing against a degranulated phenotype, on the contrary an immune-suppressive neutrophil phenotype was found. It is tempting to speculate that this phenotype is associated with the increased risk of URTI during periods of intensified training in athletes.

### **Collaborations between researchers, statisticians and flow core facility operators are of key importance**

Many of the datasets were analysed with models that were designed in collaboration with the Jansen group as well as other MVA algorithms. The advantages and disadvantages of different methods could be identified by applying and comparing them to my own datasets. Many immunologists work on research projects of which flow cytometry is one of the key techniques they use to answer their research questions. This makes it difficult to spend enough time studying the MVA methods and applying them to their datasets. Also with the increasing depth in the datasets in mind, and the increasing number of analysis methods being developed, it would be optimal for an immunology group to have one or multiple statistician(s) as a support for data analysis. Ideally, the flow core of an immunology institute employs people with knowledge of both flow cytometry as well as multidimensional data analysis techniques. These persons can help immunologists in producing good quality flow cytometry datasets, train immunologists in using multidimensional analysis methods and help out with selecting the algorithm which suits best for their specific dataset and research question.

## Will MVA replace conventional manual analysis?

As already discussed above the important condition in order to apply MVA to an MFC dataset is a good quality of the dataset, obtained from experiments with standardised protocols and cytometer settings. Therefore, when setting up a panel, testing markers and still slightly adjusting cytometer settings or compensation settings between samples, manual (gating) analysis is preferably used. On the other hand, when a MFC panel is finalised and there is a standard operating procedure to isolate the cells and perform the flow cytometry measurements, a dataset can be generated and multidimensional analysis can be performed. The multidimensional analysis can visualise the most important variances in the data and might give new insights about cell phenotypes present in the samples. Currently, the findings from MVA are still verified by manual gating. This already saves a lot of time since the researcher does not need to gate all the cells from every sample and analyse the number of cells or MFI's of cells within the gates. From the MVA results it is possible to specifically look for a cell population with known marker expressions. Currently we might live in a transition phase between manual analysis and MVA. By gaining more knowledge about MVA and more experience in MVA and by developing more comprehensive and robust MVA methods, more trust will be gained in the interpretation of the MVA method. Therefore, it is to be expected that MVA will replace manual analysis.

Multidimensional analysis methods must be implemented especially in clinical diagnostics, where standardised operating procedures are daily routine and samples are measured 24/7. In a diagnostics setting these methods can be very helpful. Most likely it will be feasible to generate a multidimensional “red flag” analysis using an MFC panel which contains markers to identify the most prevalent indicators for leukocyte phenotypes in a certain disease both in peripheral blood and in bone marrow. This is nicely illustrated in diagnostic procedures for haematological malignancies. Using control bone marrow samples the normal bone marrow cells can be defined by ECLIPSE. For every patient bone marrow sample that is measured, ECLIPSE can remove all normal-like cells. Then a patient-specific ECLIPSE model is generated and based on the distribution of patient-specific cells a “red flag” can show up, indicating that the sample should be further analysed. If this works, and larger patient datasets become available, it might even be possible to generate a group-based ECLIPSE model which can also indicate what type of disease the patient has, dependent on the location of the disease specific cells in the multidimensional ECLIPSE model.

To come back to the initial question whether large flow cytometry datasets will increase our knowledge, I would say yes, but this not unconditional. I would not advocate to just



measure as many markers as possible in one panel, as even with the currently available multidimensional analysis methods, the larger a panel gets, the bigger the chance that important information might remain or even become hidden. Therefore I would recommend to design robust panels, with the pending research question in mind. And instead of spending efforts in including as many markers as possible, preferably spend more time in setting up proper panels and correspondingly appropriate cytometer settings. Multidimensional analysis methods might even be used for selection of the most relevant markers for a certain sample or cell population. Especially PCA-based methods can provide this information in a blink of an eye, as the length of the marker loadings in FLOOD, ELIPSE and DAMACY define which markers are most important to explain the variance between the cells analysed.

Setting up high quality panels, will result in flow cytometry datasets of high quality, without compensation errors, transformation artefacts and batch effects. High quality datasets are very well suited for thorough multi-dimensional analysis.

## References

1. Irish JM. Beyond the age of cellular discovery. *Nat Immunol.* 2014;15(12):1095-1097. doi:10.1038/ni.3034
2. Chattopadhyay P, Perfetto S, Gaylord B, et al. Toward 40+ Parameter Fluorescence Flow Cytometry. In: *XXIX Congress of the International Society for Advancement of Cytometry.* ; 2014.
3. Robinson JP, Roederer M. Flow cytometry strikes gold. *Science (80- ).* 2015;350(6262):739-740. doi:10.1126/science.aad6770
4. Amir ED, Davis KL, Tadmor MD, Simonds EF, Jacob H. viSNE enables visualization of high dimensionainll data and revnotypic heterogeneity of leukemia. *Nat Biotechnol.* 2014;31(6):545-552. doi:10.1038/nbt.2594.viSNE
5. Maaten L Van Der. Accelerating t-SNE using Tree-Based Algorithms. *J Mach Learn Res.* 2014;15:3221-3245. doi:10.1007/s10479-011-0841-3
6. Van Unen V, Höllt T, Pezzotti N, et al. Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. *Nat Commun.* 2017;8(1). doi:10.1038/s41467-017-01689-9
7. Folcarelli R, van Staveren S, Bouman R, et al. Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm. *Sci Rep.* 2018;8(1). doi:10.1038/s41598-018-29367-w
8. Timmerman ME, Kiers HAL. Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika.* 2003. doi:10.1007/BF02296656
9. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: Helping to make sense of high-dimensional immunology data. *Nat Rev Immunol.* 2016;16(7):449-462. doi:10.1038/nri.2016.56
10. Jansen JJ, & Hilvering B, van den Doel A, et al. FLOOD: FLOW cytometric Orthogonal Orientation for Diagnosis. *Chemom Intell Lab Syst.* 2016;151(December):126-135. doi:10.1016/j.chemolab.2015.12.001
11. Tinnevelt GH, Kokla M, Hilvering B, et al. Novel data analysis method for multicolour flow cytometry links variability of multiple markers on single cells to a clinical phenotype. *Sci Rep.* 2017;7(1):1-11. doi:10.1038/s41598-017-05714-1
12. Van Staveren S, Ten Haaf T, Klöpping M, et al. Multi-dimensional flow cytometry analysis reveals increasing changes in the systemic neutrophil compartment during seven consecutive days of endurance exercise. *PLoS One.* 2018;13(10). doi:10.1371/journal.pone.0206175



# APPENDICES



Nederlandse Samenvatting  
Dankwoord  
Curriculum Vitae  
List of Publications





## Nederlandse samenvatting

### Introductie in flow cytometrie en multi-dimensionele analysemethoden

#### Middels flowcytometrie worden eigenschappen van cellen gemeten

Flowcytometrie is een veelgebruikte techniek voor het bestuderen van microscopische deeltjes in een vloeistof, bijvoorbeeld een vloeistof met menselijke cellen. “Flow” staat voor de Engelse term *in flow*, wat stroming (van een vloeistof) betekent. “Cyto” betekent cel en “metrie” verwijst naar de meting. Samen betekent flowcytometrie dus meting van stromende cellen. De eerste flowcytometer werd dan ook ontwikkeld voor het meten van eigenschappen van cellen in een vloeistof.

