From Department of Women's and Children's Health
Karolinska Institutet, Stockholm, Sweden

Inference of systems-level behaviors in the immune system from single-cell data

Dieudonné Nkulikiyimfura

Karolinska Institutet

Stockholm 2020

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Cover: Figure by Dieudonné Nkulikiyimfura.
Printed by US AB

ISBN 978-91-7831-774-5

Inference of systems-level behaviors in the immune system from single-cell data

THESIS FOR DOCTORAL DEGREE (Ph.D.)

by due permission of Karolinska Institutet, will be publicly defended in Lecture Hall Louis at Widerströmska, Tomtebodavägen 18A, Solna.

Friday 17^{th} of April 2020 at 9am

By

Dieudonné Nkulikiyimfura

Principal Supervisor:
Petter Brodin, Associate Professor
Karolinska Institutet
Department of Woman's and Children's Health
Division of Clinical Pediatrics

Co-supervisor(s):
Jeff Mold, Associate Professor
Karolinska Institutet
Department of Cell and Molecular Biology

Opponent:
Mauno Vihinen, Professor
Lund University
Department of Experimental Medical Science

Examination Board:
Dirk Repsilber, Professor
Örebro University
School of Medical Sciences

Jane Synnergren, Associate Professor
Skövde University
School of Bioscience
Systems Biology Research Center

Björn Wallner, Professor
Linköping University
Department of Physics, Chemistry and Biology
Division of Bioinformatics

To my daughter and wife

Abstract

Understanding the systems-level behaviours of the immune system requires profiling many parameters of its constituents as well as identifying environmental factors influencing the time evolution of immune response. Recent advances in technologies enable multiple cellular components to be measured simultaneously at an unprecedented scale and resolution across many individuals, and the resulting data could be used to meet this goal. Such data allows studying the immune system as a whole enabling a holistic approach. The goal here is to analyze such high-dimensional data by examining the measured parameters interdependence in order to infer the emergent behaviours of the immune system. This thesis involves different techniques to analyze various datasets in order to reveal new insights into mechanisms of human immunity of relevance to environmental exposures. This dissertation presents results from four studies.

I. The human immune variation is continuous, rather than described by discrete groups of individuals with similar immune cell populations. How can collective states of many immune system components describe immune variation across individuals? Using partial least squares method, we derived a set of aggregate immune cell population frequencies that define an individual's immunotype, and robustly predict diverse functional responses to cytokine stimulations. In immunotype space, individuals of younger age are similar to one another than older individuals are. Cytomegalovirus seropositivity induces a shift of one's immunotype towards a more aged immunotype.

II. Mothers transfer antibodies but we don't know what their composition looks like. What are the transferred maternal antibodies? How large is the repertoire of maternal antibodies, their specificities, and duration after birth? Using VirScan method, we assayed around 10^7 antibody-peptide interactions in mother-child dyads. The repertoire of antiviral maternal antibodies target between 5-10 different viruses, and the transferred antibodies mirrors those found in the mothers. Although IgG transfer happen principally during the final trimester, very preterm ($<$30 weeks of gestation) and term ($>$37 weeks of gestation) children receive a similar repertoire. However, the concentrations of antibodies at birth are lower in preterm than in term children, and determine how long the conferred immunity lasts for.

III. Newborns adapt to living outside the womb, suddenly exposed to new bacteria and viruses, which leads to a rapid biological change of human

newborn immune system. Can phenotypic variants of developing human newborn lymphocyte be explained regarding of trade-offs between specialist and generalist phenotypes? In immune system, individual cells face a dilemma. No single cell can be optimally suited for all possible tasks, and therefore cells specialize to perform specific tasks. For example, in the human immune system, cytotoxic lymphocyte kills virus-infected cells, and B-lymphocytes produce antibodies, etc. Using Pareto archetype analysis, we learned geometrical shapes of protein expression space from longitudinal human newborn lymphocyte data. Single B cells are arranged in a triangle, while CD4+ T cells are best represented by pentahedron. The vertices of these shapes are extreme protein expression profiles optimal for tasks and correspond to major cell subsets. Cells lie along a continuum of expression inside polytope. In triangle B cells, a 1D continuum of states describes cells specialization pattern to tasks and suggests pseudo-time trajectories in the developmental path of the newborn B cells.

