Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/67101 holds various files of this Leiden University dissertation.

Author: Unen, V. van
Title: Mucosal immunology revisited through mass cytometry : from biology to bioinformatics and back
Issue Date: 2018-11-27

CHAPTER 1

General Introduction

The immune system

The immune system protects the body against bacterial, viral, and parasitic infections. It is one of the most dynamic and plastic systems in the human body, present in nearly every tissue type. Although the immune system is generally highly efficient, disruptions of homeostasis can occur leading to a variety of immune-mediated diseases, such as diabetes mellitus type 1 or Crohn's disease. The immune system can be divided into an innate and an adaptive compartment, and their cellular components consist of a variety of different cell populations. The innate immune system represents the first line of defense which comprises of the myeloid cell lineage, including antigen presenting cells, such as macrophages and dendritic cells, and innate lymphocytes, such as natural killer (NK) cells and innate lymphoid cells (ILCs). The adaptive immune system is more specialized and typified by the expression of antigen-specific receptors on B and T lymphocytes, the latter including $CD4^+$ T cells and $CD8^+$ T cells, and the capacity to develop immunological memory providing superior protection towards pathogens. In addition, we can distinguish innate-like unconventional T cell subsets that reside more prominently at barrier sites, including TCR$\gamma\delta$ cells, NKT cells and mucosal-associated invariant T (MAIT) cells[1]. Immune processes are mediated by the crosstalk between these types of cells, tissue-resident as well as circulating immune cells, all interacting in specific micro-environmental contexts. Each of these immune cell types can be phenotypically defined through the expression of specific proteins on the cell surface, referred to as markers hereafter.

Flow cytometry

In the past decades, flow cytometry has been the benchmark technique to analyze markers expressed by individual immune cells, revealing their identity. By selecting a set of fluorochrome-conjugated antibodies that specifically recognize certain markers expressed on the cell surface, subpopulations of interest can be analyzed by measuring the fluorescent emission at the single-cell level. This has allowed us to gain a wide understanding of the composition of the immune system in health and disease[2]. However, the primary drawback of traditional flow cytometry is that the number of markers that can simultaneously be measured is limited by spectral overlap (generally 8-12) (**Figure 1A**). In addition, the design of complex flow

cytometry antibody panels is laborious and challenging due to the spectral overlap. To capture all cell populations across all major innate and adaptive lineages of the immune system simultaneously, a higher level of multiparametric analysis of single cells is required that cannot currently be met with fluorescence-based technologies.

Mass cytometry

Mass cytometry (CyTOF, cytometry by time-of-flight) is a new generation of single-cell analysis technology offering a high-throughput platform for robust characterization of immune cells by overcoming this flow cytometry-intrinsic marker limitation[3]. The CyTOF is a mass-spectrometer-flow cytometer hybrid which analyzes antibodies conjugated with heavy metal isotopes instead of fluorescent reporters[4,5]. It is thereby unhampered by interference from spectral overlap. The CyTOF allows the detection of currently up to 42 markers simultaneously, approximately 3-fold more than with traditional flow cytometry, and with much sharper peaks and less crosstalk between channels (**Figure 1B**). Theoretically, the number of markers detectable with mass cytometry could increase to about a 100, once the isotopic purification of metals and the chemical conjugation of metals to antibodies are improved, expected in the near future. In addition, flow cytometry measurements can be hindered by autofluorescence, the natural fluorescence that occurs in cells. However, since heavy metals do not occur in biological systems, mass cytometry has a strongly reduced biological background compared to flow cytometry, and is unhampered by autofluorescence. Moreover, mass cytometry has

