UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Centre hospitalier
universitaire vaudois

Mémoire de Maîtrise en médecine no 6879

Functional Analysis of Immunocompromised Patients' Leucocytes by Single-cell Mass Cytometry

Etudiant
Porret Raphaël

Tuteur
Prof. Pantaleo Giuseppe
Service d'immunologie et allergie, CHUV

Co-tuteur
Dr. Fenwick Craig
Service d'immunologie et allergie, CHUV

Expert
Prof. Münz Christian
Institute of experimental immunology, USZ

Lausanne, 31.01.2020

Raphaël Porret 31.01.2020
Medical School, University of Lausanne, Switzerland

Functional analysis of immunocompromised patients' leucocytes by single-cell mass cytometry

Author : Raphaël Porret
Supervisors: Prof Giuseppe Pantaleo & Dr Craig Fenwick
Service of Immunology and Allergy – IAL, CHUV
Expert: Prof Christian Münz
Institute of Experimental Immunology, USZ

Abstract

Immunodeficiencies make up a large group of diseases characterized by heterogeneous clinical manifestations, including life-threatening infections, autoimmunity, chronic inflammation, allergy and malignant diseases. They are classically divided in primary (PID) and secondary (SID) immunodeficiencies and they can be caused by monogenic defects or be secondary to exogenous factors, malignant or non-malignant diseases. In the last 20 years, accelerating progress has been made in identifying new forms of PIDs thanks to the advances of molecular and genetic characterizations. These disorders are either diagnosed early in life or even later, in adults. It is estimated that 1-2% of the population might be affected with any type of the whole PID spectrum.

Immune cell characterization, particularly by flow cytometry techniques, has extensively showed its importance in the clinical management of patients presenting immune deficiencies with quantitative cell defects, as well as in the understanding of the immune system. It has already improved the classification of immunological diseases, as well as contributed to improve treatment efficacy and follow-up. Recently, mass cytometry techniques have been used for diagnostic purposes, significantly increasing the breadth and depth of the functional and phenotypic characterization of a patient's immune cells, in comparison to traditional flow cytometry techniques. These advancements are driven by the great increase in measurable parameters provided by mass cytometry, which allows for all major known immune cell populations and subpopulations to be characterized with a single analysis.

The major contribution of this research resides in directly testing the functional activity and response of a patient's immune cells to different stimuli. The highly multiparametric nature of mass cytometry allows for both a broad and in depth characterization of the functional immune response using only a minimal volume of a patient's blood (1 mL) with results available within one day, thus drastically improving time to diagnosis. In addition to having a proportional and phenotypic characterization of a patient's immune cells, identifying the functionally abnormal cell population(s) will provide the clinicians with an even better understanding of their patient's immunological defect. Interpretation of the mass cytometry results along with the patient's clinical data will allow for the identification of signatures associated with specific immunological defects, new classes of immunodeficiencies and therapies that are best adapted to a specific class of an immunological disorder, hence improving the diagnosis and the benefits for immunocompromised patients.

Key words: immunology, immunodeficiency, cytometry, bioinformatics, molecular signatures

UNIVERSITY OF LAUSANNE

MASTER'S THESIS

SERVICE OF IMMUNOLOGY AND ALLERGY - IAL, CHUV
LAUSANNE, SWITZERLAND

Functional Analysis of Immunocompromised Patients' Leucocytes by Single-cell Mass Cytometry

Author
Raphaël PORRET

Supervisors
Prof. Giuseppe PANTALEO
Dr. Craig FENWICK

Expert
Prof. Christian MÜNZ

January 31, 2020

UNIL | Université de Lausanne
Faculty of Biology and Medicine

Contents

1 Introduction

"Intellectual work is an act of creation. It is as if the mental image that is studied over a period of time were to sprout appendages like an ameba - outgrowths that extend in all directions while avoiding one obstacle after another - before interdigitating with related ideas."

— Santiago Ramon y Cajal

This master's thesis focuses on providing a new in depth characterization of the human immune system in health and diseased states. The complexity of nature governing us pushes research to be more precise and have an increased depth in its observation in order to increase our comprehension of life. The approach presented in this work employs advanced analytical laboratory procedures and data analysis methods that are applied to our current understanding of the immune system. The goal is to better characterize immunodeficiencies at the molecular level of a patient's immune cells with the final goal of offering better diagnosis tools that could lead to an improved clinical management of patients through more appropriate treatments and follow-up care.

This project aims to develop a mass cytometry profiling assay for single-cell functional analysis of immune cells from patients at the CHUV with suspected immunodeficiencies. This analysis will complement the phenotypic immune cell characterization currently performed by the CHUV-IAL diagnostics lab and provide a distinct readout of a patient's functional immune profile. In addition to the evaluation of immune cell lineage markers, chemokine receptors and activation markers, cell specific expression of different cytokines will provide an important readout of either ongoing immune activation or the capacity of a patient's immune cells to respond to stimuli that include Pathogen Associated Molecular Patterns (PAMPs). These functional responses will be compared to those of healthy donors to allow for a comparative bioinformatics analysis of potential immunological defects. Initial analysis focused globally on leucocytes, followed by a directed analysis of myeloid dendritic cells to provide insights into the strength of this approach. It is anticipated that this analysis will identify signatures and highlight molecular mechanisms associated with specific immune cells in health and diseased states. It will thus provide the basis for a strong analysis pipeline to be used in further translational research work. To give an insight of molecular immunological profiles that can be detected with this analysis pipeline, a group of 15 clinically immunocompromised patients were compared to a group of 15 healthy donors.

This chapter will first give a general overlook of the immune system and immunodeficiencies. Finally, the last part of this chapter will give an insight into the different technologies available and used in this work for in-depth characterization of immune cells. These advanced technological progresses allow for increased insights into the determination of cells' properties, including cellular gene regulation and intracellular molecular signalling pathways.

1.1 The Immune System

Having a vital function in all other organs, the immune system ensures the essential function of maintaining the human body's homeostasis, positioning itself at the convergence point of systemic regulation. Indeed, being able to distinguish the self from the

non-self or from the abnormal self allows a thorough defence against microorganisms, foreign substances, products of damaged cells and abnormal cells growths (rejection of tumours). What's more, while fulfilling this task the immune system should ignore normal healthy cells, harmless molecules or symbiotic microorganisms, which would otherwise lead to unwanted consequences.

In order to achieve its function, the immune system employs numerous actors and effectors that play key roles as biological guardian of the organism. As it will be discussed below, different cells, proteins (e.g. signalling molecules, receptors, transcription factors), physical (e.g. epithelia, mucosa) and chemical (e.g. low pH) barriers are of an extreme importance to guarantee an adequate response against infections and in rejections of tumours. It is important to mention that there is a constant battle driven by the immune system in order to counter the harmful infections continually present in the environment. However, for immunologically healthy hosts, persistent signs of inflammation are uncommon and disappear as soon as the foreign agent is removed.

Thereby, there are clinically three main host's manifestations of a dysfunctional immune system, that are immunodeficiency, autoimmunity, or allergy. Each of those can be explained by an unbalance in the physiological immune homeostasis, either by a defect in the activation of the system, or by an overactivation of it.

1.1.1 Overview

The main characteristics of the immune system is that it is systemic, so that its actors can evoke mechanisms of host defence in any tissues, migrating through blood or lymphatic vessels. The innate immunity is in the first place to detect the initial triggers, and is also phylogenetically the oldest guard of plants and insects. It will then initiate the immune assault, while activating in parallel the adaptive immunity that will act in a second phase within days after the infection begins. This more complex mechanism of defence is more specialized and requires much more sophisticated cells and tissue structures (receptors, antibodies, specialized lymphoid tissues, cytokines), which have evolved only in vertebrates. The advantage of an effective, powerful and safe immune response is to have not only positive signals that amplify the response, but also to have control mechanisms. Indeed, these should prevent a loss of control which would drive to self injuries either by the extent of the reaction generated or by setting of self antigens recognitions to destruction. The homeostasis needed to guarantee a fair balance between the acceleration of an immune response and the brake mechanisms to avoid collateral damage is fine and has to be conscientiously regulated.

Figure 1 offers a general overview of the interactions between the innate and the adaptive immunity, how they are acting in respect to each other. Some of the most important steps of this coordinated response will be further explain in the following chapters.

The main actors involved in the immune response are briefly presented and described in *figure* 2. This work will mainly focus on the cellular elements of the immune response, with their associated cytokines and chemokines, whose function is explained below. Almost all of the circulating cells of the immune system are derived from hematopoietic stem cells (HSCs) in the bone marrow. Their differentiation and regulation rely on growth factors and cytokines stimulations (1).

In order to achieve a coordinated and well regulated immune response, immune cells interact with one another and with other host cells by secreting cytokines. This family of proteins is able to act on the growth and differentiation potential of cells, to activate them as well as to ensure their migration among tissues through the blood circulation or lymphatic vessels. The major proinflammatory cytokines of innate immunity are TNF, IL-1, IL-6, IL-12, IL-15, IL-18 and IFN-γ. The major anti-inflammatory cytokines are IL-10 and TGF-β. Their functions and characteristics can be found in *figure* 3.

1.1.2 Innate Immunity

The innate immunity, or native immunity, is able to develop within hours after the infection occurs. Antimicrobial peptides (AMPs), complement molecules and sentinel cells, such as macrophages, dendritic cells (DCs) or mast cells, are located in the blood, in the different epithelia or mucosa surfaces of the organism or in any other organs to respond directly when pathogenic microorganisms break the barrier and invade the host. Invading microorganisms are directly detected through some of their repeated patterns shared by classes of microbes, called pathogen-associated molecular pattern molecules (PAMPs). Damaged cells are detected through the repeated and conserved pattern of endogenous molecules released, called damage-associated molecular pattern molecules (DAMPs). The fact that those receptors are constitutive (germline encoded in all cells), they exist before that the infection takes place, and do not need a primary contact with an invader to be effective (unlike adaptive immunity receptors). This allows to trigger a fast and effective response that is sufficient to eradicate most of the infections the organism is confronted with. However, they do not allow a fine recognition of different microbes and see a limit in its efficacy by more virulent pathogens. Furthermore, their limited diversity is illustrated by the less than 100 of different types of invariant receptors that have been discovered in this family of pathogen recognition receptors (PRRs), while the two receptors present in the adaptive immunity can contain up to millions of variations of each and recognize more than 10 million different microbial molecules. That is why the adaptive system is also essential in the defence against microorganisms..

The different PRRs, their ligands, the different innate immune cells, with their characteristics, as well as blood proteins playing a significant role in the innate immunity response will now be discussed. It is also important to mention that innate immunity contains a lot of redundancy, and that many actors have the same effect. This protects against microbes that achieve to block one defence pathway, as other mechanisms of defence will be activated and the microorganisms will be nevertheless combated.

Pattern recognition receptors (PRRs)

PRRs are fundamental receptors of different structures recognizing dangerous antigens (PAMPs, DAMPs) and triggering an immune response. They are located on the surface or in the cytoplasm of the cells, as well as in the blood circulation as soluble molecules. The subsequent inflammation response elicited induces the release of different mediators (e.g. cytokines, chemokines, small molecules mediators), which activate innate immunity effectors. Different pathologies can arise when a dysfunction is present at their antigen interface or in their downstream signalling pathways. A summary of the most important PRR is given in *figure* 4.

PRRs ligands can be from different natures. PAMPs can be mainly from four different molecular structures: nucleic acids (e.g. single-stranded or double-stranded RNA, unmethylated CpG DNA sequences), proteins or peptides (e.g. pilin and flagellin), lipids (e.g. LPS, lipoteichoic acid), and carbohydrates (e.g. mannan, glucans). PAMPs usually are life-essential molecules of the microbes and thus less susceptible to mutation. This is a reason why innate immunity has achieved to be conserved among the evolution. DAMPs (also called alarmins) come from endogenous damaged cells and can also be from different nature. Those molecular patterns are summarized in *figure* 5.