IV. The variation of transcriptional responses to microbial stimulants is large among primary immunodeficiency disorder (PID) patients, and little in healthy individuals. How can functional defects in PID patients be inferred from transcriptional variation of human immune responses to bacterial and viral challenges? We deduced a collective set of genes that can predict variation of transcriptional responses to stimulant antigens in PID patients. Lastly, we identified gene variants associated with the differences in transcriptional responses between PID patients and healthy individuals allowing understanding immune functional defects in patients.

LIST OF SCIENTIFIC PAPERS

I. **Continuous immunotypes describe human immune variation and predict diverse responses**
Kevin J. Kaczorowski, Karthik Shekhar, Dieudonné Nkulikiyimfura, Cornelia L. Dekker, Holden Maecker, Mark M. Davis, Arup K. Chakraborty, and Petter Brodin
Proc Natl Acad Sci USA. National Acad Sciences; 2017 Jul 25;114(30):E6097-106.

II. **The repertoire of maternal anti-viral antibodies in human newborns**
Christian Pou*, Dieudonné Nkulikiyimfura*, Ewa Henckel, Axel Olin, Tadepally Lakshmikanth, Jaromir Mikes, Jun Wang, Yang Chen, Anna Karin Bernhardsson, Anna Gustafsson, Kajsa Bohlin, Petter Brodin
Nature Medicine, 25(4), 591–596.

III. **Analyses of phenotypic trade-offs to understand human lymphocyte development**
Dieudonné Nkulikiyimfura, Tadepally Lakshmikanth, Yang Chen, Jaromir Mikes, Ewa Henckel, Anna Karin Bernhardsson, Kajsa Bohlin, Petter Brodin
Manuscript

IV. **A transcriptional method to understand functional defects in patients with primary immunodeficiencies.**
Dieudonné Nkulikiyimfura, Christian Pou, Tadepally Lakshmikanth, Jaromir Mikes, Peter Bergman, Petter Brodin
Manuscript

* These authors contributed equally

Contents

List of Abbreviations

BCG	Bacillus Calmette-Guérin
CD	Cluster of Differentiation
CMV	Cytomegalovirus
DNA	Deoxyribonucleic Acid
EBV	Epstein-Barr virus
ELISA	Enzyme-linked Immunosorbent Assay
HSV	Herpes Simplex Virus
IAV	Influenza A virus
LPS	Lipopolysaccharidea
LV	Latent Variable
MHC	Major Histocompatibility Complex
NK	Natural Killer
PCA	Principal Component Analysis
PCHA	Principal Convex Hull Analysis
PCR	Principal Component regression
PID	Primary Immunodeficiency Disorder
PLS-DA	Partial Least Square Discriminant Analysis
PLS	Partial Least Squares
RS	Respiratory Syncytial
VDJ	Variable, Diversity and Joining
cDNA	complementary Deoxyribonucleic Acid
mRNA	messenger Ribonucleic Acid
mavAb	Maternal antiviral Antibodies
t-SNE	t-distributed Stochastic Neighbor Embedding

Chapter 1

Introduction

1.1 Basic immunology

Few years after Charles Darwin's theory on diversity and natural selection was published in 1859, infectious diseases were discovered, and the first postulate of immune system emerged as an active force of the human body defense against pathogens, and since then the theory has helped researchers in studying how our immune defenses arose [1, 2].