Cellen bestaan uit moleculen, onder andere eiwitten. Dit zijn verschillende typen eiwitten met diverse functies. Om te achterhalen welk cellen er aanwezig zijn in het bloed, meten we met behulp van flowcytometrie welke eiwitmoleculen er aanwezig zijn op de buitenkant van de cel of in de cel. Als we weten welke cellen aanwezig zijn in het bloed, kunnen we deze cellen tevens vergelijken met andere cellen. Door flowcytometrie kunnen we cellen in bijvoorbeeld het bloed van zieke patiënten vergelijken met cellen in het bloed van gezonde proefpersonen. In het onderzoek wordt flowcytometrie gebruikt om te achterhalen welke processen een rol spelen bij het ontstaan van een ziekte of de processen een rol spelen gedurende het beloop van een ziekte. In de kliniek wordt flowcytometrie ingezet voor het diagnosticeren van ziekten.

Witte bloedcellen maken deel uit van ons immuunsysteem. Het immuunsysteem beschermt ons tegen indringers zoals bacteriën en virussen. Als het immuunsysteem minder goed werkt, kunnen infecties ontstaan. In het geval van auto-immuunziekten is het immuunsysteem juist overactief en reageert het op lichaamseigen cellen en weefsels, wat leidt tot weefselschade. In de immunologie wordt flowcytometrie gebruikt om witte bloedcellen uit bijvoorbeeld ons bloed of beenmerg te analyseren. Elk type cel en zelfs kleine non-organische microscopische partikels (zoals microplastics, hier komen we in hoofdstuk 7 op terug) kunnen met de flowcytometer gemeten worden. In deze introductie zal ik specifiek de analyse van witte bloedcellen gebruiken als voorbeeld.

#### Identificatie en karakterisering van cellen middels flowcytometrie

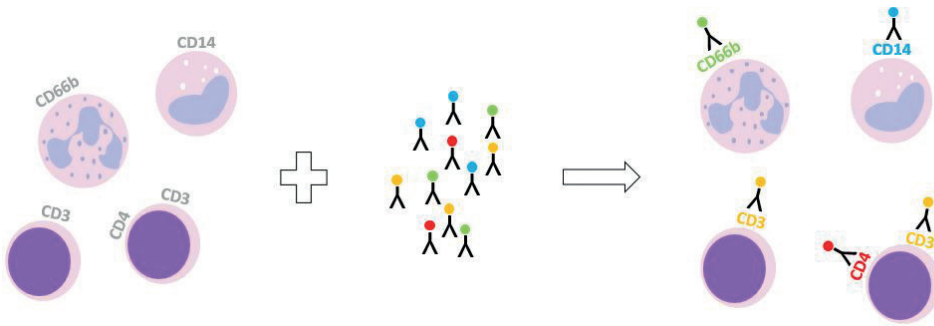
In de flowcytometer staan laserstralen gericht op een vloeistofstroom. Wanneer een cel langs de laserstraal komt wordt het licht van de laserstraal door de cel verstrooid. De mate van verstrooiing is een indicatie voor de grootte van de cel en de complexiteit van de cel. Op basis daarvan kunnen we tot op zekere hoogte verschillende soorten bloedcellen van elkaar onderscheiden. Metaforisch gesproken kunnen we op deze manier een ananas van een appel onderscheiden. Dat is niet heel gedetailleerd. We

zouden ook graag appels van peren willen onderscheiden, en liefst Jonagold appels van Pink Lady's. Om celtypes in meer detail van elkaar te onderscheiden, kan de flowcytometer meten welke eiwitten er aanwezig zijn op en in de cellen. Daarvoor worden er na het verkrijgen van patiëntencellen eerst fluorescente antilichamen aan de cellen toegevoegd welke de flowcytometer kan meten. Deze antilichamen binden zich aan specifieke eiwitten op of in cellen, dus wanneer een bepaald antilichaam op een cel gesignaleerd wordt, betekent dit dat een specifiek eiwit aanwezig is op de cel. De antilichamen kan de onderzoeker zelf uitkiezen. Meestal wordt er een mix van diverse antilichamen toegevoegd, met aan elk type antilichaam een fluorescente label met een andere kleur (Figuur 1). De antilichamen kunnen specifiek zijn voor een eiwitmolecuul die alleen voorkomt op één bepaald type witte bloedcel, of het kunnen antilichamen zijn die binden aan eiwitmoleculen die in gezonde toestand niet aanwezig zijn op de cel, maar bijvoorbeeld wel bij ziekte. De eiwitmoleculen worden ook wel markers genoemd. De lasers in de flowcytometer kunnen fluorescente signalen opwekken en detectoren in de flowcytometer kunnen diverse fluorescente signalen meten. Per cel wordt naast de verstrooiing van het licht ook de fluorescente signalen gemeten. Wanneer het fluorescente signaal van een bepaald antilichaam gemeten wordt, is de marker waaraan dit antilichaam bindt aanwezig op de cel. Er kan dan geconcludeerd worden dat die specifieke marker tot expressie komt op de cel. Op deze manier kan de onderzoeker herleiden welke celtypen er in het bloed van de patiënt zitten. Als er van één celtype meerdere cellen aanwezig zijn spreken we over een celpopulatie. Ook kan de onderzoeker aan de hand van bijvoorbeeld activatie-eiwitten bestuderen of deze cellen andere eigenschappen hebben dan de cellen in het bloed van gezonde proefpersonen. Flowcytometrie is dus een waardevolle methode voor het identificeren en karakteriseren van cellen.

### **Multidimensionele analysemethoden zijn nodig voor het verkrijgen van informatie uit grote flowcytometrie datasets**

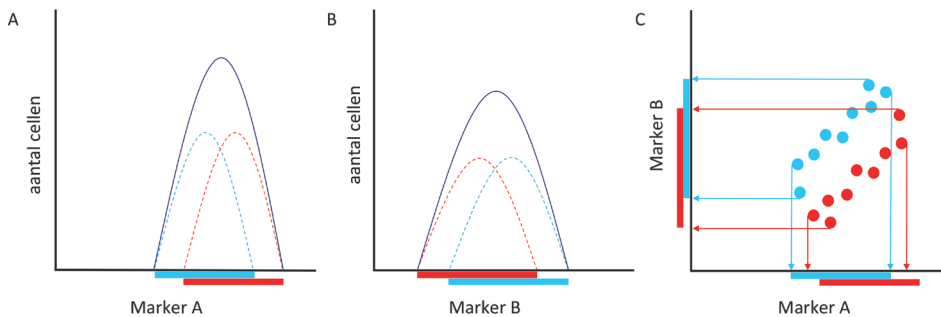
De eerste flowcytometer werd gebouwd in 1965. Aanvankelijk bevatte de flowcytometer één laser en kon er maar één marker per cel gemeten worden. De techniek ontwikkelde erg snel, en inmiddels bestaan er flowcytometers die 40 markers per cel kunnen meten. Per bloedmonster kunnen er duizenden tot miljoenen cellen geanalyseerd worden. Dit leidt tot enorm grote datasets. Het is een uitdaging om alle informatie uit dit soort complexe datasets te verkrijgen.