Complement system

The complement system is an important defence mechanism of humoral innate immunity. Moreover, plasma proteins recognize microbial surface structures and can be activated through three different pathways. Zymogens, which are proteases precursors, are activated through the proteolytic activity of other proteases and thus trigger a cascade that amplify itself at each step. During this process, different effector molecules are generated. Those are able to opsonize microbial products, to act as chemoattractants recruiting leukocytes, or to directly kill the microorganism. The complement activation pathways are summarized in *figure* 6.

Hematopoietic stem cell-derived innate immune cells

Innate immune cells are mostly derived from haematopoietic myeloid or lymphoid progenitors in the bone marrow, whereas adaptive immune cells principally come from lymphoid haematopoietic progenitors. Their major subsets are illustrated in *figure* 7, highlighting their key functions. The diversity of functions and abilities of those different cells are complementary in order to cover a wide variety of pathogens and to be able to defend the host effectively against them.

Other immune actors present in blood also play an important role in the defence against foreign organisms. Those are for example lipid mediators, peptides, amines, nitric oxide, adhesion molecules (integrins, selectins) or acute phase proteins (CRP). As this work will not focus on those molecules, their attribute and functions will not be discussed here.

1.1.3 Adaptive Immunity

Innate immunity can show some weaknesses facing some more virulent pathogens, due to its poor diversity of patterns recognition. However, in those situations it can already slow down the invasion and initiate the adaptive response, which will most of the time be able to control and overcome a strong infection. Thereby, the adaptive immunity, or acquired immunity, distinguishes itself from the innate immunity through its increased capabilities of defence against microbes and nonmicrobial products. The two main cells families constituting and coordinating this more flexible system are the T and B lymphocytes, both descending from the common lymphoid progenitor cell of the bone marrow.

Cardinal features of the adaptive immunity

The central features that make the adaptive immunity unique are firstly its specificity and diversity. Firstly, the great variety of antigen epitopes recognized by the two

main adaptive receptors (BCR and TCR) is made possible by the vast repertoire of immature lymphocytes. Each of them presents at its surface a unique rearranged receptor that corresponds to a specific pathogens antigen. Once this antigen is recognized, the corresponding immature lymphocyte will undergo clonal expansion and thus provide protection against this pathogen.

Secondly, another key feature of adaptive immunity is its ability to create a immune long-lived memory against pathogens it has already been exposed to. Each time a pathogen is seen by the system, it increases the promptness, its magnitude and the sensitivity of its response. This improves the efficacy of responses against microbes that often come across the immune system. Thirdly, the last main vital characteristic of this system is its self tolerance. Self-controlled mechanisms get rid of the lymphocytes that build responses against antigens belonging to the own self. An inappropriate preservation of self-tolerance is the first step to a wide variety of autoimmune diseases.

Antigen presentation

One of the most important mechanism which the adaptive immune system relies on is the presentation of antigens by antigen-presenting cells (APCs) to lymphocytes. Indeed, this is a way that front-line cells or any other cells of the organism have to communicate with the immune cells and inform them about the protein content of their cytosol or extracellular environment. Thanks to this method, the immune cells constantly have an overview over the state of each cell, being thus able to detect if a cell gets infected, presents malignant changes or has sampled pathogens from its surroundings. The many different functions of antigen-presenting cells are broad and will not be discussed in details in this work. However, the main molecules expressed at their surface allowing antigen presentation are part of the major histocompatibility complex (MHC or HLA) family. This includes the class I MHC (containing HLA-A, HLA-B and HLA-C), synthesized and expressed on all nucleated cells, and the class II MHC (consisting on HLA-DR, HLA-DQ and HLA-DP), which are constitutively expressed only on dendritic cells, B lymphocytes, macrophages, thymic epithelial cells, and a few other cell types.

Other HLA-independent presentation molecules include the CD1 molecule (presenting glycolipid components and glycosphingolipids, recognized by NKT cells), MHC class I-related chains A (MICA) and B (MICB) recognized by $\gamma\delta$T cells.

Initiation and development of adaptive immune responses

As shown in *figure* 1 and explained above, dendritic cells (DCs) situate themselves at the convergence point between the innate and adaptive immunity. Indeed, they display the antigens that they have captured to the naive T lymphocytes in the T cell rich areas of regional lymph nodes. On one hand, myeloid DCs capture those antigens in the organs where they reside (above all in the epithelia), and on the other hand, plasmacytoid DCs may sample antigens in the bloodstream. By getting in contact with microbial products and other cytokines such as TNF, DCs are activated and mature into potent antigen-presenting cells (APCs). This maturation goes together with a modification of their surface molecules. They start to express high levels of CCR7, a receptor specific for the two chemokines CCL19 and CCL21 that are produced in lymphatic vessels as well as in the T cell rich areas of lymph nodes. Therefore, those DCs displaying microbial products travel through the lymphatic system to the regional lymph nodes where they will present their captured antigens to the naive T lymphocytes (which also express CCR7

and therefore also circulate to the same zone where DCs are attracted to in the lymph node).

During their maturation, DCs increase their expressions of MHC molecules and costimulators of T cell activation (B7-1, B7-2, ICAM-1, IL-12). Hence, when they will be put in contact with naive T lymphocytes they will be able to activate them and trigger the adaptive immunity. It is also important to mention that the meeting between DCs and naive T lymphocytes also take place in the absence of infection or inflammation. In fact, DCs continually sample their surroundings in the different tissues and bring those harmless antigens to the lymph nodes. The difference is that by expressing self-antigens they will not express co-activating molecules of T lymphocytes. This will nonetheless cause either the death of self-reacting naive T lymphocytes, their inactivation or their differentiation into regulatory T cells. This is a significant process that stop autoimmunity by maintaining self-tolerance.

There is likewise another way of initiating the adaptive immune response. Indeed, some antigens travel directly through the lymph vessels and arrive in their soluble form into the regional lymph nodes. There, those cell-free antigens travel through conduits of fibroblast reticular cell (FRC) that are located between the sinuses and traverse the lymph node cortex. Then, they are captured by the interdigitated processes of resident DCs, taken up by macrophages or by B cells. DCs will display their antigens to T cells, while macrophages will present them to resident B cells in the follicles.

We commonly divide the adaptive immunity in two types: the humoral and the cell-mediated immunity. By cell-mediated immunity it is meant the type of adaptive immunity that is mediated by T lymphocytes. Its three main objectives are to stimulate phagocytes in order to enhance the destruction of microorganisms (by CD4+ helper T lymphocytes), to kill infected or malignant cells (by CD8+ cytotoxic T lymphocytes) and thirdly to prevent an uncontrolled immune response by inhibiting the response (by CD4+ regulatory T cells, Treg). Humoral-mediated immunity involves B lymphocytes, which is complementary to the cell-mediated immunity as it can target other types of microorganisms, as for example extracellular pathogens or toxins. Humoral means that it is mediated by antibodies that are found in the blood and other mucosal secretions. B lymphocytes differentiate into plasmocytes, which are the cells capable of secreting different classes of antibodies.

Antibodies

Antibodies, or immunoglobulins, are basically the same molecule as the already discussed BCR present at the B cell surface. The only difference is that they have lost their transmembrane exon by alternative splicing in order to become soluble and be able to circulate freely in secretions or blood circulation. Their main mechanisms of action are the neutralization or opsonization of microbial products, NK cells sensitization, mastocytes and eosinophils sensitization, and complement activation.

Synthesis of innate and adaptive immunity

The fundamental differences between innate and adaptive immunity are the ability of adaptive effectors to specifically recognize pathogens, increase the specificity and affinity to those antigens while creating an immune memory. Furthermore, what is particularly

interesting to observe is the great interdependence between innate and adaptive immunity. Even if fundamentally different mechanisms are present to detect antigens and coordinate the immune response, each of those two immune arms needs the other one in order to fulfill a suitable response against danger. Indeed, when virulent pathogens cannot be contained and mastered only by innate immunity, adaptive immunity is there as a backup. It is a real synergy between those two subsystems, regulating human body homeostasis.

Having an effective immune response also means that regulatory mechanisms have to exist to avoid a loss of control of those inflammatory reactions and to regain immune homeostasis once the danger has been eradicated. Different ways of preventing an overresponse exist at different levels. Regulatory cells, inhibitory receptors and molecules as well as apoptotic destiny or destruction are the different ways to achieve it. It should be considered as important as powerful tools to eradicate microbes in order to avoid autodestruction of host tissues.

Evolution has provided crucial tools to fight against microbes from our environment, foreign substances, dysregulated cells that turn into tumours. However, as it will be discussed later, this strong arms can undergo dysregulations and when they are not optimally controlled, diseases occur like immunodeficiencies, autoimmunity or allergy.

1.1.4 Immune Cells Molecular Characterization

Immune cells populations are defined using cell-surface markers, especially cluster of differentiation (CD) molecules. Each defined surface molecule has its corresponding CD number. Those surface molecules can have diverse biological functions, acting for instance as receptors (e.g. CD21 is a receptor for the C3d complement molecule), coreceptors (e.g. CD3 acts as a coreceptor for the TCR), or antigen presenting peptides (e.g. CD1c presents non-peptid antigens to T lymphocytes). The combination of different discriminative markers thus define cells from a specific population and allows clustering cells according to their phenotypic resemblance. The different functions, cells expression and molecular structure of CD molecules involved in this work are synthesized in *figure*8.

Once an immune cells phenotypic recognition has been made, it is possible to look at the functional state of each cell. Activation markers (e.g. HLA-DR, CD69), cytokines expression (e.g. IL-1, TNFα), phosphorylation states of intracellular signalling pathways (e.g. STAT transcription factor) or cell division markers (e.g. Ki67) are markers of cell activation. It is thus possible to identify differences in the activation state of a patient's cells in comparison to healthy donors. It is also possible to evaluate the response of a person to vaccine, or the impact of other exogenous factors (e.g. infections, treatments) on the different cells populations.

Figure 9 provides a non-exhaustive overview of the expression and non-expression of molecules at the surface of the major known immune cells subsets. Depending on the combination of those proteins, immune cells families are isolated and characterized.

CD45 marker allows to differentiate the granulocytes neutrophils and eosinophils (CD45-CD66b+) from all other leucocytes (CD45+CD66b-). CD3 is used to differentiate T lymphocytes from other leucocytes. Then, to further divide T cells, CD4, CD8,

CD27, TCR$\gamma\delta$,CCR7, CD45RA and CD45RO are the most appropriated markers to isolate CD4+ T lymphocytes, CD8+ T lymphocytes, TCR$\gamma\delta$ T cells and NKT cells subsets (2). B cells are characterized by CD19, CD45 and CD45RA markers expression. Their subsets can be further down-clustered with CD21 and CD27. NK cells are characterized by the absence of CD3 marker and expression of CD7 and CD56 (3). CD16 will be utilized to subcluster this family, identifying NK cells sub-populations that have the ability to respond to antibody-coated cells (4). Dendritic cells are characterized by the absence of CD3, CD7, CD14, CD19, CD20, CD56 and CD66b markers, while expressing HLA-DR surface molecule. To identify myeloid dendritic cells within DCs CD11c is used, further identifying conventional mDC with CD1c and inflammatory mDC with CD16. Plasmacytoid dendritic cells are defined within DCs with CD123 (5). Monocytes are CD11c+, CD33+ and HLA-DR+. CD14 and CD16 are used to differentiate between classical, inflammatory and non-classical monocytes (6). Finally, basophils express CD123 and are distinguished from pDCs by their lack of HLA-DR.

Once the major cell populations have been identified, it is possible to go even deeper and be more precise in the definition of subsets in each population. As this work will mainly focus on dendritic cells phenotypes, activation state and response to exogenous stimuli, *figure* 10 illustrates for the three major DCs subpopulations their main surface molecules, PRRs, expressed genes and the downstream cytokines released.

1.2 Immunodeficiencies

Immunodeficiencies are a large group of diseases characterized by heterogeneous clinical manifestations. An increased susceptibility of life-threatening infections can be observed in patients presenting an immune deficiency, as well as autoimmune and chronic inflammation disorders, allergy and malignant diseases (7), (8), (9). This great variety of manifestations reflects the wide spectrum of activity of the immune system in maintaining biological homeostasis.