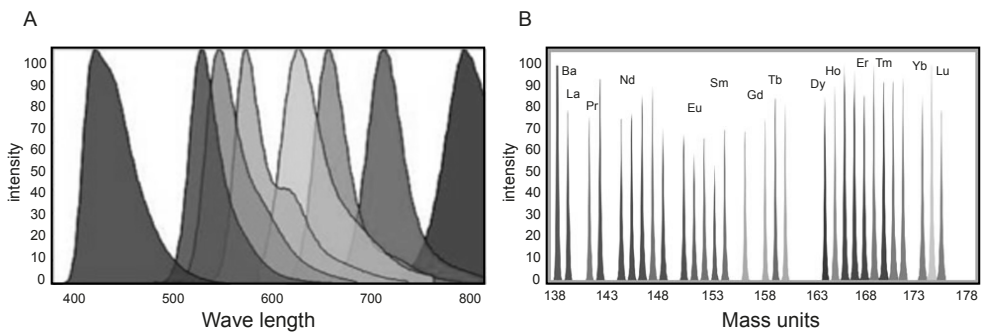

Fig. 1. Detection of antibody reporters by flow cytometry and mass cytometry (**A**) Example of emission spectra of several fluorescent reporters with flow cytometry. (**B**) Mass spectrum of 30 purified heavy metal isotopes with mass cytometry. Adapted from Maecker et al. (2015) and Tanner et al. (2013).

a minimal variation in the intensity of the individual metal reporters, unlike flow cytometry. A schematic overview of the mass cytometry workflow[6] is depicted in **Figure 2**. Mass cytometry has therefore the ability to measure dozens of markers

on millions of cells collected within a reasonable timeframe (1 million cells per hour). Disadvantages of mass cytometry compared to flow cytometry are that the metal sensitivity is lower than the brightest fluorochromes (detection limit of 350 antibodies per cell), it has a lower throughput (300 cells per second), the sample recovery is incomplete (60%), there is no ability to recover viable analyzed cells (i.e., cell sorting) and light-based measures such as forward and side scatter cannot be measured. Nevertheless, mass cytometry is now widely adopted as it offers an unprecedented resolution in the analysis of cellular diversity of the immune system.

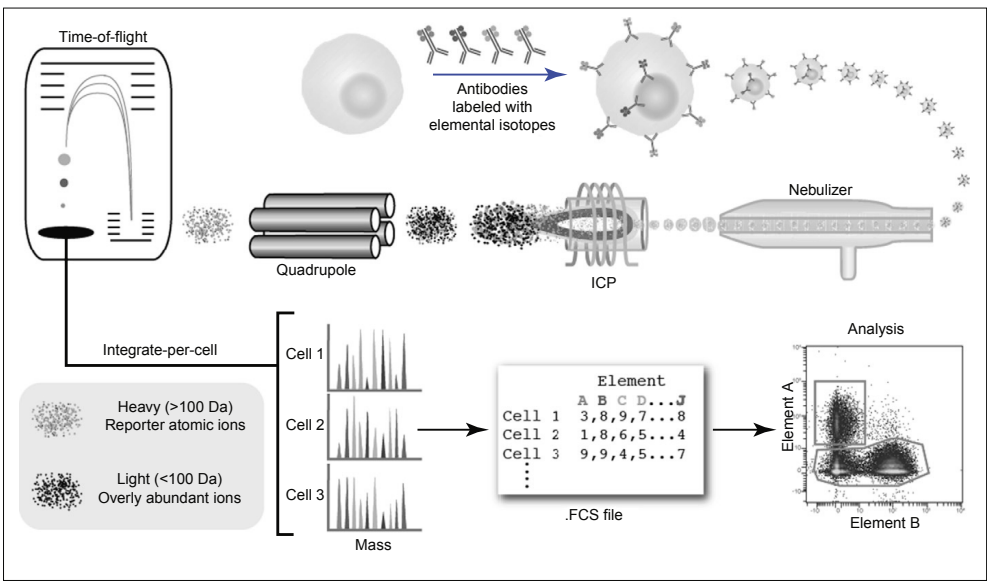

Fig. 2. Mass cytometry allows single-cell quantification of heavy metal isotope reporters A single-cell suspension is labeled with heavy metal isotope-conjugated antibodies, followed by introduction into the nebulizer where it is aerosolized. The aerosol droplets are directed into the inductively coupled plasma (ICP) torch where the cells are vaporized, atomized and ionized. Low mass ions are removed in the Quadrupole, resulting in a cloud of ions enriched for the heavy metal isotopes. The ion cloud then enters the Time-of-Flight (TOF) chamber where the ions are separated on the basis of their mass to charge ratio as they accelerate towards the detector. The time-resolved detector thus measures a mass spectrum that represents the identity and quantity of each metal ion on a single-cell basis. This is thus proportional to the number of antibodies originally bound per cell. Data is generated in .fcs format and analyzed in third-party software programs. Adapted from Bendall et al. (2012).