**Figuur 1. Fluorescente antilichamen binden aan cel-specifieke receptoren.**

Schematische weergave van de proefopzet voor het “kleuren” van cellen voor een flowcytometrie meting. Cellen hebben bepaalde receptoren op hun oppervlak, ook wel markers genaamd. De verschillende markers hebben naast een naam ook een cluster of differentiation nummer, oftewel CD-nummer, waarmee ze doorgaans worden aangeduid. De onderzoeker voegt antilichamen toe die binden aan specifieke receptoren in de celmix. Cellen die CD66b tot expressie brengen worden neutrofielen genoemd; CD14-positieve cellen zijn monocyten; CD3-positieve cellen worden ook wel T-cellen genoemd. Er zijn meerdere typen T-cellen, waaronder de CD4-positieve T-cel. Deze antilichamen hebben verschillende kleuren zodat ze door de flowcytometer onderscheiden kunnen worden.



**Figuur 2. Het weergeven van correlaties tussen verschillende markers zijn essentieel voor identificeren van verschillen tussen groepen cellen**

A) Schematische representatie van de expressie van marker A door twee groepen cellen ookwel celpopulaties genoemd (rood en blauw); B) Schematische representatie van de expressie van marker B door twee groepen cellen (rood en blauw). Op basis van de weergave in A en B is geen onderscheid te maken tussen de 2 groepen cellen wanneer ze tegelijk in één monster gemeten worden, aangezien de meting zal resulteren in de donkerblauwe piek. C) weergave van de expressies van marker A (x-as) en B (y-as) tezamen in één grafiek. De groep blauwe en groep rode cellen worden nu als het ware “uit elkaar getrokken” en lijken wel degelijk te verschillen van elkaar.

### **Conventionele analyse van flowcytometrie data is tijdrovend, arbeidsintensief, inefficiënt, subjectief en foutgevoelig**

De conventionele analyse van flowcytometrie data bestaat uit 1-dimensionele (1D) analyse of 2-dimensionele (2D) analyse. Bij 1D analyse worden de signalen van 1 marker op de x-as en de cel-aantallen op de y-as geplotted (Figuur 2A, B). Voor 2D analyse worden de cellen in een zogenaamde *scatter plot* weergegeven, waarin per cel weergegeven wordt wat de fluorescente intensiteit voor marker A is (weergegeven op de x-as) en de fluorescente intensiteit voor marker B (op de y-as). Een 2D plot geeft dus de correlatie tussen de 2 gemeten markers weer (Figuur 2C). Vaak wordt er overlap gevonden in de expressies van de markers tussen 2 celpopulaties. Middels 1D analyse kan het in dat geval lijken alsof de celpopulaties niet (significant) van elkaar verschillen. Echter subtiele verschillen tussen de expressies van 2 markers kunnen in 2D analyse wel degelijk leiden tot de identificatie van verschillende celpopulaties (zie ook de uitleg bij Figuur 2). 2D analyse geeft dus meer informatie dan 1D analyse.

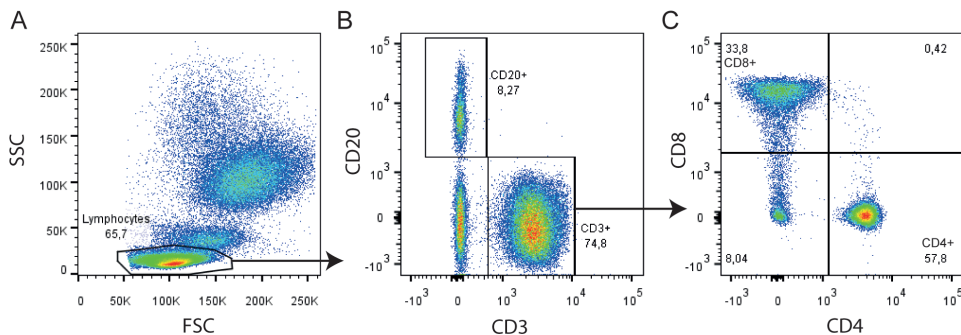
Tegenwoordig worden er in de meeste experimenten 10 tot 30 antilichamen per cel gemeten. Om de correlaties tussen al deze antilichamen te begrijpen, dienen er veel 2D grafieken geanalyseerd te worden. Bijvoorbeeld, wanneer er 10 antilichamen gemeten worden, kunnen de correlaties tussen al deze antilichamen in 45 grafieken weergegeven worden. Vaak bevat een dataset meerdere controle- en patiënten-samples, en voor elk monster kunnen die 45 grafieken interessante informatie bevatten. Bij grote datasets is analyse op deze manier erg tijdrovend, arbeidsintensief en inefficiënt.

Een ander nadeel van de conventionele flowcytometrie analyse is dat een onderzoeker manueel cellen van interesse selecteert in een 1D of 2D grafiek en alleen deze geselecteerde cellen in een volgende grafiek weergeeft voor verdere analyse. Dit manueel selecteren van cellen gebeurt door het plaatsen van een “*gate*”, of begrenzing, in een 2D plot rond de cellen van interesse (Figuur 3). Door het gaten van cellen worden andere niet-geselecteerde cellen door de onderzoeker buiten beschouwing gelaten, terwijl de dataset interessante informatie over deze cellen kan bevatten. Er blijft dus informatie verborgen op basis van de keuzes van de onderzoeker tijdens analyse. Het manueel selecteren van cellen is subjectief en leidt tot variatie in resultaten tussen laboratoria en onderzoekers. Ook is de kans op het maken van fouten bij de analyse van datasets met veel samples, waarin veel gates geplaatst en gecontroleerd dienen te worden, groot.

### **Multidimensionele flowcytometrie analysetechnieken werken geautomatiseerd, snel en objectief**

Om bovenstaande problemen op te lossen zijn er verscheidene multidimensionele analysemethoden ontwikkeld voor flowcytometrie datasets. Het doel is om geheel objectief en geautomatiseerd zo veel mogelijk informatie uit de dataset weer te geven in