Immune deficiencies are divided into primary and secondary entities. About 300 primary immunodeficiencies (PID) caused by monogenic defects are known, whereas secondary immunodeficiencies (SID) can be caused by exogenous factors (e.g. infections, drugs), or other primary disease conditions, such as the nephrotic syndrome (7).

Red flags during clinical evaluation of a patient should encourage to search for an immunodeficiency. It is the case if the patient presents four or more infectious episodes each year that need an antibiotic therapy (e.g. otitis, bronchitis, sinusitis), if recurrent infections or infections that need a prolonged antibiotic therapy are noted, when two or more severe bacterial infections occur (osteomyelitis, meningitis, septicemia, cellulitis), when two or more pneumonia radiologically are proven in a period of three years, when the familial medical history is positive for a primary immunodeficiency, and finally for any infections with unusual localisation, or with an unusual pathogen.

1.2.1 Primary Immunodeficiencies

In the last 20 years, accelerating progress has been made in identifying new forms of PIDs thanks to the advances of molecular and genetic characterizations (10). These disorders are either diagnosed early in life or even later, in adults. It is estimated that

1-2% of the population might be affected with any type of the whole PID spectrum (11).

PID classification

The general classification of PID first identifies the type of immunity that is affected (e.g. innate or adaptive immunity) or if it is a problem within the downregulators of immune responses. Then, the different primary immune deficiencies diseases are grouped according to their mechanism. *Figure* 11 highlights the main types of PID, giving some disease examples belonging to each class. This classification orients the clinical approach when an immunodeficiency is suspected, based on immunological knowledge. For instance, opportunistic infections (such as recurrent infections by *Candida*), as well as recurrent viral infections, indicate that a problem exists in the adaptive cell-mediated immunity, involving T lymphocytes, those cells being the normal protectors against such infections. However, recurrent infections of the respiratory tract would rather suggest a defect in the humoral immunity, involving B lymphocytes. Finally, invasive bacterial infections indicates a problem in the innate immunity, opening as differential diagnosis, for instance, an asplenia, defects in Toll-like receptor signalling or complement deficiencies (12).

Other classifications exist, amongst all the IUIS Phenotypic Classification for Primary Immunodeficiencies, grouping PID by pathogenesis and giving more precise clinical features (13). This classification can be easily used at the bedside, correlating clinical findings to differential diagnosis.

Clinical manifestations

As mentioned above, clinical manifestations have a high diversity and PIDs can mimic various diseases, thus delaying the diagnosis. First, PIDs can increase the susceptibility to infections and allow opportunistic infections. Increased sinopulmonary infections, gastrointestinal infections, septic arthritis, bacterial meningitis and sepsis can be at the front line and should make suspect an underlying immunodeficiency.

Second, autoimmunity can be another manifestation of a PID. Autoimmune haemolytic anemia (AIHA), immune thrombocytopenia (ITP), rheumatoid arthritis, vitiligo, vasculitis, systemic lupus erythematous or inflammatory bowel disease can be consequences of an immune defect and can affect up to one out of four patients (14). The pathogenesis linking autoimmunity with immunodeficiency is not completely understood, but it could outline a defect in the regulatory mechanisms of the immune response. Third, allergic reactions can also be a manifestation. It can present itself as a difficulty to control asthma for example, food or other materials allergies, as recurrent or complicated sinusitis/otitis, or even as eczema. Finally, malignant disorders are also linked to PIDs. There is a higher susceptibility to develop non-Hodgkin lymphomas, as well as other gastric cancers.

Diagnosis

Different paraclinical exams are available to diagnose an immune deficiency. Simple routine checks can be perform to look for indirect signs of PID. Blood cell counts (neutrophil, lymphocyte and eosinophil counts) can reveal a cells population deficiency. Chest x-ray can be used to evaluate the thymic shadow as well as costochondral junctions; bone x-ray allows the evaluation of metaphyseal ends and can show abnormalities

in the context of a syndromic immunodeficiency (e.g. cartilage hair hypoplasia).

Then, as soon as an immune deficiency disease is suspected, immunoglobulins serum levels should be analysed as it could highlight an hyper- or hypo-globulinemia. When a suspicion of chronic granulomatous disease is made, dihydrorhodamine fluorescence assay and nitroblue tetrazolium assay can be made to look at the reactive oxygen species produced by neutrophils. The complement system can be functionally evaluated by dosing the activity of different proteins involved in the classic and alternative complement pathways. Those proteins are CH50, AP50 or MBL50. Immunonephelometry can dose quantitative antigens fractions C3 and C4. Levels of regulatory proteins of the complement pathways can also be determined, as well as the fragments of complement systemic activation (sC5b-9 and Bb). Then, asplenia could be sought through abdominal ultrasonography. Finally, phenotyping the different lymphocytes subfamilies with cytometry techniques can also be performed and bring to light abnormal absolute or relative cell counts (12).

Making a precise diagnosis of immunodeficiency can result to be very difficult in cases of non-straightforward immune cells alterations. Complementary examinations are essential to define the type of the immune deficiency, as the clinical features are often not specific and only makes us suspect such a disorder. For instance, identifying a quantitative cell defect in a severe combined immunodeficiency (SCID) can be easily done with flow cytometry, clearly outlying the differentiation defect leading to the lack of a specific population. However, in other cases, a simple immune cell analysis does not allow a precise characterization of the deficiency. When genes or molecules regulating immune cell functions are affected, simple immunophenotyping studies can return normal results and deeper analyses are needed to identify the immune problem(s). Those analysis using mass cytometry and bioinformatics techniques are the main point of this work and will be described later. It is supposed to unveil new immune deficiencies that could not be revealed until now, looking at the unclear immune signatures that cannot be classified actually.

Management approaches

The management of patients suffering for a primary immunodeficiency mainly depends on the type of immunodeficiency present. General concepts are presented here (sources: (15)). It goes hand in hand with a prompt vigilance of all organs to detect associated autoimmune, malignant or other disorders.

Firstly, a special attention in the prevention and treatment of infections should be made in order to avoid complications and sequelas. Viral infections (e.g. EBV, VZV and adenovirus) should be screened periodically and regular vaccination should be performed. A proper dental care should be insured, in order to avoid buccal infections and gateway to systemic breach. Then, immunocompromised patients respond not as well as immunocompetent persons to antibiotherapy. It is hence important to adjust the dosis and duration of treatment in this group of patients in order to guarantee a safe recovery and avoid re-infection. Usual antibiotics dosis can be doubled or tripled, depending on the situation.

Antibioprophylaxis can also be considered, as for instance azithromycin (16). Furthermore, depending on the immune deficiency, other antibiotics can be prescribed to

avoid opportunistic microorganisms infections like *Pneumocystis jirivecii*. In this case, cotrimoxazol can be given three times a week. In the case of an increased susceptibility to mycobacterias, like in the MSMD, prophylaxis against *Mycobacterium* can also be given.

Intravenous immune globulin therapy can be considered amongst all in B cell immunodeficiencies, combined immunodeficiencies or in other immune dysregulations. Specialized immune globulins injections for CMV, VZV or RSV can also be evaluated to strengthen the patients' defence against those pathogens.

Finally, in the most extreme cases, an immune reconstitution can be considered. It is usually the last step in the treatment approach, or reserved to the most severe PID presentations. It includes an hematopoietic cell transplantation (HCT) (17), gene therapy (18) or thymic transplant in a DiGeorge syndrome (19) or forkhead box N1 (FOXN1) deficiency (20).

1.2.2 Secondary Immunodeficiencies

Secondary immunodeficiencies (SIDs) are far more common than PIDs. There are several external conditions that can lead to decreased immune competences by affecting any arm of immunity. Malnutrition, drugs, metabolic diseases, environmental conditions or infectious diseases can induce an immunodeficiency. Physiological conditions as extreme ages (newborn and advanced age) can also be accompanied by a decrease in the host immune defences (21).

The general principle for the management of SIDs is to cure or eradicate the underlying condition that causes immune dysfunction. When this elimination is not possible, then prophylaxis against infections and other measures (e.g. vaccination, antibioprophylaxis) as discussed for PIDs can be undertaken to limit the risk of severe infections that can lead to permanent functional losses. As this work will not focus on those conditions, no further details will be given.

1.3 Immunophenotyping

Immune cell characterization has extensively showed its importance in the clinical management of patients, as well as in the understanding of not only immunological diseases, with their classification and treatment, but also in other specialties like oncology or infectiology (22), (23), (24), (25), (26). The key technique to be able to analyse the phenotype and activation state of immune cells is using immunophenotyping methods. This chapter will outline the general principles of immunophenotyping, from the laboratory techniques used to bioinformatics methods available to handle the huge amount of data gathered with cytometric methods.

1.3.1 Cytometric methods

Cytometry is the art of measuring and acquiring cells properties, that is to bring to light the different molecules expressed and displayed by cells, as well as their cytoplasmic content. To reveal the content and the appearance of a cell, different approaches can be used. Nowadays single-cell mass cytometry techniques have shown indispensable new contributions in the diagnosis of immunocompromised patients allowing a deep insight in the proteomics of any cells. The great increase in measurable parameters that can

be seen at once in a cell provides a fast and large immune system screening in order to get rapidly a picture of the phenotypic aspect of the different patient's leukocytes. Furthermore, the implementation of this recent technique has also brought new prospects in the understanding of various functional states that might be impaired in immunocompromised hosts.

Flow cytometry

The well-established flow cytometry technology (FACS) using fluorescent staining of cellular proteins allows for a reliable quantitative and qualitative characterization of different immune cell population in order to study their relative and absolute abundance (27), (28), (29). FACS techniques are well positioned in assisting the clinician in his diagnosis of immunological pathologies, hence identifying abnormal levels of different immune cells populations in the blood (30), (31).

Any type of tissues cells can be analysed with this technique, giving a quick overview of the cells content and phenotypes in a wide variety of organs: bone marrow, cerebrospinal fluid, blood, urine and any other solid organ. In fact, this method is already routinely used in many different clinical specialities, like immunology, oncology or haemoatology. Its strength resides also in its standardized use in diagnostics, high rate of cells analysis and capability of sorting cells. However, when the need comes of inspecting more parameters in a single cell, it becomes challenging to evaluate over 12-15 markers due to the compensations needed to correct for the overlap in the fluorescence spectra of the different fluorophores used and autofluorescence. This is particularly problematic when an in deep immune cells profiling is sought.

Mass cytometry

Single-cell mass cytometry (CyTOF) techniques have been applied recently to enhance phenotypic and functional understanding of immune impairments (32), (33). It is a technique that locates itself between flow cytometry and mass spectrometry. More than 40 cell markers can be profiled in parallel with this innovative technology thanks to the detection of pure rare earth metal isotopes, amongst all from the lanthanides group (e.g. europium, neodymium, gadolinium, holmium), conjugated to antibodies that are detected by an inductively coupled plasma (ICP) mass spectrometer.

The way mass cytometry works is as follows. Metal-chelating polymers are loaded with pure rare earth metal isotopes and then covalently attached to the antibodies. Those metals are normally not found in normal biological systems. Therefore, each isotope will correspond to a specific extra- or intracellular cell marker that will then be possible to detect. Cells are washed to remove unbound antibodies, then the cell suspension is introduced into the mass cytometer that atomizes and ionizes the cells. A radio frequency quadrupole ensures the filtration to remove all low mass elements that are normally present in biological samples (carbon, nitrogen, and other elements with an atomic mass smaller than 80 Da). Finally, each isotope reaches a time-of-flight (TOF) detector, which allows a quantitative isotopes signature for each cell of our sample (32). Data are finally stored into a flow cytometry standard (fcs) file and can be analyzed by manual gating and by unsupervised analysis software packages. *Figure* 12 summarizes the different steps required for data acquisition.