In subsequent years, immunologists realized that our immune defense depends on a system of diversity with many specialized cells to tasks. The immune system is commonly distinguished in two categories according to their defense mechanisms: innate immunity refers to a non-specific defense mechanism with an immediate response. The discovery of B and T-cell receptors [3, 4] that recognize proteins of pathogens has led to a new form of immunity, adaptive immunity, which refers to antigen-specific immune response. Ehrlich and Metchnikoff pioneered two competing defense mechanisms consisting of this latter: the humoral system of antibodies and cell-mediated immunity from white blood cells, for which they received the Nobel prize in 1908.

It took a half-century to gain a comprehensive understanding of achieving diversification of immune responses to all existing pathogens or molecules. Burnet suggested that lymphocytes are key evolutionary players being selected in the body to undergo a process called clonal selection [5]. Later in the 1970s, the clonal selection theory was complemented by a process called somatic Variable, Diversity and Joining (**VDJ**) recombination [6], where Tonegawa demonstrated how maturing B cells could produce a vast array of antibodies. According to these processes, developing lymphocytes rearrange gene segments and result in a highly diverse repertoire of immunoglobulins (Igs) and T-cell receptors found in B-cells and T-cells, respectively, and then

maintained in the body as an immune memory. However, it was until the year 1990 after identifying Major Histocompatibility Complex (**MHC**)-locus genes [7] that the whole process was fully understood.

From a functional point of view, in the late 19th century, immunity emerged as a host defense. Immune functions were associated with wound healing [8], pain signaling [9], clearance of cancer [10], and many more [11]. Recently, studies have identified new immune components such as dendritic [12] and Natural Killer (**NK**) [13] cell lineages, and T helper cell subsets [14]. Such developments in defining molecular and cellular networks have tremendously advanced our understanding of defense mechanisms of immune function.

1.2 Motivation

It is heuristically understood how pathogenic and non-pathogenic signals modulate the immune response of the innate and adaptive immune system. Apart from discriminating between self- versus non/altered-self, the immune system maintains homeostasis by balancing between mounting an appropriate response to some stimuli while remaining unresponsive to others, for example, microbes that are crucial to development, nutrition, and more. This is only achieved through a network of interactions [5, 15] between many specialized immune cells and soluble mediators.

1.3 Scope and outline of the dissertation

In this thesis, a systems approach is adopted to study human immune system behaviors. This approach does not depend on formulating any specific research hypothesis. Otherwise, it would require drawing on ideas and intuitions from our knowledge of the immune system that comes from decades of work on mice [16]. As we will see below, this could be limiting given remarkable differences between mice and humans. In general, our intuition of how the human immune system works is based on observed behaviours. To understand these, immunological observables, e.g., cell frequencies, immune phenotypes, and perturbed genes expressions, are explained in a given condition or over time. Moreover, different contexts (health against diseases, stimulated against non-stimulated) are considered for comparison purposes.

This thesis is organized as follows. It consists of seven chapters. The first chapter introduces briefly basic immunology with a special focus on the immune functions. The second chapter describes the concepts of systems immunology as well as technological advances allowing this. In chapter 3, an introduction to the development of the human immune system related to

the present thesis is briefly given. In chapter 4, the goals of this thesis are described with respect to the presented results. The fifth chapter describes the experimental work leading to generating data and computational methods to analyze the data. Chapter 6 summarizes the results presented in the constituent papers. Finally, conclusions and future perspectives are given in Chapter 7.

Chapter 2

Systems Immunology

2.1 Introduction

The immune system is an intricate network of many cells and molecules interacting to generate an immune response against infection and maintain self-tolerance to host tissues [17–24]. Such a network of interactions is dependent upon changes in immune cell composition over time. Many specialized cell types execute distinct immune functions. They exist in a continuum of development and communicate and modulate each other while preventing autoimmune conditions. During an immune response, changes in immune cell frequencies occur, and this could result in an altered balance between activation and inhibition that causes a complete state-of-shift of the entire immune system.