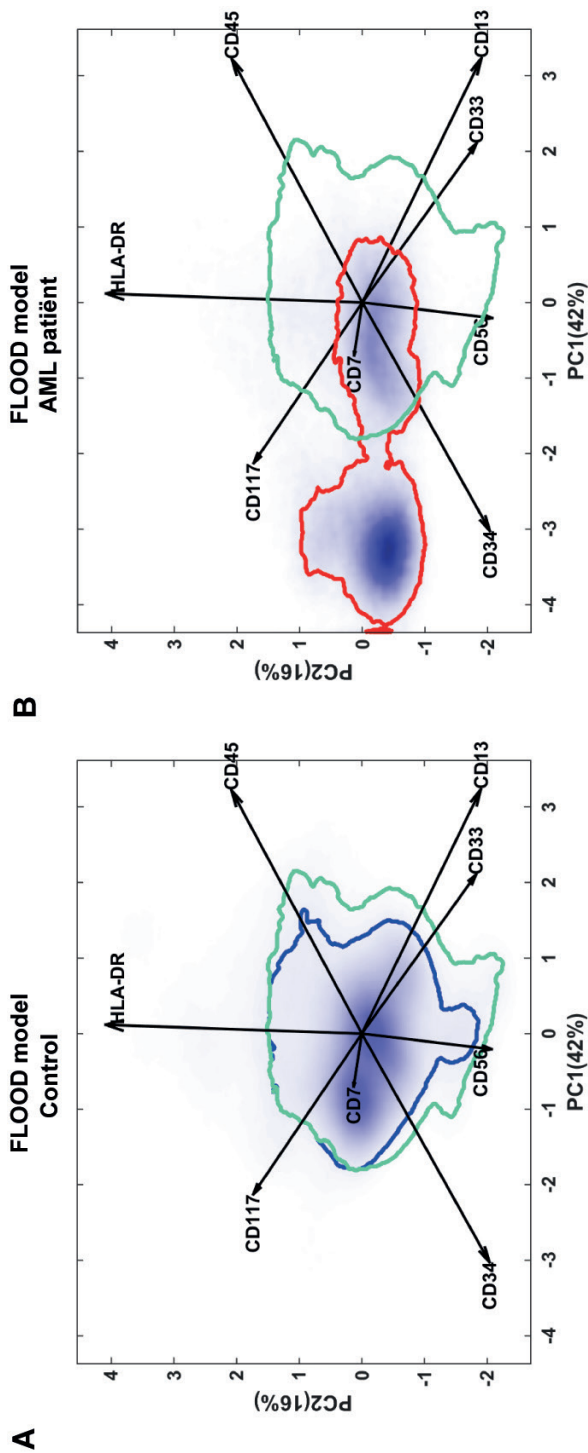
een 2D vlak, zonder dat de resultaten afhankelijk zijn van de keuze of interesses van de onderzoeker. FLOOD is zo'n multidimensionele analysemethode, ontwikkeld voordat ik aan mijn promotietijd begon door onderzoekers van de Radboud Universiteit in Nijmegen en immunologen uit de Koenderman groep. Middels FLOOD wordt per celsample de patiënt-specifieke variatie weergegeven in één grafiek en dit kan vergeleken worden met de variatie in controlesamples in een 2<sup>e</sup> grafiek. De verschillende gemeten markers worden weergegeven als vectoren, de richting van de verschillende vectoren geeft aan hoe de antilichamen met elkaar gecorreleerd zijn. Vectoren die in dezelfde richting wijzen zijn positief gecorreleerd aan elkaar, terwijl vectoren die tegengestelde richting op wijzen negatief gecorreleerd zijn aan elkaar. Als vectoren in een hoek van 90 graden staan, zijn deze markers niet aan elkaar gecorreleerd. De positie van een cel in de grafiek ten opzichte van de vectoren geeft aan welke markers wel of niet tot expressie komen per cel. Als een cel in de richting van de pijl van de vector gepositioneerd is, wordt de marker relatief hoog tot expressie gebracht.



**Figuur 3. Conventionele analyse middels handmatig gaten van cellen.**

De conventionele flowcytometrie analyse bestaat uit het handmatig selecteren van cellen van interesse. A) Weergave van witte bloedcellen op basis van grootte (FSC) en complexiteit (SSC). De celpopulatie bestaande uit kleinere cellen met weinig complexiteit werd geselecteerd door er handmatig een begrenzing, ofwel gate, omheen te tekenen. Deze kleinere cellen worden lymfocyten genoemd. B) De lymfocyten worden verder geanalyseerd op basis van de expressie van marker CD3 en CD20. Lymfocyten die CD20 tot expressie brengen zijn B-cellen en lymfocyten die CD3 tot expressie brengen worden ook wel T-cellen genoemd. De T-cellen worden verder bestudeerd in C. C) Er bestaan verscheidene typen T-cellen, welke CD4 of CD8 tot expressie brengen, of T-cellen die beide markers niet tot expressie brengen.

Het gebruik van multidimensionele analysemethoden voor flowcytometrie datasets is relatief nieuw, en er zijn nog veel uitdagingen. Het is ontzettend belangrijk om de experimenten gestandaardiseerd uit te voeren, de data juist te behandelen en de analyses secuur uit te voeren, aangezien de uitkomsten van de analyses anders misleidend kunnen zijn.



**Figuur 4. FLOOD resultaten van beenmergsamples van een gezonde controle en een leukemie patiënt.**

FLOOD analyse waarin 80% van de cellen van gezonde controle beenmerg (blauw omrand gebied) in A vergeleken worden met 80 % van de cellen van een AML patiënt in B (rood omrand gebied). De groene omranding geeft aan waar 80% van de cellen van de gehele controlegroep gelokaliseerd zijn. De pijlen kunnen als een kompas gebruikt worden om de cellen te karakteriseren. Hoe langer de pijl, hoe belangrijker de marker is voor het verklaren van variatie tussen de verschillende samples. In B hebben de cellen links in de FLOOD plot bijvoorbeeld een hoge expressie van CD34, CD117 en CD7. Tegelijkertijd hebben de cellen links in B een lage CD45, CD13 en CD33 expressie, aangezien ze in tegenovergestelde richting van de pijlen liggen.

Een andere uitdaging is de identificatie van zeldzame ziekte-specifieke cellen. In veel multidimensionele modellen ligt de focus vooral op de meestvoorkomende cellen in een sample of dataset. Zeldzame celpopulaties kunnen overschaduwd worden door de overgrote hoeveelheid aan normale cellen die aanwezig zijn in het bloed van zieke personen. Als patiënten met leukemie of met een lymfoom (kanker van witte bloedcellen in het beenmerg) behandeld worden, is het belangrijk om te vervolgen hoeveel restziekte er over blijft in het beenmerg. Het vinden van restziekte is namelijk gecorreleerd met terugkeer van de ziekte en een slechte prognostische uitkomst, ook voor hele kleine celaantallen.

Een groot interessegebied van onze onderzoeksgroep is de variatie in verschillende soorten neutrofielen. Neutrofielen de meest voorkomende witte bloedcellen in het menselijk bloed. Ze zijn zeer goed in het doden van bacteriën voorkomen zo dat wij infecties krijgen. Aanvankelijk werd gedacht dat de neutrofielen in ons bloed allemaal dezelfde functie hebben. Recent ontdekten mijn voorgangers echter dat neutrofielen zeer verschillende eigenschappen kunnen hebben. Middels flowcytometrie werden tijdens een acute ontstekingsreactie twee extra neutrofiel celpopulaties in het bloed ontdekt met andere eiwitten op het oppervlak vergeleken met de gezonde situatie. In gezond bloed vindt men neutrofielen die CD16 en CD62L allebei hoog tot expressie brengen (afgekort CD16+CD62L+ neutrofielen). Tijdens acute ontsteking komen er CD16-CD62L+ neutrofielen tevoorschijn, dit zijn jonge cellen die nog beter bacteriën kunnen doden dan de CD16+CD62L+ neutrofielen. Ook komen er neutrofielen tevoorschijn met een CD16+CD62L- expressiepatroon, deze cellen zijn minder goed in het doden van bacteriën en hebben mogelijk een ontstekingsremmende functie. Het uitvoeren van experimenten met neutrofielen is altijd een uitdaging. Het zijn erg gevoelige cellen die zodra ze afgenomen worden uit het bloed geactiveerd raken en daardoor andere eiwitten op het oppervlak dragen. Bij een flowcytometrie experiment dienen de cellen dus zo snel mogelijk na afname gemeten te worden. Ook was het niet mogelijk om neutrofielen in te vriezen, aangezien ze dan dood gaan.

## Samenvatting van de studies beschreven in dit proefschrift

Het doel van mijn onderzoek was het verleggen van de grenzen van flowcytometrie methoden en van multidimensionele data analyse.