Recently, mass cytometry techniques have been applied for diagnostic purposes, providing a significant increase in the breadth and depth of the proportional and phenotypic characterization of a patient's immune cells, in comparison to traditional flow cytometry techniques. These advancements are driven by the great increase in measurable parameters provided by mass cytometry which allows for all major immune cell populations and subpopulations to be characterized with a single analysis. It brings undoubtedly the prospect of improved diagnosis and therapies for immunodeficiencies and autoimmune diseases.The main drawbacks of mass cytometry are its impossibility to sort and collect cells as in flow cytometry due to the fact that each cell is destroyed when passing through the instrument. Furthermore, the speed of data acquisition is also lower than using FACS techniques (34).

A single-cell mass cytometry panel consisting of 35 different metal isotope conjugated antibodies is currently being used by CHUV diagnostics department to provide an in depth immunological profile patients' immune cells. Suspected immunodeficiency, non-response to a vaccine or autoimmune diseases are the primary criteria for the use of this new diagnostic tool. Prior to the accreditation of this assay in Switzerland, an extensive cross-validation study between flow and mass cytometry was performed by the Service of Immunology and Allergy of the CHUV, showing a strong statistical correlation between these two analytical methods, which also appeared to be the case in the literature (35). This reinforces the validity of using the increased analysis power of mass cytometry.

Thanks to this technique, it is thus possible to compare various samples in different stimulation conditions. The process from blood samples to computable substantial data requires meticulous attention and management of data measurement variations. Preparation of blood samples should be standardized between the different experimental conditions in order to avoid unwanted bias. In a similar manner to flow cytometry, acquisition through mass cytometer can also induce a certain variability due to instrument performance variations over time that can induce signal intensity fluctuations. To face this challenge, barcoding of samples can be used, which also reduces the number of cytometer runs and spares conjugated antibody reagents used for staining cells (36).

1.3.2 Data pre-processing

The data measured by mass cytometry are transformed into numerical values in order to be stored in a standardized file format called flow cytometry data file standard (FCS). Each event (meaning each cell analysed) has its parameters stored one after the other in a list called data set. It can be conceptualised as a large matrix containing for each cell all its parameters measured. The first step of data pre-processing is therefore a matter of importing the database in the programmed desired for further use. Different softwares can handle FCS files. The ones that are going to be used in this work are Cytobank, FlowJo, R, and Matlab. The two first ones are the most adequate to perform manual gating and isolate cells subsets manually.

As it has been mentioned above, an important challenge that has to be face is instruments' performance variations from one run to another, and as well after calibration or cleaning. In general, the instrument's sensitivity decreases in relation with the acquisition time. A proposed method that showed encouraging results is the so-called normalization

method with bead standards. This consists of providing quality assurance by correcting short-term and long-term signal fluctuations to get cleaner, better interpretable biological differences. To face this challenge, polystyrene beads standards linked with a combination of heavy metal isotopes are added to each cell suspension ready to be passed through the cytometer. Then, when comparing samples, the beads are spotted and their mean intensity is calculated and will serve as the reference to re-adjust and normalize the measurements obtained from the stained antibodies. This makes it possible to consistently compare different data acquired over a large period of time, taking into account the cytometer's variations in its measurements, as well to quantify the instrument's performance (37).

Before starting unsupervised analysis, normalized mass cytometry acquired data also have to be transformed. Indeed, it is known that signals measurements of cellular molecules by flow cytometry follows a normal logarithmic distribution. When comparison of large distances are wished, it is important to scale transformed the data. Hence, performing a linear transform to a logarithmic scale is essential in flow cytometry data processing to get an entire overview of data distribution according to each of their parameters intensity. When it is desirable to handle zero and negative values, hyperbolic inverse sine (*arcsinh*) transform is used, as logarithms of negative values are not allowed. Moreover, it allows for the data distribution to be more symmetric, reduce calculation bias for large numerical distances that are not experimentally significant, thus increasing the interpretation of data around zero value allowing comparisons (38), (39).

Then, the next step in data pre-processing consists in cleaning the data set. Events corresponding to single cells are conserved through a gating strategy to remove events containing abnormal DNA content including cellular debris (e.g. dead cells) with low levels of DNA or cell doublets with elevated DNA intensity. After accomplishing this exercise, barcoded samples are separated into individual sample datasets using distinct isotopic markers chosen to define each sample.

Once these pre-processing steps have been achieved to isolate cells from patients and healthy donors into single FCS files, primary analysis is based on standard gating practices manually separating major known immune cell populations. This method relies on our current phenotypic understanding of immune cells. Once this first step has been achieved, unsupervised analysis of individual populations allows for an unbiased evaluation of potential differences between different sets of samples (i.e. healthy donors and patients). Secondary analysis procedure is driven on all cells, directly analyzing them with semi- or unsupervised clustering methods. The large variety of surface proteins on human leucocytes provides many different possibilities to isolate cell populations. Indeed, no fully standardized immunophenotyping exists, even if suggestions are made to find a consensus, especially in manual gating (2). Once the desired populations in each sample has been semi-automatically isolated, the downstream analysis of this second approach can begin.

1.3.3 Bioinformatics methods using R programming

In order to be able to interpret the relevant information of this large amount of data produced by the CyTOF analysis, some essential tools in bioinformatics are required (40),

(41). This new discipline was born with the need of analyzing large data produced by genomics (42). Nowadays, it finds its application in many other biological fields, situating itself at the converging point of biology, computer science, mathematics and statistics. Therefore, this will be the key point of the analyses.

Indeed, powerful methods exist in R programming, amongst all using *Bioconductor* packages, to highlight cell subsets characteristics in high dimensional mass cytometry data (43), (44), (45). It provides different tools to analyze and help interpret multidimensional large datasets. Supervised, semi-supervised and unsupervised cells clustering methods are suitable and available for this data processing (38). For instance, FlowSOM package allows its user to visualize the data, as well as building a minimal spanning tree and a meta-clustering (46). Once the metaclustering is performed, different possibilities of analysis pipelines are conveniently available: manually down-clustering, visual representation with dimensionality reduction (47) and intra-clusters properties evaluation with a differential analysis (38). Clustering algorithms allow for the separation of specific cell populations based on their phenotypic and/or functional profiles, as shown in *figure* 13.

Different algorithms and functions are used in this work. The main ones are now going to be detailed.

Heatmaps

Heatmap is a powerful way to visualize median marker intensities to compare between different variables and samples. For instance, it gives a quick graphical overview of marker expression in a cell population compared to others, representing this intensity in a pixels matrix as colors. It facilitates thus the visualization and comparison of values from multiple samples.

Self-organizing maps and minimal spanning trees

Events in high-dimensional data sets are challenging to cluster without neglecting the multidimensionality, as it is often done by manual two-dimensional gating. Clustering is the art of organizing and grouping events according to their protein expression profiles similarities. To achieve this task, different algorithms exist using different mathematical approaches, such as k-means clustering. The choice of the method influence the results and can bring bias or create irrelevant subsequent biological interpretations when inadequate (48).

FlowSOM clustering approach, an opensource R package, has shown superior performances, allowing a good clusters separation comparable to manual gating, and fast runtimes in comparison to different other clustering methods. Therefore, it makes it suitable for large multi-dimensional datasets analysis (40). It relies on the building of a self-organizing map, which is a type of artificial neural network, using competitive learning. Using Euclidean distance, a vector is compared to already existing nodes (initially randomly assigned with points from the dataset), and the best matching unit (e.g. the closest node) is selected. It creates a two-dimensional interdependent nodes representation of the multi-dimensional data, allowing a clustering according to the high-dimensional similarities of each event (49). Then, a minimal-spanning tree is built, which allows a visual representation of the obtained clusters. It links the closest clusters together, in a way that the sum of the different connections from a node is minimal (46).

t-distributed stochastic neighbor embedding (tSNE) algorithm

tSNE algorithm creates a low dimensional map from a high dimensional space. Similarity matrix is built from distance matrix, based on Euclidean distance and local density. Then, it maps the different points of the dataset in lower dimensions using student's t-distribution. By dimension reduction, the Kullback-Leibler divergence (or relative entropy, which is the measure of dissimilarity between two probability distributions) is minimized, in order to get a final two dimensional map displaying each point next to its nearest neighbours, according to higher dimensional proximity. It is thus a nonlinear dimension reduction algorithm, able to capture non-linear relationships between points of the dataset. This makes it different from conventional principal component analysis based on covariance matrices, which assumes linear relationships between parameters.

Then, to get the visual representation of this dimensional reducted high dimensional single-cell data, a scatter plot displaying all events (or cells) is represented using viSNE. Therefore, as discussed above, the distance in the two dimensional space between to points reflects the Euclidean distance in the high-dimensional space. An additional dimension can be represented by colouring each cell according to their clusters that they have been assigned to with FlowSOM for instance, or according to the expression level of different markers intensities (22). The main drawback of viSNE is the so-called "crowding problem" (22). Not all events can be mapped as it would saturate the visual represented field.

Differential analysis

Once different clusters have been defined and isolated, a differential analysis to evaluate inter-clusters variability can be undertaken. Classical statistical methods are not appropriated and optimal to handle non-normal distributed data, and usually do not incorporate random effects in their designs. However, data acquired by single-cell mass cytometry can present those two characteristics.

First, variations among individuals do exist, which add a random effect in the measurements. To handle this, linear mixed models integrate random effects and are appropriated. In this way, this model can be used when comparing markers expression between different cell populations. Second, the acquired data do not necessarily follow a normal distribution, as they can contain binary values or counts. Furthermore, when handling proportions with lower cell counts, the uncertainty of the calculations is higher. Generalized mixed models are suitable to handle the non-normal distribution of data, as for instance using a logistic regression as binomial distribution. Those models are also able to catch a randomly distribution due to inter-individuals evaluation. This can therefore be used to compare cell populations proportions (38), (50).

2 Methodology

2.1 Patients and controls

Patients with suspected immunodeficiencies (n = 15) have been selected. They presented higher susceptibility to infections and at the request of their personal physician, their blood was sent to the diagnostics laboratory of immunology for a quantitative analysis of their immune cells. More clinical information or demographical information are

not known. Control subjects (n = 15) were healthy donors from the laboratory or volunteers from the CHUV hospital occupation consultation. There were 11 women and 5 men who were of matched age with the selected patients.

2.2 Data acquisition

2.2.1 Blood preparation

Blood samples of the above mentioned patients (P) and healthy donors (HD) have been processed and analysed. The first step was to prepare, stimulate, fix and freeze them. In total, 2x 200 μL are used for each patient and healthy donor, separated into tube A and tube B. 5 healthy donor samples and 5 patient samples are processed together, as they are batched and barcoded, as explained below.

Blood stimulation and incubation

Whole blood will be stimulated with LPS, at the concentration of 200 ng/mL, and R848, at the concentration of 1 μg/mL, in the tube B. On the other hand, tube A will consist of unstimulated blood. Both tubes received golgi plug at a concentration of 1 uL/mL and were left in incubation for 4 hours at 37°C, without agitation.

Fixation and freeze

Then, in order to freeze and keep the samples stables during all blood samples acquisition, both tubes for each patient and HD were treated with SMART tube stabilizer (280 μL per well) and let incubate for 10 minutes at room temperature. Then, they were directly put on ice and freeze at -80°C for at least 1 hour.

2.2.2 Blood thawing, staining and analysis through mass cytometer

Thawing and red blood cells lyse

Frozen blood was thawed in water at 10°C and then red blood cells were lysed with 4 mL thaw-lyse buffer diluted 1'000 times in DDWater. A 10 minute incubation at room temperature was followed by a 600xg centrifuge for 7 minutes at room temperature. Finally, supernatant was discard and cells were washed with 1 mL of phosphate-buffered saline (PBS) containing 0.02% saponin.

CD45 and palladium barcoding

In order to use multiplexing during mass cytometry acquisition, barcoding of the samples was used using CD45 and isothiocyanobenzyl-EDTA (palladium) to chemically conjugate the fixed cells. CD45 antibodies linked to different heavy metal isotopes were used to code and batch 5 different samples together, whereas two isotopes of palladium are used to code the unstimulated and stimulated conditions. Batching and barcoding samples together are important to reduce variability between the different cytometer acquisitions. Furthermore, it allows to save on the amount of antibodies used for the downstream staining. This is why this approach has been chosen in this work.