Data analysis

Conventional approaches for flow cytometry data analysis typically rely on the manual interpretation of a large number of 2-dimensional plots by selecting subsets of interest from parent populations, a strategy called 'gating'. The high number of measurable single-cell markers with mass cytometry, however, brought a daunting

increase in the complexity of the data. With every additional marker simultaneously analyzed the amount of information obtained increase exponentially. Hypothetically, if a cell type would be characterized by the presence or absence of a given marker, even ignoring its quantitative nature of expression, a 40-antibody panel would yield 240 (1 trillion!) potential combinations of marker expression profiles. Even if one would go through all the 2-dimensional plots that could be generated, this does not reveal multidimensional relationships and single-cell correlations. In addition, it has been shown by various multicenter studies, such as the Human Immune Genome Project, that manual gating is one of the largest variables in the outcome of flow cytometry-based experiments[7,8]. Therefore, such approaches are not scalable in the context of high-parametric marker expression data across millions of cells, suffer from individual user bias and require prior knowledge of the cell type of interest. Consequently, bioinformatics tools are required to extract relevant information from the generated high-dimensional datasets. Many algorithmic methods have recently been described to facilitate the analysis of mass cytometry data in an unbiased manner[7], and these can roughly be divided into either being clustering-based or dimensionality reduction-based method.

An example of a clustering-based method specifically developed for mass cytometry data analysis is SPADE, unsupervised hierarchical clustering with minimum spanning tree projection[9] (**Figure 3A**). With SPADE, cells are grouped into a pre-defined number of nodes based on phenotypic similarity for all markers simultaneously, and depicted in a dendrogram displaying the corresponding higher-order relatedness between those nodes. While SPADE provides an overview of the heterogeneity and the relatedness of cell populations of the immune system, it does not allow analysis at the single-cell level. Consequently, rare cells are difficult to visualize with the SPADE analysis.

The t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis[10,11] is a nonlinear dimensionality reduction method and has recently been implemented for the analysis of mass cytometry data[12]. This method generates a 2-dimensional map where cells with similar multidimensional phenotypes are placed close to each other, while maintaining single-cell resolution (**Figure 3B**). It does so by taking all marker expressions into account simultaneously. Unlike principal component analysis (PCA), t-SNE effectively captures nonlinear relationships in the high-dimensional data, thereby preserving subpopulations of cell types with subtle differences in marker expression profiles. The presence or absence of multiple markers reveal

each cell's subset identity and result in a robust characterization of cell phenotypes that is captured and visualized in a single 2-dimensional t-SNE map. Notably, t-SNE dimensionality reduction alone does not assign cells to groups. Therefore, this approach was extended by introducing ACCENSE[13], a tool that rather than by conventional manual gating, identifies phenotypically distinct subsets based on the t-SNE map using a density-peak algorithm (**Figure 3C**).

However, while non-linear methods like t-SNE do retain local data structure with single-cell resolution, they are limited by the number of cells that can be analyzed. In cytometry studies, this poses a problem, as datasets usually contain information on millions of cells. Therefore, substantial numbers of cells needs to be removed by random downsampling to make dimensionality reduction computationally feasible and reliable.