In **Hoofdstuk 2** beschrijven we een nieuwe methode om neutrofielen in te vriezen en op een later moment te ontdooien voor flowcytometrie meting. Dit kan goed van pas komen bij onderzoek in afgelegen gebieden, waar geen goed uitgerust laboratorium met flowcytometer beschikbaar is. Deze invriesmethode hebben we gebruikt voor de studie

in **Hoofdstuk 3**, waar we 30 amateur fietsers volgden gedurende een intensieve 8-daagse fietstocht van Italië naar Nederland, de Tour for Life. Tijdens deze fietstocht namen we bloedsamples af van de fietsers. In een bestelbus, uitgerust met een primitief lab, hebben we de samples bewerkt en vervolgens ingevroren om later, bij terugkomst in het UMC Utrecht, te bestuderen op de flowcytometer. We analyseerden de flowcytometrie data van neutrofielen op conventionele wijze en middels FLOOD. We vonden een toename in verandering over de tijd waarbij de neutrofielen gedurende de nacht niet meer herstelden tot hun normale staat. Er werden meer CD16-CD62L+ en ook meer CD16+CD62L-neutrofielen gevonden over de tijd. Daarnaast hadden alle neutrofielen (onafhankelijk van de CD16 en CD62L expressie) een verhoogde expressie van LAIR-1, CBRM1/5 en CD49d expressie en een verlaagde expressie van CD11b. De gevonden veranderingen suggereren een mogelijke ontstekingsremmende functie van neutrofielen tijdens een intensieve periode van duursporten. Dit kan verklaren waarom professionele sporters een verhoogde kans op bovenste luchtweginfecties hebben ten opzichte van amateur sporters.

Voordat een multidimensionele flowcytometrie analyse toegepast kan worden, dient de data getransformeerd te worden. Een bekende transformatie is logaritmisches transformeren. Deze type transformatie kan helaas niet toegepast worden op flowcytometrie data aangezien er negatieve waarden aanwezig zijn; logaritmische transformaties kunnen alleen verricht worden op positieve data. Er zijn andere type transformaties die toegepast kunnen worden, zogenaamde bi-exponentiële transformaties. In **Hoofdstuk 4** laten we zien dat het van belang is dat deze transformaties nauwkeurig uitgevoerd worden omdat er anders artefacten in de data optreden. Deze artefacten zijn gevaarlijk, aangezien ze misleidende resultaten kunnen geven en daardoor misleidende conclusies over de dataset. In Hoofdstuk 4 beschrijven we een nieuwe geautomatiseerde methode, genaamd OTflow, om correcte transformatie uit te voeren waarbij geen artefacten ontstaan.

In **Hoofdstuk 5** presenteren we de nieuwe multidimensionele analysemethode ECLIPSE om patiënt specifieke cellen op een geautomatiseerde en snelle manier te identificeren en te karakteriseren. Middels ECLIPSE worden normale cellen gedefinieerd aan de hand van controle samples en deze worden geheel automatisch uit de dataset verwijderd. Door het verwijderen van de normale cellen wordt er gefocust op ziekte-specifieke cellen en kan meer informatie over deze cellen weergegeven worden in het ECLIPSE model. Het verwijderen van normale cellen op basis van de hoog-dimensionele expressieniveaus is uniek, er zijn geen andere multidimensionele methoden die dit doen.

In **Hoofdstuk 6** laten we zien dat ECLIPSE specifiek goed toepasbaar is voor het identificeren van ziekte-specifieke zeldzame cellen. Ter illustratie gebruiken we een dataset van controle beenmergsamples en beenmergsamples van patiënten met lymfoom. Na behandeling

kunnen we bij 7 van de 7 patiënten de restziekte terugvinden middels ECLIPSE analyse en we vinden in 5 van de 7 gevallen ook meer maligne cellen terug dan wanneer we op de conventionele manier middels gating de data analyseren. Dit bewijst dat ECLIPSE een veelbelovende methode is voor het detecteren van zeldzame ziekte-specifieke cellen.

Middels flowcytometrie kunnen niet alleen cellen gemeten worden, maar ook andere kleine microscopische partikels. In de studie in **Hoofdstuk 7** hebben we onder andere flowcytometrie gebruikt voor het detecteren van microplastics om antwoord te krijgen op de vraag wat het effect is van microplastics op de gezondheid van de mens. Het is een actueel thema in het nieuws dat er een toenemende hoeveelheid plastic afval gevonden wordt in onze oceanen en oppervlaktewateren. Deze plastic voorwerpen breken langzaam af tot microscopisch kleine plastic partikels, microplastics. De microplastics worden teruggevonden in schelpdieren en vissen, maar ook in tafelzout, bier en honing. Middels ons voedsel krijgen wij deze microplastics dus binnen. Eerder onderzoek heeft bewezen dat microplastics onze darmwand kunnen passeren. We hypothesiseerden dat neutrofielen de eerste witte cellen zullen zijn die de microplastics in ons lichaam tegenkomen. In het lab hebben we onderzocht wat er gebeurt als neutrofielen met microplastics in contact komen. Het was schokkend om te zien dat de neutrofielen microplastics opnemen, net zoals ze bacteriën in zich opnemen alvorens ze deze doden. Ook zagen we een verhoogde sterfte van neutrofielen. Wanneer neutrofielen in contact werden gebracht met grotere microplastics (van ongeveer dezelfde grootte als de neutrofiel zelf) ging 97% van de neutrofielen binnen 45 minuten dood. Bij kleinere microplastics ging ongeveer 50% van de neutrofielen dood binnen 45 minuten. De cellen die overleefden hadden nog wel een goede bacteriedodende functie. Dit zijn schokkende resultaten, verder onderzoek dient uitgevoerd te worden om te bestuderen wat de effecten van de microplastics in een levend organisme zijn.





## Dankwoord

Dit proefschrift was er niet in deze vorm geweest zonder de hulp en steun van heel veel mensen.

**Leo**, als onderdeel van mijn sollicitatie gaf ik 5 jaar geleden een presentatie over de resultaten van mijn masteronderzoek voor jouw groep. Toen ik vroeg of er nog vragen waren, gaf jij meermaals aan dat je het niet helemaal begreep. Toch nam je me aan als PhD student, wat ik op mijn beurt niet zo goed begreep. Inmiddels begrijp ik dat jij juist enthousiast wordt als je iets niet helemaal begrijpt. En jouw enthousiasme is aanstekelijk. Ik heb ontzettend veel geleerd in de 4 jaar onder jouw begeleiding en ik hoop ook altijd net zo bevlogen te zijn in mijn werk als jij bent in jouw vak.

**Nienke**, ik heb veel bewondering voor jouw kritische blik, doorzettingsvermogen en enthousiasme. Ik herinner me ons eerste gesprek nog, via Skype. Je gezellige Zaanse accent, vrolijke houding en tomeloze enthousiasme over de onderzoeksprojecten stelden me meteen op mijn gemak. In de 4 jaar die daarop volgden heb je meer een academicus van me gemaakt en ben je op veel vlakken een voorbeeld voor me geweest. Daarnaast stond je tijdens de afrondende fase van dit proefschrift praktisch dag en nacht voor me klaar, dat waardeer ik heel erg.

**Pim**, bedankt voor je input betreffende het microplastics aandeel van mijn promotietraject. In dit proefschrift is dat deel helaas kleiner geworden dan we allemaal van tevoren verwacht hadden. De microplastics studie was voor ons super nieuw en experimenteel en daardoor heel uitdagend en leerzaam. De experimenten hadden een hoog *trial and error* gehalte en ze leidden helaas niet altijd tot interpreteerbare resultaten. Pas tegen het einde van mijn promotietijd begon dit eindelijk te lopen. Gelukkig zal hoofdstuk 7 niet het laatste zijn wat er geschreven zal worden over de effecten van microplastics op het immuunsysteem, wordt vervolgd!