The heavy metal isotopes used with CD45 marker were yttrium-89, praseodymium-141, platinum-194, platinum-195, and platinum-198. The staining in 50 μL PBS-saponin solution lasted 30 minutes at room temperature in agitation.

Then, after washing the cells, palladium barcoding with palladium-104 (Pd-104) and palladium-108 (Pd-108) isotopes was performed by suspending the cells in a 80 μL dimethyl sulfoxide (DMSO) solution containing the corresponding Pd isotope at 400 nM for 30 minutes at 4°C. Pd-104 corresponds to the unstimulated (US) condition, and Pd-108 corresponds to the stimulated (S) condition. After those first 30 minutes of incubation, the solutions were quenched using 400 μL CSM and incubated for 10 more minutes at 4°C.

Finally, the samples were pooled together in a patients tube and another healthy donors tube. Those two tubes will be processed in parallel and will not be mixed together. After a washing with 1 mL CSM, they are re-suspended in 1 mL CSM, ready for the next processing step.

90% granulocytes withdrawal

As the analysis will not focus specifically on the granulocytes, which constitute the major part of peripheral blood leukocytes, 90% of them were removed to save antibodies and decrease the later acquiring time through mass cytometry. Therefore, 100 μL of sample were taken from the suspension and kept aside. A column for magnetic cells separation, using magnetic anti-CD15 antibody beady, was used to specifically deplete the granulocytes from the sample. Thus, the cells were suspended in a mix containing 2.5 μL beads and 97.5 μL CSM, and were incubated during 15 minutes at 4°C. Then, they were washed and re-suspended in 500 μL CSM and they were passed through the magnetic column. Once all the suspension had passed through, the column was washed with 1 mL CSM, and the new suspension put together with the formerly kept aside 100 μL suspension, which let at the end only 10% of the total granulocytes.

Extracellular staining

The cells were re-suspended in the antibody staining mix diluted in CSM, for a final volume of 50 μL. The mix of antibodies used with their corresponding bound heavy metal isotopes are shown in *figure* 14, for a total incubation time of 30 minutes at room temperature in agitation. Then, there were washed and fixed with 2.5% PFA and let incubate for 5 minutes at room temperature in order to permeabilize the cells and allow a further intracellular staining. 1 mL of CSM-S were used to wash the cells and maintain the permeabilization that had been done.

Intracellular staining

The next step was the intracellular staining, using the antibody mix shown in *figure* 14. The intracellular antibody mix diluted in CSM-S for a final volume of 50 μL was used to re-suspend the cells. A 30 minutes incubation at 4 degrees took place before the next wash.

Iridium staining and data acquisition through mass cytometer

Finally, an iridium staining was performed to mark cellular DNA. Hence, the cells were mixed with a solution containing 450 μL CSM-S, 1 μL iridium and 1.6% PFA. The final sample was then passed through the mass cytometer for data acquisition.

2.3 Data processing

2.3.1 Data pre-processing

Preprocessing of FCS files

Once the different FCS files have been acquired through the mass cytometer, a bead normalization is undertaken in order to get rid of signal variability, using the bead normalization package from nolanlab written in Matlab. During CyTOF acquisition, different calibration beads have been used, that are Cerium-140, Europium-151 and -153, and Lutetium-175. Manual gating in each beads domain is undertaken, and the events situating themselves at the intersection of the different gates will be interpreted as the beads by the program, and their intensity will be used as normalization reference. The Matlab code creates new FCS files with the initial data without the events corresponding to the gated beads.

Cleaning the data set and separating the barcoded samples

In order to separate the acquired FCS files, containing batched samples, into different file representing each sample for each condition individually, FlowJo has been used for the gating and the exportation of the new single FCS files. During this pre-processing step, it has also been used to clean the dataset and remove the cell debris, doublets or other noisy events that will not be interpreted as a single cell. *Figure* 15 shows the gating strategy (illustrated by healthy donors 1 to 5) to achieve the individualization of each condition for each healthy donor or patient. Moreover, the gating into stimulated and unstimulated cells only shows the use of 104 as a barcoding marker, and not 108 for instance. This underscores the importance of using both 104 and 108 barcoding markers for unstimulated and stimulated conditions since there is not a clean separation between some of the populations including healthy donor 5.

Manual gating strategy

Their are two further approaches in the bioinformatical analysis pipeline. The first one consists of utilizing all cells data and drive semi-unsupervised clustering approaches, as explained below. In this case, the FCS files created in the last step are sufficient and no further pre-processing is required. However, the second approach consists of gating manually the major known immune cells populations, and drive an unsupervised clustering algorithm for each of the isolated populations, to identify sub-families. There are some opinions that fully unsupervised clustering methods might not be optimal and have to be used with extreme caution as statistical relationships predominate above biological relevant data. Thus, to explorate the different approaches and compare their accordance, manual gating has been performed in this work. Then, it has focused on the myeloid dendritic cells subpopulation to drive further unsupervised clustering algorithms and differential analysis to compare between the different conditions and the different subpopulations.

Thus, *figure* 16 shows the manual gating strategy used to isolate the major known immune cells populations, using Cytobank software. It should be mentioned that the presented gating strategy for the separation of NK cells was not optimal and due to a reduced staining of CD7 in some patient samples (NK cells being CD3-/CD7+/CD56+). Generally, the NK cells can be isolated directly from the CD3- cell population by their expression of CD7 and elevated expression levels of CD56. Here, the gating strategy

in *figure* 16 for NK cells showed sequentially the removal of: CD3+ T cells, CD19+ B cells, CD14+ monocytes and HLA-DR+ dendritic cells, and removal of CD123+ basophils before gating on the NKG2+ NK cells.

2.3.2 Bioinformatical analysis pipeline

As said earlier, the analysis has been driven on two populations of cells. The first one was on all of the cells, whereas the second one was on the manually gated myeloid dendritic cells subset. The bioinformatical analysis was done using R programming and its different packages available for high dimensional biological data analysis. Many are provided by Bioconductor, an open source platform for bioinformatics. The general pipeline is the same for those two subsets analysis. Different comparisons will then be possible with the manually or unsupervised clustering methods, comparing each population (e.g. healthy donors and patients) between each other, or comparing between each of the condition (e.g. unstimulated or stimulated) for each of the cell populations.

Another important notice is that the data have been transformed before driving any clustering or other analytical method. Hyperbolic inverse sine (arcsinh) has been applied with a cofactor 5. This has been made in order to have a more symmetric distribution between the different markers intensities measured and to make them comparable to each other. This is specially important when clustering or other differential analysis have to be made. This analysis pipeline and the R code have been greatly inspired by (38).

Immune cells populations clustering

FlowSOM and ConsensusClusterPlus libraries have been used in order to identify cells populations based on their similarities regarding some chosen surface markers. Those methods have been preferred to others as they have been considered as the most efficient in terms of speed of calculation and performance in a comparison of different clustering algorithms (40). Clustering surface markers have been set based on biological considerations, choosing the markers that can best discriminate the different cells populations. Those are represented in *figure* 17, being different when looking and clustering all the cells or only the mDCs subpopulation.

After having built the minimal spanning tree using FlowSOM, a further metaclustering step is performed using ConsensusClusterPlus algorithm. This has as an effect to group once again by similarities the 100 nodes identified by FlowSOM into only 20 groups, and assign to each cell its group affiliation.

Immune cells populations visualization and further cluster merging

The two-dimensional visual representation of the cells is obtained using tSNE algorithm, and the package in R programming called Rtsne. It is possible then to add a third dimension colouring each cell of the obtained tSNE graph according to the cluster it has been assigned with ConsensusClusterPlus. Then, a heatmap is generated in order to visualize the different molecular properties of the different populations.

Finally, in the case of automatically analysing all of the cells together, a last clustering step is done to merge some of the identified 20 clusters into reduced known immune cells populations. This is done manually, based on the heatmap obtained showing the

expression of the different surface markers of each population.

It is important to mention that all of those clustering steps are done looking at all cells together, that means mixing healthy donors and patients cells. This is done in order to get more robustness when assigning a family to a cell, as more cells are constituting the reservoir for the clustering. However, when a visualization is made, healthy donors can be separated from patients, as well as stimulated conditions can be separated from unstimulated conditions. Once all cells have been classified, it is possible to perform differential analysis to compare the populations and conditions between each other.

Differential analysis

As mentioned earlier when talking about the available bioinformatic approaches to statistically compare groups, a right statistic model has to be chosen. In this work, two major statistical models are used, and significance was determined for a p-value < 0.05, obtained with a t-test to test the null hypothesis. In both cases, it is essential to take into account the inter-individual variations that should not be considered as pathological differences. Therefore, the two models will have to consider a random effect depending on the number of samples analysed and the corresponding overdispersion.

Statistical tests will be driven on a large high dimensional data, comparing each marker expression for each sub-population, between different groups. However, some of the results emerging from this large statistical analysis will be significant by chance. That is why those false positive have to be consider. A method proposed is to use the false discovery rate (FDR) approach with corresponding adjusted p-value to multiple testing proposed by Benjamini and Hochberg in 1995 (51). A FDR of 5% is used in this work.

When comparing proportions of population sizes, a binomial distribution is not adequate to model the dispersion of the data. Therefore, a generalized linear mixed model (GLMM) was performed using logistic function as a binomial distribution. On the other hand, when comparing intensities of marker expression between cell populations, a binomial distribution is adequate. Thus, to take into account the random effect from inter-individuality, the linear mixed model (LMM) was used. Those two models are available in the package lme4 provided by the CRAN project.

3 Results

3.1 All cells analysis

The first results presented give a rough idea about the data, in order to screen for data quality. A barplot showing for each sample and for each condition the cells events allows to have a main idea on the consistency in the size of each of them. It is represented in *figure* 18.

The next figure has been constructed to identify the statistically most discriminating surface markers for all the cells, that is the markers that present the highest variability in each sample. The score is called non-redundancy score (NRS) and it is based on a principal component analysis. The markers presented on the left have a greater NRS, which means that they might be more discriminating that the ones on the right of the graph. The resulting graph is shown on *figure* 19. It should be noted that this graph does not take into account biological relevant markers, it is only based on a statistical point of view

and it should therefore not be taken as an absolute truth. However, when comparing it to the markers that had been selected for their biological significance in distinguishing the prevalent immune cell populations (see *figure* 17), it is reassuring that most of those chosen situate themselves on the left part of the graph.

3.1.1 Cells populations clustering

FlowSOM minimal spanning tree is represented on *figure* 20. Once it is obtained, as said earlier, ConsensusClusterPlus undertakes a further meta-clustering to merge those 100 clusters into 20. As a result, the heatmap shown in *figure* 21 is obtained, displaying for each cluster the median intensity of its markers expressions. Based on this heatmap, the final manual down-clustering step was performed, and the heatmap of *figure* 22 shows the expression intensities for each of the known and identified immune cells populations. Remaining cells debris identified, antibodies aggregations or abnormal clusters have been removed (as for example cluster 14 on *figure* 22.

3.1.2 Cells populations visualization

Figure 23 shows the obtained tSNE plot of the data, where each cell is coloured according to its meta-cluster that has been assigned. *Figure* 24 represents the known identified major immune cells populations. *Figure* 25 allows a quick overview of the populations stratified by samples, in order to detect a possible lack of one of the population in a sample. The tSNE plot of healthy donors (containing both unstimulated and stimulated conditions) was compared to the tSNE plot of the patients group (also containing unstimulated and stimulated conditions), which is shown in *figure* 26. Finally, once the clustering was over, a comparison of the proportions between this automatic clustering method and the manual one was performed. The *figure* 27 shows the obtained resulting graphs.

3.1.3 Differential analysis

Before starting the differential analysis, different heatmaps have been made to have an overview of the expression of different activation markers and cytokines expressions in each of the groups and conditions. *Figure* 28 represents the intensities of the activation markers screened for each of the groups and conditions, whereas *figure* 29 represents the cytokines expressions.