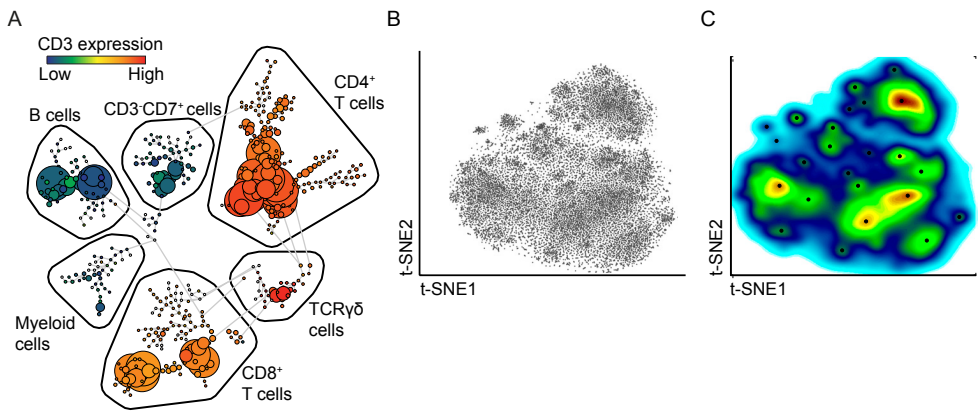

Figure 3. Computational tools implemented for mass cytometry data analysis (**A**) A SPADE tree analysis of a peripheral blood sample. Size of the nodes is proportional to the respective number of clustered cells. Color bar represents CD3 marker expression. (**B**) A t-SNE map showing murine CD8⁺ T cells. Each black dot represents a single cell. (**C**) A composite map depicting the local probability density of cells as embedded in panel B. Black dots represent centers of phenotypic subpopulations and were identified using a standard peak-detection algorithm. Color represents cell density. Adapted from van Unen et al. (2016) and Shekhar et al. (2013).

Intestinal immune system and pathology

The intestine contains a single epithelial cell layer separating the external environment or lumen from the underlying tissue, and represents the largest compartment of the immune system[14]. It has the task to provide protection against pathogens yet remaining tolerant to unharmful microbiota. To accomplish this, a complex intestinal immune system has evolved existing in homeostasis with the

microbiota. In the intestine, there is a substantial amount of organized lymphoid tissue and it contains large populations of innate and adaptive immune cells. It is continuously exposed to foreign antigens and other environmental antigens from the diet and the microbiota. The intestinal immune system therefore has the task to discriminate between harmful and beneficial antigens. This role is particularly impressive when one considers the vast mucosal surface and changing landscape along the gastrointestinal tract that the intestinal immune system must monitor. For example, the small intestine, whose primary function is in nutrient digestion and absorption, is anatomically highly distinct from the colon, whose function is in water absorption and final stages of the digestive process. This results in a gradual change in distribution of environmental factors along the length of the intestine[14]. Therefore, recent evidence demonstrated a proficient regional specialization within the intestinal immune system along the gastrointestinal tract[14,15].

Although strong protective immunity is essential to prevent invasion by pathogens, equivalent immune responses against dietary proteins or commensal bacteria can lead to chronic diseases. A complex interplay of regulatory mechanisms and continual crosstalk between various immune cell types normally prevents such unwanted responses. A disturbance of intestinal tolerance can therefore rewire immune cell composition and functionality, resulting in chronic inflammation and increased risk for infection.

Although the role of several immune subsets in driving intestinal pathology has been established, a system-wide approach that simultaneously interrogates all major lineages on a single-cell basis was lacking. High-dimensional mass cytometry is a powerful tool for dissecting the entire immune landscape, given that the accompanied challenges for efficient data analysis have been overcome. Bettering our understanding of the cellular composition of the immune system in immune-mediated diseases is key to obtain mechanistic insight and develop improved diagnostics and targeted therapeutic approaches.

Outline

At the start of my PhD project in 2013, mass cytometry was not yet implemented in the Netherlands. I was given the unique opportunity to implement high-dimensional mass cytometry as the first adopter in the Netherlands. We were the first to apply mass cytometry to the analysis of the composition of the immune system in biopsy material from patients with a variety of inflammatory intestinal diseases (**Chapter**

2). This revealed previously unrecognized immune cell heterogeneity, and identified immune subsets associated with the pathogenesis of complex intestinal disorders.