**Jeroen**, bedankt voor de super fijne samenwerking. Een significant deel van mijn promotietraject heeft zich in Nijmegen afgespeeld, in jouw groep. Ik voelde me altijd welkom bij jullie, ik heb gelachen tijdens de lunch- en koffiepauzes en ik heb genoten van de discussies die we gevoerd hebben. We waren het niet altijd meteen eens en soms was het lastig om precies te begrijpen wat de ander bedoelde, met zulke verschillende achtergronden. We spraken bijna verschillende talen. Uiteindelijk hebben we elkaars taal leren begrijpen, en zijn de eerste stappen gezet om ook elkaars taal te spreken. Bedankt voor je geduld, voor je extreem snelle feedback en je vertrouwen dat deze samenwerking zou slagen. Ik hoop dat de samenwerking tussen de Jansen groep en de Koenderman groep zich nog lang voort zal zetten.

Ook wil ik **Prof. Rob de Boer, Prof. Jürgen Kuball, Stefan Nierkens, Prof. Joost Sluijter** en **Prof. Dick Vethaak** bedanken voor het beoordelen van mijn proefschrift en het plaatsnemen in de oppositie tijdens mijn verdediging.

Verder wil ik mijn dank uiten aan de partners van het **IMPACT consortium** waarmee ik samen heb mogen werken en in het bijzonder **Freek, George, Heather, Liron, Patrick** en **Ruud**. Bedankt voor de samenwerking tijdens experimenten, jullie input en jullie kritische blik op mijn resultaten.

**Oscar**, bedankt voor de prettige samenwerking bij het transformatie project en de Lorentz meeting. Het was uitdagend om van iemand vanuit een andere invalshoek vragen te krijgen, dat heeft me erg gemotiveerd. Daarnaast heb ik de borrels en etentjes tijdens het bespreken van het project ook als gezellig ervaren.

**Gerjen** and **Rita**, you truly are my 3<sup>rd</sup> and 4<sup>th</sup> paranimphs. Without you my thesis, in this form, would not exist. Seriously 7 of the 8 chapters would not exist. I learned so much from you. I am sure it must have been frustrating from time to time to get an email from me entitled "Matlab error". But you never showed that to me. You were always very patient and answered all my questions. I enjoyed our meetings, our skype calls from Utrecht/Rotterdam/Antwerp/Nijmegen and the night out at the Nijmeegse Vierdaagsefeesten. Also the CYTO congress in Prague was so much fun with you! Gerjen, I hope to see you again at Lowlands or another festival this summer. Rita, we will for sure meet to finish the last submissions, and after that we should celebrate together!

**Erinke**, bedankt voor de super fijne samenwerking 4 jaar lang, met de LPS-BM studie als afsluiter. Je bent een waar organisatorisch en wetenschappelijk wonder. Dank voor je geduld met mijn ietwat minder gestructureerde werkwijze. Het was altijd leuk met jou erbij tijdens de LTI-borrels, halloween feesten, hotpot avond en het congres (en de Amarone!) in Verona. Hopelijk zien we elkaar nog in de toekomst, wellicht in Oxford?

**Na**, thank you for helping me whenever I needed you. Even when you were very busy yourself, you always made time for a chat or a helping hand. And at least as important: I want to thank you for the countless Chinese dumplings, Taiwanese sweets and other food you shared with me, and with Marco. One day I will come to China to taste everything! It was really fun to have you as a colleague and friend. I will always remember the family Christmas celebration with you. I hope to see you again, in China or anywhere else in the world.

**Lilian** en **Roy**, bedankt voor het brengen van wat leven in de brouwerij! In hoeverre dat nou geholpen heeft voor dit proefschrift, weet ik niet zo zeker... Niet alleen als vrienden,

maar ook als collega's zijn jullie super lief en behulpzaam. En jullie plantenkennis oversteeg al mijn verwachtingen. Einde bericht.

**Marwan**, ik wil jou, en ook Mar1, graag bedanken. De samenwerking tijdens de LPS-BM studie verliep heel soepel en je hebt me meermaals uit de brand geholpen, bedankt daarvoor. Daarnaast ben je een gezellige collega en erg grappig, ook zonder dat je het zelf soms door hebt.

**Bart**, bedankt voor de eerste belangrijke stappen die jij gezet hebt richting multidimensionale analyse en de connecties die je met Jos en Twan gemaakt hebt rond de start van de Tour for Life. Ik ben vereerd dat ik beide projecten van jou heb mogen "adopter".

**Corneli, Deon, Lei, Karin, Pien, Pieter** en **Tamar** dank voor de gezelligheid tijdens de lunchpauzes, labuitjes, nieuwjaarsdiners en Sinterklaasvieringen. Ook bedankt voor jullie hulp op het lab als ik weer eens geen idee had waar ik de microtubes of de V-vormige 96-wells platen kon vinden.

**Anne, Emily, Fabienne, Jochem, Margot** en **Mo** bedankt voor de samenwerking bij verschillende projecten en jullie enthousiasme, ik heb veel van jullie geleerd. **Tom**, thanks for the pleasant collaboration and your enthusiasm regarding the microplastics project. You did a great job!

**Twan, Jos, Erik** en **Margot**, tijdens de start van mijn PhD waren jullie er dagen en nachten lang bij. Een goed begin is het halve werk zegt men. Daarvoor wil ik jullie bedanken. Het was hard werken maar vooral ook heel erg gezellig in Italië, Frankrijk, België en Nederland. Ik zal het avontuur van de Tour for Life studie nooit vergeten!

**Kiki**, de lange avonden achter de Aria in het AZU waren super saai maar met jouw aanwezigheid werd er gelukkig ook veel gelachen.

**Andries**, bedankt voor de prettige samenwerking, gedurende mijn gehele PhD, en vooral voor je hulp en input betreffende de toepasbaarheid van ECLIPSE voor het detecteren van zeldzame cellen. Ik hoop dat de samenwerking voortgezet wordt!

**Jeroen** en **Pien**, bedankt voor jullie hulp en ondersteuning als CFF en bij het uitvoeren van technische experimenten. **Jeroen**, ik wil je ook bedanken voor de discussies over compensaties, multidimensionale analyses en gating. We waren het niet altijd met elkaar eens en het was juist goed, ik heb daar veel van geleerd.

Mijn PhD commissieleden **José** en **Michiel**, bedankt voor de jaarlijkse gesprekken en het meedenken over mijn projecten en planning. Dat was zeker niet overbodig, integendeel, het heeft mij gestimuleerd om meer over mijn proefschrift na te denken, ook al leek dat toen nog zo ver weg.

Lieve borrelcommissieleden **Lucas, Maarten, Mariona, Marlot, Marwan, Rianne** en **Nila**, bedankt voor de gezelligheid en de borrels!

Aan de beste kamergenootjes: **Anne-Mieke, Fleurieke, Francesca, Helen, Inêsz, Kuldeep, Cowboy Maarten, Dr. Maarten, Sweet Maarten, Ruben, Sofie**, bedankt voor de rare filmpjes, slechte grappen, het delen van frustratie-momentjes en de discussies tussen het werken door.