Then, the relative abundance of the clustered immune cells populations was compared between the healthy donors and patients in the unstimulated condition, to see if there was a significant difference of population abundance between those two groups. Boxplots with jittered points are represented in *figure* 30. The differential analysis using the generalized mixed model showed that the CD4+ T cells population was significantly lower in proportion to other cells populations in the patients group compared to the healthy donors group, with a calculated p-value of 0.0015.

The differential analysis of cytokines expression in the different immune cells populations between the patients group and the healthy donors did not find any statistically

significant results for a p-value ¡ 0.05, neither when comparing the populations in the unstimulated condition, nor in the stimulated condition. Boxplot with jittered points representing the different cytokines expression in the two populations in each of the conditions is represented in *figure* 31. However, when looking at activation markers, both in the unstimulated and stimulated conditions there were statistically significant differences in the activation markers expressions intensities in the patients group, as illustrated in *figure* 32 and 33.

3.2 Myeloid dendritic cells analysis

This part of the analysis is performed on the manually gated population of myeloid dendritic cells. As performed for the analysis of all cells, quality control graphs are suitable to have a quick overview of our data during automatic clustering and analysis. *Figure* 34 shows the number of events present in each of the samples and conditions. The NRS for all samples and chosen discriminative surface markers is displayed on *figure* 35.

3.2.1 Cells populations clustering

Figure 36 shows (A) the minimal spanning tree obtained after running FlowSOM algorithm on the manually gated mDC population, and (B) the CD1c expression intensities in each cluster which corresponds to the classical dendritic cells within the mDC population. The heatmap showing the markers intensities of each cluster after the ConsensusClusterPlus algorithm is displayed in *figure* 37. Another representation of the median marker and cytokines intensities among the 20 meta-clusters is given in *figures* 38 and 39.

3.2.2 Cells populations visualization

The tSNE graph obtained illustrating two-dimensionally the meta-clustering obtained is illustrated in *figure* 40. *Figure* 41 displays the tSNE plots of the healthy donors compared to the one of the patients, whereas *figure* 42 allows a quick overview of the populations stratified by samples, in order to detect a possible lack or increase of cells in one of the population in a sample. *Figure* 43 focuses on the separate tSNE graph of one particular patient and a particular healthy donor.

3.2.3 Differential analysis

First, the analysis has been made over the whole set of mDCs, without considering the clusters identified. For the activation markers, no significant differences in their expression have been detected in the unstimulated condition. In the stimulated condition, *figure* 44 shows the markers that have a statistical significant difference in their expression between the patients and healthy donors groups. When looking at the cytokines expression between healthy donors and patients, there was a statistically significant (p-value < 0.5) decrease of IL-1b, IL-6, IL12p40 and TNFα expression in the patients group compared to the healthy donors in the unstimulated condition. In the stimulated condition, the statistically significant results are shown in *figure* 45.

Now, performing the differential analysis among the 20 clusters of mDC population, *figure* 46 shows the median activation markers expression intensities , whereas *figure* 47 shows the median cytokines expression intensities.

Then, the relative abundance of the clustered mDCs subpopulations was compared between the healthy donors and patients in the unstimulated condition, to see if there was a significant difference of population abundance between those two groups. The differential analysis using the generalized mixed model is shown in *figure* 48, and box-plots with jittered points are shown in *figure* 49.

The differential analysis of cytokines expression in the different mDC subpopulations between the patients group and the healthy donors is shown on *figure* 50 for the unstimulated condition, and on *figure* 51 for the stimulated condition. Moreover, when looking at activation markers, both in the unstimulated and stimulated conditions there were statistically significant differences in the activation markers expressions intensities in the patients group, as illustrated in *figure* 52 and 53.

4 Discussion

4.1 All cells analysis

The first general graph obtained, *figure* 18, illustrating the number of cells for each sample and condition shows a certain variability in the events present in our data. That can have different explanations. First, there can be an intrinsic difference in the immune cell number of each patient and healthy donor, coming from inter-individual variability or, in cases of patients, from pathological immune cells counts. Second, it can also be due to different loss of cells during laboratory manipulations and thus can be a result of non-accuracy in the handling of cells. This variability should not have non-desirable repercussions as the clustering was conducted on all cells together, that is healthy donors and patients mixed, and the differential analysis compared proportions of populations, as well as markers expression in each of the population. Thus, no absolute cell counts analysis were undertaken.

Figure 19 has mostly been used to strengthen the selection of surface markers that were used to delimit cells lineage during the automated clustering. Those markers were chosen based on the literature, choosing the ones that were able to discriminate the best cells populations. NRS score confirmed that most of them were statistically great discriminators.

FlowSOM minimal spanning tree, illustrated in *figure* 20 shows already a consistent cells clustering, as it is possible to distinguish the different major known cell populations. Furthermore, it is also an interesting way to keep an eye on the automated clustering. Indeed, unsupervised clustering can turn out to be dangerous in a sense that one is not always aware of what is happening. Hence, it is primordial to get intermediate plots that allow a semi-supervision of the analysis.

After the ConsensusClusterPlus meta-clustering to 20 final clusters, the major cells populations could be identified in *figure* 21. However, the 20 clusters were not perfectly adequate to identify all subfamilies of immune cells. Thus, a meta-clustering to a final

number of clusters larger than 20 could be a possibility to improve the final manual clustering. Indeed, if for instance subpopulations of myeloid dendritic cells were wished to be found, it would not have been straightforward using the initial analysis of all immune cells combined. As such, a new FlowSOM analysis should be undertaken specially on this isolated population to focus the clustering on this subset.

Moreover, the results obtained after ConsensusClusterPlus meta-clustering in *figure* 21 show that some populations are biologically unrelated and irrational. This is the case for the minor population 14 (0.023% of all cells) that most likely corresponds to an aggregate of antibody, cell debris and DNA that was not eliminated in pre-filtering the data. In any case, it has been decided to keep this population aside and not consider them in the downstream analysis.

tSNE plots are great visual methods to get a two-dimensional representation of high-dimensional data. The more separated are the clusters, the more discriminative has been the clustering and is gratifying. However, there is some caution to have when interpreting a tSNE plot. Moreover, this algorithm will represent two points that are really close in the high dimension as well close in the two dimension. Thus, those points will be represented close to each other. But, the opposite is not always correct. Two points that are close in the two dimensional plot are not necessarily close in the high dimension. It is thus recommended to run it a few times, taking different initial randomly disposed points. It will then give more robustness to its interpretation. Another drawback of tSNE is its computational time. When tSNE are stratified by samples as in *figure* 25, it gives a quick look on the data and allows a rapid identification of the samples which are lacking any major population. Comparing the tSNE plots of healthy donors and patients in *figure* 26, no obvious loss of population appeared in the patients group, as well as no new populations could be identified. Finally, the comparison between immune cells populations proportions that were manually and semi-automatically obtained shows a good concordance of those two methods, as illustrated in *figure* 27. There is a tendency of the semi-automatic method to over-classify T cells into CD4+ and to under-classify them into CD8+, compared to the manual expert gating. This can be explained by the fact that the indium metal 113 used for CD8 also has about 10% staining in the CD4 indium 115. For manual gating, this is easy to see and correct for, but the automatic gating has more of a problem. This is however the only example of an antibody that stains in two metals at once (i.e. CD8 that will have 113 and 115). Given what is known in the literature for defining the major immune cell populations in *figure* 27, it can be concluded that manual gating using the expertise of a trained immunologist is better at defining these populations. However, in general the unsupervised analysis was fast and reasonably good in this task as well.

It is important to be aware that immunodeficiencies can clinically present themselves with similar symptoms, but having different underlying immune deficiencies. When interpreting the differential analysis results, it is essential to be aware that in this work the patients group has been selected based uniquely on clinical criteria. The patient set being evaluated is therefore heterogeneous with respect to their cellular phenotype. Thus, after obtaining our results, it would be interesting to perform additional analysis to group patients according to their immune cell specific deficiencies. It is hence essential to have a bigger dataset of patients to get significant and robust results. For instance, most immunodeficiencies are B cell related. When looking at the B cell populations throughout the

patients in *figure* 25, it can be seen that patients 3, 4, 10 and 15 have low proportions of B cells compared to the healthy donors and other patients. Thus, having a bigger dataset of patients, it would be possible to cluster those different patients showing this phenotype to drive further analysis and to try to find out why they have this decrease of B cells.

When looking at the cytokines expressions from the different cells populations, no statistically significant results were found, even though in *figure* 29 it seems to be less IL-12 in the mDCs of patients compared to healthy donors. However, in the following section focusing on mDCs, differences will be seen. That can be a hint that a combination of those two analysis approaches would be ideal to optimally compare the populations. Moreover, another possibilities would have been to drive the analysis of clustering and subpopulations identification specifically on the mDC subset, in a similar way as in the approach using manual gating. Then, comparing the subpopulations of mDC might have bring to light differences. However, when analysing the expression of activation markers in the different populations as shown in *figures* 32 and 33, both in the unstimulated and stimulated conditions there was a statistically significant increase in activation markers in different immune cells populations. This might be interpreted as an over-activation of the immune system in the patients groups in a context of permanent inflammation, due to a disease state or a dysfunctionality in another arm of the immune system that induces the over-activation of other cells populations. Once again, it would be interested to group the patients according to their molecular signatures and to look for underlying deficiencies in those cells.

This analysis provides a rapid overview of the statistical differences that are observed in the different immune cell populations between the healthy donor and patient populations. However, interpretation of this data with the mindset of an immunologist is required to help identify areas where unsupervised analysis is overinterpreting small differences. For example, looking at *figures* 32 and 33, CD86 (B7-2 receptor) is shown to be overexpressed in most immune cell populations of patients compared to healthy donors. Automated analysis even noted this in T cell populations that should not significantly express CD86. Therefore, a combination of unsupervised and supervised verification is a necessary part of the analysis pipeline.

To conclude this first part of all cells analysis, only few significant differences could be seen between the patients and the healthy donors group. It shows the importance to go even deeper in the analysis, thus concentrating on one cell subpopulation. This approach, as it will be shown below, will allow to bring to light relevant differences and ideas for further analysis.

4.2 Myeloid dendritic cells analysis

The initial plots obtained, as in *figure* 34, to have a rough look on our data shows as well variations in the number of cells in each sample. As mentioned above, this can be attributed to a heightened immune and/or inflammatory response. A further reason could be due to the manual gating variability that does exist. As proportions will be analysed, it should not bias our results interpretations. On the NRS plot shown in *figure* 35, it is interesting to see that the marker discriminating the most between all of the cells events is CD1c, defining the conventional dendritic cells subset, an important mDC subfamily that will be discussed below.

The minimal spanning tree coloured according to the CD1c expression (*figure* 36) gives the position of the clusters corresponding to the conventional dendritic cells. This is a way to verify and semi-supervise the quality of the clustering, showing that meaningful biological differences are taken into account and that the well-known subpopulations are being detected with this approach.

Figures 38 and 39 give a direct histogram representation of the surface markers and cytokines intensities among the 20 mDC clusters obtained with FlowSOM and ConsensusClusterPlus. Different molecular signatures can be clearly seen between each of the clusters. It is also possible to see that clusters 10 and 12 express a higher level of CD66b, which is normally associated with granulocytes. Therefore, it is most likely that some of the granulocytes have not been gated out during the initial manual gating. With respect to the cytokines expression, subpopulations of mDCs can be identified that secrete higher levels of TNFα, IL-1b, IL-12p40, or IL-6. Comparisons between the healthy donors and patients group was therefore relevant to look for expression intensities differences between those two populations. Furthermore, it has to be noted that those statistically significant differences in cytokines expression could only be seen when cells were manually subdivided into separate clusters.

The heatmap shown in *figure* 37 is of an extreme importance to distinguish different subpopulations. Clusters 15, 18, 19 and 20 have the phenotype of conventional dendritic cells (CD1c+ CD16-), whereas clusters 4, 6, 9, and 12 have the phenotype of inflammatory dendritic cells (CD1c- CD16+).