2.2 Why systems-level view of the human immune system?

2.2.1 Emergent properties of the immune system

Despite this intricate network of interactions between immune components, reductionist experiments have dominated immunology and hindered the study of signaling in the immune system.
From a behavioural perspective, the immune system evolved to mediate a coordinated set of defensive functions via cell-cell interactions and soluble mediators -cytokines- that would optimize successful adaptation [25–27] and survival. The immune system can

- Recognize antigens and induce immune responses
- Retain memories of previously encountered antigens

- Make decision on the best strategy to cope with pathogens

For this latter, like any sensory system, the immune system senses and recognizes molecules of variable compositions, properties, and sources [28, 29]. With this sensing capability, the immune system is not only able to respond or sustain interaction but also to simultaneously deal with several pathogens at the same time without conflicting with maintaining interactions with both microbes and host. This multitasking ability of the immune system requires the simultaneous recall of multiple patterns for reaching homeostasis.

In summary, the immune system performs tasks as a decentralized system. The collective actions of all the different cell populations determine the emergent behaviours of a whole network. However, such behaviours cannot be predicted from analyses of the individual cell populations one by one [30–33].

2.2.2 Applications of basic immunology in clinical problems

The human immune system is actively involved in combating infections and immune system dysfunctions are involved in immunodeficiency disorders and autoimmunity. Moreover, diseases like cancer [34], cardiac diseases, degenerative diseases [35], psychiatric diseases [36], and diabetes are not directly immune related but have a basis of immunological mechanisms [37].

With these conditions, pharmacological control targeting at modulating the host immune response instead of the causative agent itself may be useful in treating diseases. Immunomodulatory treatments such as hematopoietic stem cell transplantation [38], monoclonal antibodies [39], immunosuppressants [40], inflammatory disease [41], and checkpoint therapy [42–45] have contributed significantly for this purpose. While these treatments have been useful, however, responses to these therapies are different across individual patients with unexpected side effects [46, 47]. This exemplifies inadequacy in the current understanding of the human immune system and importance of interindividual variability, indicating its theoretical basis should be updated before being applied in the clinics. Our current knowledge of the immune system has been inferred from extensive studies on a mouse model of complex human disease. For example, type 1 diabetes has been cured multiple times in mice [48] and not in humans.

Several reasons tempting to explain this unsuccessful application of mouse model of diseases to the clinical conditions have been speculated. These include unavoidable compromises of many disease induced-approaches [49], phenotypic imbalance of immune components between humans and mice [50] explained by an evolutionary distance, and the overall difference in lifespan between mice and humans.

Such unsuccessful applications have stimulated a shift from mouse model studies to human studies [51, 52], to better understand human conditions and stimulate progress in developing novel therapies. Herein, one could perform systems analysis of blood that includes sampling white blood cells (lymphocytes) and cytokines to probe communications between cells and tissues. Continuous discovery of various surface markers and cytokines facilitates this. Because of that, immunological health could be defined first and serves not only as metrics to the physicians but also in investigating mechanisms of diseases per se [16, 51].

2.3 Technological advances allowing human system-immunology

Cellular functions and phenotypes depend on a coordinated behavior of many interacting intra-cellular components: genes or proteins and their products, and not on a single gene or protein. These cellular components can be partitioned into modules [53] where they interact with each other, and their combined performances determine the function and states of a cell [54]. With this modular approach [54], the number of degree of freedom that must be considered has been drastically reduced. As a result, cellular characterization requires multiple parameters to be measured simultaneously.

Besides, cells are frequently reacting to different stimuli by changing their phenotype due to induced changes in expression patterns of proteins or genes. Moreover, cells from the same tissue can be inherently heterogeneous in space. Identifying the different types of cells within a differentiated cell population is not possible from one or two of cellular parameters. As a rule of thumb, to understand system-wide cellular function in various contexts, one needs to profile multiple cellular parameters and with single-cell resolution [55, 56].

As a result, the key goals of biology have then been relating to the structure of genes/protein networks with the functions that they perform. In recent years, an effort has shifted from understanding individual genes to understanding all genes or proteins in determining cell functions. Great strides were made in advancing novel technologies, so-called omics-technologies (Proteomics, transcriptomics, and genomics). The introduced technologies have enhanced the throughput, scalability, and the amount of data generated from them. They allowed taking a holistic approach to studying the immune system.