Ook wil ik alle lopers van de **UMCU LSDR hardloopclub** bedanken. Ruim 3,5 jaar elke donderdag een uur samen trainen en lol maken, ik kan het iedereen aanraden: even ont-stressen! En in het bijzonder **Willem**, bedankt. Je hebt je geweldig ingezet voor ons cluppie. Mede dankzij jou is mijn plezier in hardlopen nog meer gegroeid.

Aan alle leden van het **Roparun Team 69**: dank jullie wel voor de twee bijzonder mooie en gezellige ervaringen. Op naar de 3<sup>e</sup>!

**Joost, Anton** en **Stefan**, door de onderzoeksmaster Molecular Medicine en mijn masterstage onder jullie begeleiding is mijn wens om een PhD te gaan doen ontstaan. De sfeer in het lab, de interessante discussies en ingewikkelde experimenten vond ik geweldig. Bedankt dat jullie dat met me gedeeld hebben.

Aan alle **collega's van de kindergeneeskunde in het Maasstad ziekenhuis**, bedankt voor het warme welkom in de kliniek. Ik heb ontzettend genoten van een geweldig skiweekendje, fijne kaasfondue-avonden en de collegialiteit op het werk. **Michael** en **Jennifer**, bedankt voor jullie begrip en flexibiliteit rond het afronden van mijn proefschrift en het last minute vrij geven voor een schrijfdag.

Lieve vrienden, jullie hebben allemaal een belangrijk deel uit gemaakt van mijn promotietijd. Bedankt voor jullie interesse in wat ik doe, voor de support en voor het delen van alle mooie momenten en ervaringen waardoor ik de lab-stress helemaal kon vergeten. **Anne, Banafsheh, Bernard, Carina, Emiel, Ilse, Jacco, Jan, Jesse, Julian, Linda, Lisa, Mariska, Marjolijn, Merel, Nienke, Pim, Roos, Rolf, Sara, Sabine, Sam, Sevrin, Simone, Thomas, Tessa, Tonnie, Willem, Wouter, Wytske** jullie zijn geweldig! **Linda en Sabine**, mijn eerste studievriendinnetjes. Voor mijn gevoel stonden jullie aan de wieg van mijn promotietraject, eerst als mede-BMWers voor een jaar. Wat hadden we

een lol tijdens de practica en de Mebiose-borrels. Dat jaartje BMW heeft me gemotiveerd om naast geneeskunde door te gaan in het onderzoek. **Linda**, uit het oog maar zeker niet uit het hart. Toen ik je vertelde dat ik voor deze PhD plek ging solliciteren, stuurde je mij meteen allemaal artikelen toe. En in tijden dat ik het de afgelopen 5 jaar wat moeilijker had, wist ik dat ik altijd met jou kon bellen om mij te helpen mijn gedachten op een rijtje te zetten. Onze vriendschap overleeft de afstand wel, maar ik hoop echt dat we elkaar de komende jaren vaker zullen zien. **Sabine**, bedankt voor de gedeelde klaag-momenten over onderzoeksprojecten en de academie, al dan niet tijdens een hardlooprondje langs de Amstel. Maar bovenal bedankt voor de gezellige avonden in de kroeg of samen op de bank met een glas wijn pratend over alles wat belangrijk is en vooral ook over de onbelangrijke dingen in het leven.

**Wyts**, mijn langste vriendin, ik ben onze ouders zo dankbaar dat zij ons bij elkaar brachten. Een *match made in heaven*. Bedankt voor het regelen van een huisje voor mij in Utrecht tijdens het laatste jaar van mijn promotie, voor de Spotify playlists waarop het schrijven soms ineens een stuk gemakkelijker ging, voor het luisteren naar wat me bezig hield, voor de fijne afleiding in de vorm van de ontelbare festivals, concerten en stapavonden, maar vooral voor de bloedzusterschap die ik altijd voel.

**Sara en Carina**, we begonnen als studiegenoten geneeskunde, werden huisgenoten en super goede vriendinnen in Rotterdam. Twee van ons namen een kijkje in Utrecht, alle drie hebben we even genoten van Amsterdam, maar nu zijn we weer samen in Rotterdam. Dat jullie daar zijn, voelt voor mij echt als thuiskomen. Bedankt voor jullie oprechte interesse in mij ("Maar Sel, hoe gaat het nou echt?!"), de ontelbare etentjes en voor de avonden uit en nachten feesten. Ik ben jullie dankbaar dat jullie er voor me zijn. Ik verheug me er al op om jullie proefschriften te ontvangen!

**Roos en Anne**, samen met Wyts begonnen we als de welbekende *Knysna Hermanus band*, opgericht in Zuid-Afrika. De afgelopen 4 jaar hebben we een mooi tournee gemaakt. Van Porto tot Putten, en van muziekfestivals in Rotterdam tot op de Veluwe, niks was te gek. In die tijd hebben jullie mijn PhD-*struggles* en overwinninkjes van dichtbij meegemaakt. Jullie wisten me nog eens extra te motiveren door een weekend-tripje cadeau te doen als mijn proefschrift af is. Zo fijn dat dat moment nu aangebroken is!

Lieve **Daniël**, mijn broer en paranimf, ik ben me ervan bewust hoe bijzonder het is om jou in mijn leven te hebben en al bijna 29 jaar alles met jou te kunnen delen. Zo ook de voortgang van mijn onderzoeksprojecten. Je hebt abstracts en presentaties van mij gereviewd en kwam altijd met hele leuke en nuttige ideeën als ik even niet wist hoe ik iets wilde verwoorden. Bedankt voor het luisteren, meedenken, inspireren, motiveren en voor

de ontelbare flauwe grappen. Nog niet in brons gebeiteld, maar nu wel zwart op wit: Je bent mijn broer en tevens goede vriend.

Lieve **mama** en **papa**, bedankt dat jullie er voor mij zijn. **Mama**, jij hebt me nooit het gevoel gegeven dat je iets van mij verwachtte. Integendeel, je vroeg vaak juist of ik zeker wist dat ik dit wilde, of ik niet een keer een makkelijkere weg wilde kiezen. Jouw onvoorwaardelijke liefde en vertrouwen voel ik altijd en dat gevoel drijft me om nog net iets meer uit mezelf te halen. **Papa**, je hebt mij al jong geleerd dat het heel belangrijk is om plezier te hebben in je werk. Bij elke keuze rondom studie en werk heb ik dat het zwaarst mee laten wegen en zo ben ik hier gekomen. Bedankt voor het doorgeven van jouw ambitie en super brede interesse, beiden zijn goed van pas gekomen de afgelopen jaren.

Lieve familie: **Oma Elles, Mariëtte, Sandra, Rosa, Lucas, Barbara, Oskar, Marion, Oma Marianne, Inge, Klaas, Susanne, Jan, Wolfram, Myrte, Thomas, René, Anouk** een paar regels zijn te weinig om jullie allemaal te bedanken, ieder van jullie is er op een unieke manier voor mij. Bedankt voor jullie vertrouwen in mij en voor jullie steun in alle mogelijke vormen.

Lieve **Jeanette, Alfred, Lieke, Monique** en **Gerard** jullie hebben elke stap in mijn jonge carrière gevolgd en jullie zijn altijd heel geïnteresseerd om te horen waar ik mee bezig ben. Dat waardeer ik erg. **Alfred**, bedankt voor het ontzettend mooie ontwerp van de omslag van dit boekje. Ik vind het echt bijzonder hoe je met het kleine beetje informatie en het idee wat ik je gegeven heb, zoiets moois hebt gemaakt. **Jeanette**, bedankt voor de fijne en waardevolle gesprekken. Je stelt vaak precies de goede vragen om me aan het denken te zetten.