The large datasets generated with mass cytometry, however, posed a problem for proper data handling and analysis. Here, I have worked closely together with computational scientists at the LUMC and at the Technical University Delft for the development of appropriate tools for analysis and visualization of the large datasets generated. This has resulted in the development of Cytosplore **(Chapter 3**) and Hierarchical Stochastic Neighbour Embedding (HSNE) (**Chapter 4**), the latter a computational approach that overcomes the scalability limits of t-SNE-type approaches, thus enabling the analysis of tens of millions of cells without the need for subsampling of the data.

With these improved computational tools, we applied mass cytometry to biopsy material along the gastrointestinal tract from undiagnosed and untreated patients with inflammatory bowel disease (**Chapter 5**). This resulted in the stratification of immune cell infiltrate types in inflammatory bowel disease.

Finally, by visualizing the dynamics of the t-SNE computation over time, our research revealed extensive heterogeneity as well as multi-lineage differentiation trajectories of ILCs, important regulators of tissue integrity, in the human fetal intestine (**Chapter 6)**. This study was one of the first where exploratory mass cytometry-based approaches to determine heterogeneity in the immune system were translated into functional analyses of newly identified immune subsets.

In **Chapter 7** the major findings of this thesis are discussed in the light of the current literature and the availability of imaging mass cytometry, allowing the simultaneous analysis of 36 markers on tissue sections at subcellular resolution. The implications of this for future research are discussed.

References

1 Fan, X. & Rudensky, A. Y. Hallmarks of Tissue-Resident Lymphocytes. Cell 164, 1198-1211, doi:10.1016/j.cell.2016.02.048 (2016).

2 Simoni, Y., Chng, M. H. Y., Li, S., Fehlings, M. & Newell, E. W. Mass cytometry: a powerful tool for dissecting the immune landscape. Curr Opin Immunol 51, 187-196, doi:10.1016/j.coi.2018.03.023 (2018).

3 Bandura, D. R. et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem 81, 6813-6822, doi:10.1021/ac901049w (2009).

4 Tanner, S. D., Baranov, V. I., Ornatsky, O. I., Bandura, D. R. & George, T. C. An introduction to mass cytometry: fundamentals and applications. Cancer Immunol Immunother 62, 955-965, doi:10.1007/s00262-013-1416-8 (2013).

5 Maecker, H. T. & Harari, A. Immune monitoring technology primer: flow and mass cytometry. J Immunother Cancer 3, 44, doi:10.1186/s40425-015-0085-x (2015).

6 Bendall, S. C., Nolan, G. P., Roederer, M. & Chattopadhyay, P. K. A deep profiler's guide to cytometry. Trends Immunol 33, 323-332, doi:10.1016/j.it.2012.02.010 (2012).

7 Mair, F. et al. The end of gating? An introduction to automated analysis of high dimensional cytometry data. Eur J Immunol 46, 34-43, doi:10.1002/eji.201545774 (2016).

8 Maecker, H. T., McCoy, J. P. & Nussenblatt, R. Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol 12, 191-200, doi:10.1038/nri3158 (2012).

9 Qiu, P. et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol 29, 886-891, doi:10.1038/nbt.1991 (2011).

10 van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. J Mach Learn Res 15, 3221-3245 (2014).

11 van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. J Mach Learn Res 9, 2579-2605 (2008).

12 Amir el, A. D. et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol 31, 545-552, doi:10.1038/nbt.2594 (2013).

13 Shekhar, K., Brodin, P., Davis, M. M. & Chakraborty, A. K. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). Proc Natl Acad Sci U S A 111, 202-207, doi:10.1073/pnas.1321405111 (2014).

14 Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. Nat Rev Immunol 14, 667-685, doi:DOI 10.1038/nri3738 (2014).

15 Agace, W. W. & McCoy, K. D. Regionalized Development and Maintenance of the Intestinal Adaptive Immune Landscape. Immunity 46, 532-548, doi:10.1016/j.immuni.2017.04.004 (2017).