Looking at the tSNE plots stratified by samples, some patients seem to present conventional DCs that have undertaken a transition from a weakly (cluster 19) to a highly (cluster 20) activated state, like patient 01, or patient 02 for instance (represented in *figure* 43). Indeed, they seem to have an increased cluster 20, which present higher activated surface molecules like HLA-DR, CD38, CD69, and CD86 (see *figure* 37), in comparison to healthy donors. Looking at the mDC cytokines expression, there is a statistically significant decrease of the expression of pro-inflammatory cytokines IL-1b, IL-6, IL-12p40, and TNFα when they are stimulated by LPS and R848. This might indicate that those DCs are exhausted from being constantly activated, and thus less capable of secreting pro-inflammatory cytokines when they are stimulated. To further this investigation, TLR4 (pathogen recognition receptor - PRR - for LPS), TLR7 or TLR8 (PRRs for R848) intracellular signaling pathways phosphorylation state could be investigated. This could highlight upstream molecular defects responsible for this dysfunction.

Some functional immune cells defects have already been raised in primary immunodeficiencies, particularly in what concerns dendritic cells. For instance, it has already been shown that TLR7 and 9 can present defects in CVID (52). It would thus be interesting to see if similar findings can be highlighted in this patients groups. Another deficiency of dendritic cell function in CVID was a deficiency in IL-12 secretion (53). This highlights the relevance of our approach in examining immune cells populations and bringing to light molecular and cellular mechanisms responsible for their pathological states.

4.3 Analysis pipeline general discussion

This project shows the power of using high dimensional mass cytometry analysis in combination with bioinformatics algorithms. Using this analysis pipeline, a better understanding of immunodeficiencies molecular and cellular mechanisms can be achieved. This master's thesis illustrates and justifies the place of such an approach. It is important to keep in mind that manual gating, used in the semi-unsupervised workflow in order to manually isolate known immune cell populations, can be subject to inter-operator variability, which would alter its reproducibility. However, the different immune cell populations are referred to as lineages and help form a framework for the more in depth clustering within these lineages (e.g. T cells, NK cells, B cells, monocytes). Since these lineages are well defined, there would not be much room for inter-operator variations. Nevertheless, aspiring to a fully automated clustering method would definitely remove a certain subjectivity in the populations gating, as well as increase data processing speed and analysis robustness.

Many patients suffering from any diseases in the primary immunodeficiencies spectrum do not have a diagnosis, or have a general diagnosis to characterize their troubles. A likely explanation is that we have no applied the right tools for an in-depth understanding of the underlying causes of these diseases. However, now this approach can help to unveil new pathophysiological mechanisms and potentially bring clarity in the diagnosis of immunodeficiencies subtypes. Immune profiles will be discovered that will try to be correlated with specific clinical pictures, evolution, treatment and prognosis in order to best clinically manage immunodeficient patients. It is as well a great opportunity to discover new immunodeficiencies, as well as to contribute to explain the variety of clinical manifestations related with immunodeficiencies. Regarding the management approach of immunodeficient patients, better characterizing each immunodeficiency might also allow to directly choose the right treatment, for example the choice of an aggressive hematopoietic stem cell transplant therapy compared to intravenous immunoglobulins.

More differences were seen when manual gating was performed to identify the major known immune cells before the unsupervised analysis. It shows the importance of this method even if it relies on the expertise of the "manual gater". Further improvements and optimisation in the automatic gating have to be made to achieve a robust, fully objective and fast immune cells automated clustering in harmony with biological considerations and relevancies.

As mention earlier, the patients group contained patients presenting different kind of immunodeficiencies, as the only including condition was that they should present a clinical suspicion of a immunodeficiency. Thus, it would be of great interest to include patients with known mutations identified by NGS that affect specific immune cell populations such as B cells. That way, the profile of B cells in these two patient populations can be compared to identify the suspected immunodeficiency patients with B cells that have a highly similar profile to the patients with a specific mutation.

In the analysis pipeline, many other analysing tools have not been used in this project but might present as well high interest. For instance, machine learning algorithm and other artificial intelligence approaches should also be of a great help to bring light to this large amount of data. Those include fully automated exhaustive projection pursuit (EPP)

clustering (54), quadratic form (QF)-based cluster matching algorithm (55), Citrus (56), deep learning (57), neural network (58), or other machine learning systems (59).

It is a real challenge to make sense in the analysis of such large data. Lots of information are generated with such methods, and it is sometimes difficult to identify what is relevant and what is not. Strong biological knowledge is essential to identify and target results that are meaningful. For instance, it is not because a statistically significant result is found that it means that it is biologically meaningful, and even less that it is clinically relevant. The interpretation of this amount of results can be therefore challenging.

The main goal of this work was to set up an analysis pipeline to go deeper in the functional analysis of immune cells. Therefore, one major limitation of this work is that the patients analysed were not chosen with strict criteria. Only one criterion applied, which was the clinical suspicion of presenting an immunodeficiency, without highlighting an obvious quantitative cell defect. Thus, no conclusions can really be made about the obtained results, and the population of patients should be better selected. Another limitation is that no correlation with the clinical picture of the patients were made. This is thus a perspective for future work on this subject, trying to exactly match a immunodeficiency with specific clinical pictures.

The major contribution of this research resides in directly testing the functional activity and response of a patient's immune cells to different stimuli. The highly multiparametric nature of mass cytometry allows for both a broad and in depth characterization of the functional immune response using only a minimal volume of a patient's blood (1 mL) with results available within one day, thus drastically improving time to diagnosis. In addition to having a proportional and phenotypic characterization of a patient's immune cells, identifying the functionally abnormal cell population(s) will provide the clinicians with an even better understanding of their patient's immunological defect. Interpretation of the mass cytometry results along with the patient's clinical data will allow for the identification of signatures associated with specific immunological defects, new classes of immunodeficiencies and therapies that are best adapted to a specific class of an immunological disorder, hence improving the diagnosis and the benefits for immunocompromised patients.

5 Perspectives

The analysis pipeline shown in this work is just the first step towards the improvement in the clinical management of immunocompromised patients. When an altered deep immune cells profiling is observed, as it was in the conventional dendritic cells subpopulations, the analysis can first be completed by looking at the exome of the corresponding patient using next generation sequencing. This molecular characterization is thought to be essential in order to find a genetic alteration that could be responsible for the perturbed inflammatory profile of the patients' immune cells. Then, and this is the main point of all translational researches, the different subtypes of alterations in the molecular signatures of patients' immune cells will be source of new identification of immunodeficiency subtypes. Knowing for each biological immune deficiency subtype which is its corresponding clinical phenotype, will allow to chose rapidly for the most adequate treatment, as well as to know what is to expect in the clinical evolution and

how the medical follow-up should be adapt. In this way the condition of immunocompromised patients could be improved.

Better understanding the immune system is compulsory if it is desired to offer a better medical follow up to immunocompromised patients and to rationnally develop better therapies. However, it is not only limited to the better care of immunodeficient patients; indeed, the central place of the immune system in the organism makes it interact with many different other systems. Having a better molecular comprehension of it will also offer new prospects in other medical fields.

Finally, this approach opens other analysis perspectives in looking at the different interactions the immune system can have and its different modulators. For instance, the interactions between the immune system and the microbiome is essential in the human body's homeostasis. Indeed, the innate immune system holds a close communication with commensal organisms at the host-microbiome interface. They both influence each other; for instance, it has been shown that myelopoiesis depends on the complexity of the intestinal microbiota, or that ILCs owe their proper functions to commensal microbial colonization (60). This has not been further developed in this work, but it will definitely be an important aspect to take into account in future investigations and which may help us understand better immunological diseases and their pathophysiology.

6 Acknowledgement

I am particularly grateful to Prof Giuseppe Pantaleo, chief of the division of immunology and allergy of the CHUV, and Dr Craig Fenwick, director of the immunology and allergy immune monitoring platform at the CHUV, without whom this work could not have been done and who integrated me quickly in the laboratory team to initiate me with the laboratory work. I would also like to thank Prof Fabio Candotti, clinical director of the vaccine and immunotherapy center at the CHUV, Prof Matthieu Perreau, division of immunology and allergy at the CHUV, and Dr Denis Compte, division of immunology and allergy at the CHUV, for their support. I am also especially indebted to Ms Line Esteves-Leuenberger, Ms Manon Geiser, Ms Madeleine Suffiotti, Mr Michael Moulin, Ms Nadine Moreira Do Rosario, Mr Alex Farina, Dr Jaap Herderschee, co-workers at the Vaccine and Immunotherapy Center at the division of Immunology and Allergy of the CHUV, for their support, teaching and generous time.

References

[1] V HA, Moss PAH. Essential Haematology, 6th Edition. Wiley-Blackwell; 2011.

[2] Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. Nature Reviews Immunology. 2012 Mar;12(3):191–200. Available from: http://www.nature.com/articles/nri3158.

[3] Poli A, Michel T, Thérésine M, Andrès E, Hentges F, Zimmer J. CD56 ^{bright} natural killer (NK) cells: an important NK cell subset. Immunology. 2009 Apr;126(4):458–465.

[4] Desbois M, Rusakiewicz S, Locher C, Zitvogel L, Chaput N. Natural killer cells in non-hematopoietic malignancies. Frontiers in Immunology. 2012;3.

[5] Collin M, Bigley V. Human dendritic cell subsets: an update. Immunology. 2018 May;154(1):3–20. Available from: http://doi.wiley.com/10.1111/imm.12888.

[6] Ziegler-Heitbrock L. The CD14+ CD16+ blood monocytes: their role in infection and inflammation. Journal of Leukocyte Biology. 2007 Mar;81(3):584–592.

[7] Abbas AK, Lichtman AH, Pillai S. Cellular and Molecular Immunology. Elsevier; 2017.

[8] Recher M. Déficit immunitaire primaire - lorsque nous sommes nos gènes. Swiss medical forum. 2019;19. Available from: https://doi.org/10.4414/fms.2018.08014.

[9] Schmidt RE, Grimbacher B, Witte T. Autoimmunity and primary immunodeficiency: two sides of the same coin? Nature Reviews Rheumatology. 2018 Jan;14(1):7–18. Available from: http://www.nature.com/articles/nrrheum.2017.198.

[10] Notarangelo LD. Primary immunodeficiencies. Journal of Allergy and Clinical Immunology. 2010 Feb;125(2):S182–S194. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0091674909011622.

[11] Modell V, Knaus M, Modell F, Roifman C, Orange J, Notarangelo LD. Global overview of primary immunodeficiencies: a report from Jeffrey Modell Centers worldwide focused on diagnosis, treatment, and discovery. Immunologic Research. 2014 Oct;60(1):132–144. Available from: http://link.springer.com/10.1007/s12026-014-8498-z.

[12] Jameson JL, editor. Harrison's principles of internal medicine. Twentieth edition ed. New York: McGraw-Hill Education; 2018.

[13] Bousfiha A, Jeddane L, Picard C, Ailal F, Bobby Gaspar H, Al-Herz W, et al. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. Journal of Clinical Immunology. 2018 Jan;38(1):129–143. Available from: http://link.springer.com/10.1007/s10875-017-0465-8.

[14] Fischer A, Provot J, Jais JP, Alcais A, Mahlaoui N, Adoue D, et al. Autoimmune and inflammatory manifestations occur frequently in patients with primary immunodeficiencies. Journal of Allergy and Clinical Immunology. 2017 Nov;140(5):1388–1393.e8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0091674917302269.

[15] Stiehm ER. UpToDate: Primary immunodeficiency: Overview of management;. Accessed: 2019-11-30. https://www.uptodate.com/contents/primary-immunodeficiency-overview-of-management?search=primary%20immunodeficiency&source=search_result&selectedTitle=3~150&usage_type=default&display_rank=3#H2253771881.