Ook wil ik alle **Greevens** (en dat zijn er nogal wat) bedanken voor de gezelligheid tijdens verjaardagen, de vakantie op Kreta, koningsmarkten, avondjes uit, festivals, neven- en nichtenweekenden en andere activiteiten die we samen beleefd hebben.

**Dunya** en **Nigiri**, mijn harige monsters, bedankt voor de gezelligheid, de kopjes, jullie zachte vachtjes en de welkome afleiding tussen het schrijven door. Zonder jullie was ik gek geworden, 4 maanden schrijvend thuis. Of is het feit dat ik jullie in mijn dankwoord noem juist het bewijs van gekte...?

Lieve **Marco**, ik weet dat mijn keuzes in de afgelopen 5 jaar niet altijd jouw keuzes zijn geweest, toch heb je me op alle vlakken gesteund. Daar ben ik me zeer bewust van. Je regelde een rustige schrijfplek voor mij in Antwerpen, kookte voor me en zorgde dat al het andere door kon gaan ook al was ik alleen maar aan het werk of aan het schrijven. Jij

wist me precies op de juiste momenten nog even te pushen om verder te gaan, ook al zag ik het zelf even niet zitten. Zonder jou was dit proefschrift er nu niet geweest. Het voelt voor mij niet meer dan logisch dat jij tijdens mijn verdediging aan mijn zijde staat. Met jou is alles beter. Bedankt voor je liefde.





## Curriculum Vitae

Selma van Staveren was born on the 12th of May 1987 in Utrecht. After completing her secondary school at College de Heemlanden in Houten in 2005, she studied Biomedical Sciences for one year at the University of Utrecht. Subsequently she was enrolled in the medical curriculum at the Erasmus University in Rotterdam. Since she still had great interest in science, she followed the research master Molecular Medicine at the Erasmus University in Rotterdam, parallel to her medical studies. In 2010 she received an award for talented master students to finance an extra year of research internship. During her master she completed two research internships. The first internship was under supervision of dr. Marieke von Lindern and Prof. dr. Ivo Touw where she examined the effect of Wnt signalling on DNA damage induced bone marrow failure. During the second internship she studied the initiation of X chromosome inactivation under supervision of Prof. dr. Joost Gribnau. In 2011 she obtained her master's degree in science which was followed by her medical degree in 2014, both at the Erasmus University of Rotterdam.

In 2014 Selma started her PhD on multidimensional analysis methods for flow cytometry at the PhD Programme Infection & Immunity which is part of the Graduate School of Life Sciences of the University of Utrecht. The research was supervised by Prof. dr. Leo Koenderman, Prof. dr. Pim de Voogt and co-promotors dr. Nienke Vrisekoop en dr. Jeroen Jansen. The main focus of her PhD was the development of a multidimensional flow cytometry analysis method for fast, objective and automatic identification of disease-specific (rare) cells. Next to this, she also examined the effect of microplastics on human health. Pitching about the latter subject, she won the audience award during the Dutch Finals of the Famelab Pitch contest. Thereafter she was invited as a speaker during the European Researchers' Night 2017 at the Natural museum of History in London. Also during this event she won the audience award for best pitch. The results of her PhD project have been described in this thesis.

In 2018 Selma returned to Rotterdam to work as a paediatric resident at the Maasstad Ziekenhuis in Rotterdam.



## List of publications

### Published

Barakat TS, Loos F, van Staveren S, Myronova E, Ghazvini M, Grootegoed J, Gribnau J. The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. *Mol Cell*. 2014;53(6). doi:10.1016/j.molcel.2014.02.006

ten Haaf T, van Staveren S, Oudenhoven E, Piacentini MF, Meeusen R, Roelands B, Koenderman L, Daanen HAM, Foster C, de Koning JJ. Prediction of functional overreaching from subjective fatigue and readiness to train after only 3 days of cycling. *Int J Sports Physiol Perform*. 2017;12. doi:10.1123/ijsspp.2016-0404

Tinnevelt GH, Kokla M, Hilvering B, van Staveren S, Folcarelli R, Xue L, Bloem AC, Koenderman L, Buydens LMC, Jansen JJ. Novel data analysis method for multicolour flow cytometry links variability of multiple markers on single cells to a clinical phenotype. *Sci Rep*. 2017;7(1):1-11. doi:10.1038/s41598-017-05714-1

de Ruiter K, van Staveren S, Hilvering B, Knol E, Vrisekoop N, Koenderman L, Yazdanbakhsh M. A field-applicable method for flow cytometric analysis of granulocyte activation: Cryopreservation of fixed granulocytes. *Cytom Part A*. 2018;93(5). doi:10.1002/cyto.a.23354  
Folcarelli R, van Staveren S, Bouman R, et al. Automated flow cytometric identification of disease-specific cells by the ECLIPSE algorithm. *Sci Rep*. 2018;8(1):1-18. doi:10.1038/s41598-018-29367-w

ten Haaf T, van Staveren S, Iannetta D, Roelands B, Meeusen R, Piacentini MF, Foster C, Koenderman L, Daanen HAM, de Koning JJ. Changes in choice reaction time during and after 8 days exhaustive cycling are not related to changes in physical performance. *Int J Sports Physiol Perform*. 2018;13(4). doi:10.1123/ijsspp.2017-0218

van Staveren S, ten Haaf T, Klöpping M, et al. Multi-dimensional flow cytometry analysis reveals increasing changes in the systemic neutrophil compartment during seven consecutive days of endurance exercise. *PLoS One*. 2018;13(10). doi:10.1371/journal.pone.0206175

ten Haaf T, Foster C, Meeusen R, Roelands B, Piacentini MF, van Staveren S et al. Submaximal heart rate seems inadequate to prescribe and monitor intensified training. *Eur J Sport Sci*. 2019. doi:10.1080/17461391.2019.1571112

Giustarini G, Vrisekoop N, Kruijssen L, Wagenaar L, van Staveren S et al. Trovafloxacin-Induced Liver Injury: Lack in Regulation of Inflammation by Inhibition of Nucleotide Release and Neutrophil Movement. *Toxicol Sci*. 2019. doi:10.1093/toxsci/kfy244

### Accepted manuscripts

Tinnevelt GH, van Staveren S, Wouters K, Wijnands E, Verboven K, Folcarelli R, Koenderman L, Buydens LMC, and Jansen JJ. A novel data fusion method for the effective analysis of multiple panels of flow cytometry data. *Sci Rep*. 2019

### Submitted manuscripts

van Staveren S\*, Folcarelli R\*, Tinnevelt GH, Cadot E, Vrisekoop N, Buydens LMC, Koenderman L, Jansen JJ#, Oscar van den Brink#. Transformation of flow cytometry data with OTflow: A necessity for multivariate analysis.

van Staveren S\*, Folcarelli R\*, Bloem AC, Postma G, Tinnevelt GH, Vrisekoop N, Buydens LMC, Jansen JJ#, Koenderman L#. Transformation of flow cytometry data with OTflow: A necessity for multivariate analysis.

Folcarelli F, Tinnevelt GH, Hilvering B, Wouters K, van Staveren S, Postma GJ, Vrisekoop N, Buydens LMC, Koendeman L, Jansen JJ. Multi-set pre-processing of Multicolour Flow Cytometry Data.

\* or #: These authors contributed equally to this manuscript