[16] Milito C, Pulvirenti F, Cinetto F, Lougaris V, Soresina A, Pecoraro A, et al. Double-blind, placebo-controlled, randomized trial on low-dose azithromycin prophylaxis in patients with primary antibody deficiencies. Journal of Allergy and Clinical Immunology. 2019 Aug;144(2):584–593.e7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0091674919303884.

[17] Slatter MA, Gennery AR. Hematopoietic cell transplantation in primary immunodeficiency – conventional and emerging indications. Expert Review of Clinical Immunology. 2018 Feb;14(2):103–114. Available from: https://www.tandfonline.com/doi/full/10.1080/1744666X.2018.1424627.

[18] Bailey SR, Maus MV. Gene editing for immune cell therapies. Nature Biotechnology. 2019 Jun;Available from: http://www.nature.com/articles/s41587-019-0137-8.

[19] Markert ML, Devlin BH, McCarthy EA. Thymus transplantation. Clinical Immunology. 2010 May;135(2):236–246. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1521661610000483.

[20] Markert ML, Marques JG, Neven B, Devlin BH, McCarthy EA, Chinn IK, et al. First use of thymus transplantation therapy for FOXN1 deficiency (nude/SCID): a report of 2 cases. Blood. 2011 Jan;117(2):688–696. Available from: https://ashpublications.org/blood/article/117/2/688/28144/First-use-of-thymus-transplantation-therapy-for.

[21] Chinen J, Shearer WT. Secondary immunodeficiencies, including HIV infection. Journal of Allergy and Clinical Immunology. 2010 Feb;125(2):S195–S203. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0091674909013220.

[22] Amir EaD, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nature Biotechnology. 2013 Jun;31(6):545–552.

[23] Hartmann FJ, Bernard-Valnet R, Quériault C, Mrdjen D, Weber LM, Galli E, et al. High-dimensional single-cell analysis reveals the immune signature of narcolepsy. The Journal of Experimental Medicine. 2016 Nov;213(12):2621–2633.

[24] Montaldo E, Zotto GD, Chiesa MD, Mingari MC, Moretta A, Maria AD, et al. Human NK cell receptors/markers: A tool to analyze NK cell development, subsets and function: Human NK Cell Receptors/Markers. Cytometry Part A. 2013 Aug;83A(8):702–713.

[25] Becher B, Schlitzer A, Chen J, Mair F, Sumatoh HR, Teng KWW, et al. High-dimensional analysis of the murine myeloid cell system. Nature Immunology. 2014 Dec;15(12):1181–1189.

[26] Xu Y, McKenna RW, Karandikar NJ, Pildain AJ, Kroft SH. Flow Cytometric Analysis of Monocytes as a Tool for Distinguishing Chronic Myelomonocytic

Leukemia From Reactive Monocytosis. American Journal of Clinical Pathology. 2005 Nov;124(5):799–806.

[27] Rundberg Nilsson A, Bryder D, Pronk CJH. Frequency determination of rare populations by flow cytometry: A hematopoietic stem cell perspective: Flow Cytometry and Infrequent Cells. Cytometry Part A. 2013 Aug;83A(8):721–727.

[28] Hensley-McBain T, Heit A, De Rosa SC, McElrath MJ, Andersen-Nissen E. Optimization of a whole blood phenotyping assay for enumeration of peripheral blood leukocyte populations in multicenter clinical trials. Journal of Immunological Methods. 2014 Sep;411:23–36.

[29] Moskalensky AE, Chernyshev AV, Yurkin MA, Nekrasov VM, Polshchitsin AA, Parks DR, et al. Dynamic quantification of antigen molecules with flow cytometry. Journal of Immunological Methods. 2015 Mar;418:66–74.

[30] Brown M, Wittwer C. Flow Cytometry: Principles and Clinical Applications in Hematology. Flow Cytometry. 2000;(8):9.

[31] McKinnon KM. Flow Cytometry: An Overview. Current Protocols in Immunology. 2018 Jan;120(1). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpim.40.

[32] Bjornson ZB, Nolan GP, Fantl WJ. Single-cell mass cytometry for analysis of immune system functional states. Current Opinion in Immunology. 2013 Aug;25(4):484–494.

[33] Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. Cell. 2016 May;165(4):780–791.

[34] Bendall SC, Simonds EF, Qiu P, Amir EaD, Krutzik PO, Finck R, et al. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. Science. 2011 May;332(6030):687–696. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1198704.

[35] Gadalla R, Noamani B, MacLeod BL, Dickson RJ, Guo M, Xu W, et al. Validation of CyTOF Against Flow Cytometry for Immunological Studies and Monitoring of Human Cancer Clinical Trials. Frontiers in Oncology. 2019 May;9. Available from: https://www.frontiersin.org/article/10.3389/fonc.2019.00415/full.

[36] Mei HE, Leipold MD, Schulz AR, Chester C, Maecker HT. Barcoding of Live Human Peripheral Blood Mononuclear Cells for Multiplexed Mass Cytometry. The Journal of Immunology. 2015 Feb;194(4):2022–2031. Available from: http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1402661.

[37] Finck R, Simonds EF, Jager A, Krishnaswamy S, Sachs K, Fantl W, et al. Normalization of mass cytometry data with bead standards. Cytometry Part A. 2013 May;83A(5):483–494.

[38] Malgorzata Nowicka, Carsten Krieg, Lukas M Weber, Felix J Hartmann, Silvia Guglietta, Burkhard Becher, et al. CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets [version 2; referees: 2 approved]. F1000Research. 2017 Nov;6(48).

[39] Mair F, Hartmann FJ, Mrdjen D, Tosevski V, Krieg C, Becher B. The end of gating? An introduction to automated analysis of high dimensional cytometry data: Highlights. European Journal of Immunology. 2016 Jan;46(1):34–43. Available from: http://doi.wiley.com/10.1002/eji.201545774.

[40] Weber LM, Robinson MD. Comparison of Clustering Methods for High-Dimensional Single-Cell Flow and Mass Cytometry Data. bioRxiv. 2016 Nov;.

[41] Saeys Y, Gassen SV, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. Nature Reviews Immunology. 2016 Jul;16(7):449–462.

[42] Hagen JB. The origins of bioinformatics. Nature Reviews Genetics. 2000 Dec;1(3):231–236.

[43] Diggins KE, Ferrell PB, Irish JM. Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data. Methods. 2015 Jul;82:55–63.

[44] Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, et al. flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics. 2009;10(1):106.

[45] Chen H, Lau MC, Wong MT, Newell EW, Poidinger M, Chen J. Cytofkit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline. PLOS Computational Biology. 2016 Sep;12(9):e1005112.

[46] Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data: FlowSOM. Cytometry Part A. 2015 Jul;87(7):636–645.

[47] Laurens van der Maaten, Geoffrey Hinton. Visualizing Data using t-SNE. Journal of Machine Learning Research. 2008 Nov;9:2579–2605.

[48] Kriegel HP, Kröger P, Zimek A. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Transactions on Knowledge Discovery from Data. 2009 Mar;3(1):1–58. Available from: http://portal.acm.org/citation.cfm?doid=1497577.1497578.

[49] Sammut C, Webb GI, editors. Encyclopedia of Machine Learning. Boston, MA: Springer US; 2010. Available from: http://link.springer.com/10.1007/978-0-387-30164-8.

[50] Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution. 2009 Mar;24(3):127–135. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0169534709000196.

[51] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995 Jan;57(1):289–300. Available from: http://doi.wiley.com/10.1111/j.2517-6161.1995.tb02031.x.

[52] Yu JE, Knight AK, Radigan L, Marron TU, Zhang L, Sanchez-Ramón S, et al. Toll-like receptor 7 and 9 defects in common variable immunodeficiency. Journal of Allergy and Clinical Immunology. 2009 Aug;124(2):349–356.e3. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0091674909008100.

[53] Cunningham-Rundles C, Radigan L. Deficient IL-12 and dendritic cell function in common variable immune deficiency. Clinical Immunology. 2005 May;115(2):147–153. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1521661604003730.

[54] Meehan S, Kolyagin GA, Parks D, Youngyunpipatkul J, Herzenberg LA, Walther G, et al. Automated subset identification and characterization pipeline for multidimensional flow and mass cytometry data clustering and visualization. Communications Biology. 2019 Dec;2(1). Available from: http://www.nature.com/articles/s42003-019-0467-6.

[55] Orlova DY, Meehan S, Parks D, Moore WA, Meehan C, Zhao Q, et al. QF-Match: multidimensional flow and mass cytometry samples alignment. Scientific Reports. 2018 Dec;8(1). Available from: http://www.nature.com/articles/s41598-018-21444-4.

[56] Bruggner RV, Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. Proceedings of the National Academy of Sciences. 2014 Jul;111(26):E2770–E2777.

[57] Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, et al. Predicting HLA class II antigen presentation through integrated deep learning. Nature Biotechnology. 2019 Nov;37(11):1332–1343. Available from: http://www.nature.com/articles/s41587-019-0280-2.

[58] Amodio M, van Dijk D, Srinivasan K, Chen WS, Mohsen H, Moon KR, et al. Exploring single-cell data with deep multitasking neural networks. Nature Methods. 2019 Nov;16(11):1139–1145. Available from: http://www.nature.com/articles/s41592-019-0576-7.

[59] Tomic A, Tomic I, Rosenberg-Hasson Y, Dekker CL, Maecker HT, Davis MM. SIMON, an Automated Machine Learning System, Reveals Immune Signatures of Influenza Vaccine Responses. The Journal of Immunology. 2019 Aug;203(3):749–759. Available from: http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1900033.

[60] Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. Nature. 2016 Jul;535(7610):65–74. Available from: http://www.nature.com/articles/nature18847.

[61] Clark R, Kupper T. Old Meets New: The Interaction Between Innate and Adaptive Immunity. Journal of Investigative Dermatology. 2005 Oct;125(4):629–637. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022202X15324763.

[62] Chaplin DD. Overview of the immune response. Journal of Allergy and Clinical Immunology. 2010 Feb;125(2):S3–S23. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0091674909028371.

[63] Liew FY, Xu D, Brint EK, O'Neill LAJ. Negative regulation of Toll-like receptor-mediated immune responses. Nature Reviews Immunology. 2005 Jun;5(6):446–458. Available from: http://www.nature.com/articles/nri1630.

[64] Moreth K, Iozzo RV, Schaefer L. Small leucine-rich proteoglycans orchestrate receptor crosstalk during inflammation. Cell Cycle. 2012 Jun;11(11):2084–2091. Available from: http://www.tandfonline.com/doi/abs/10.4161/cc.20316.

[65] Pantaleo G. Lecture notes in immunology, 2nd year medical school. Service of Immunology and Allergy - IAL, CHUV, Lausanne, Switzerland; 2015.

[66] Calandra T. Lecture notes in innate immunity, 2nd year medical school. Service of Infectious diseases, CHUV, Lausanne, Switzerland; 2015.

[67] Karasuyama H, Mukai K, Tsujimura Y, Obata K. Newly discovered roles for basophils: a neglected minority gains new respect. Nature Reviews Immunology. 2009 Jan;(9):9–13. Available from: https://www.nature.com/articles/nri2458.

[68] Sokol CL, Medzhitov R. Emerging functions of basophils in protective and allergic immune responses. Mucosal Immunology. 2010 Mar;3(2):129–137. Available from: http://www.nature.com/articles/mi2009137.

[69] Voehringer D. Protective and pathological roles of mast cells and basophils. Nature Reviews Immunology. 2013 May;13(5):362–375. Available from: http://www.nature.com/articles/nri3427.

[70] Van Acker HH, Capsomidis A, Smits EL, Van Tendeloo VF. CD56 in the Immune System: More Than a Marker for Cytotoxicity? Frontiers in Immunology. 2017 Jul;8. Available from: http://journal.frontiersin.org/article/10.3389/fimmu.2017.00892/full.

[71] Strandmann EV, Shatnyeva O, Hansen HP. NKp30 and its ligands: emerging players in tumor imune evasion from natural killer cells. Annals of Translational Medicine. 2015;3. Available from: http://dx.doi.org/10.3978/j.issn.2305-5839.2015.09.08.