Technological advances have strived to capture cellular components' interactions at different length scales. In immunology, a conventional technology for such cellular analyses is flow cytometry. This technology allows analyzing over 200,000 cells [57] for 15 [58, 59] parameters using fluorescently-labeled antibodies. However, given the limited light spectra, and the already overlapping emission signals, a limited number of different cell populations can be distinguished and characterized phenotypically and functionally. To overcome this limitation, Mass cytometry or Cytometry by time-of-flight (CyTOFTM, Fluidigm Inc) has been developed (detailed description is given in the section below) [60]. Mass cytometry is an antibody-based technology allowing 50 parameters to be quantified simultaneously. As a result of this, we are now able to study phenotypes [61] and function states [62] of all immune cell populations present in human blood. Applications have been extended recently to the studies of messenger Ribonucleic Acid (**mRNA**) [63].

These technologies have focused on capturing the complexity of intercellular signaling comprehensively at the protein level. They provide broad coverage across all cell types in the blood. Clusters of functionally similar cells are prone to overlap in their protein expression profiles rather than being distinct [16, 63]. For an in-depth characterization of cellular interactions, system-wide transcriptional approaches have given more insights into well-defined and rare immune cell populations and their intracellular communication networks [30]. For example, next-generation sequencing has paved a way in studying gene regulatory circuits in immune cells [64], immune cell type heterogeneity [65], and identifying cell fate and clonality of immune cells and their specificity [66, 67].

Depending on a given condition, disease, or stimulation, these high-dimensional technologies offer analyses of overall immune system composition and enable a subsequent analysis of interdependencies between immune components. On the contrary, an individual's immune system encounters many pathogens during life. As a result, understanding an individual's exposome, i.e., collective environmental exposure, is essential. Also, here novel technologies have begun to provide some additional information. For example, the application of next-generation sequencing to measure the microbial composition of fecal, skin, tissue, and blood samples are increasingly done [68]. Also, by analyzing serology in high-throughput, an individual's infectious disease history can be investigated. For example, the VirScan assay [69] measures IgG antibodies to 94,000 peptides representing the entire proteomes of 1000 viral strains known to infect human cells. This method has been used in our projects and will be described in more detail below.

2.4 Advances in data analysis enabling systems immunology

In today's high-dimension biological data deluge, one needs storage and a new mathematical framework that is both solvable and computationally efficient. The nature of raw data from the technologies described above is explained in terms of probes estimates. It contains a lot of noise and technical artifacts given dynamical range of increased number of parameters. As a result, the preprocessing of data is often performed before making any biological interpretation. Also, unlike traditional statistics that deal with a small number of parameters with very little noise, retaining sufficient information from noisy biological data is much more challenging [63, 70].

To deconvolve a cell mixture in a tissue, advances in technology have focused on measuring multiple cellular parameters in a single cell, e.g., mass cytometry and single-cell transcriptome sequencing. Several tools based on unsupervised clustering algorithms have been developed [71–73]. For example, Citrus [74] has been used to delineate cellular subpopulations that display condition-specific behavior between patients and healthy controls. However, given the random realization of high dimensional noise, classical clustering algorithms utilizing dimensionality reduction are prone to discovering (illusory) structures that do not exist in the raw data.

However, data-driven clustering techniques assume that cells cluster in discrete groups; therefore, they may inaccurately model some biological processes given a discrete nature of this approach. An example is cell differentiation, where a cell is undergoing continuous changes from stem cell to a mature phenotype. To circumvent this limitation, tools are developed to learn developmental trajectories from single-cell data [75–77] and have identified transitional cell states and mechanisms that govern their differentiation [78].

From a dynamical point of view, the immune response evolves, and the description of this time evolution is of great interest in systems immunology. The immune system would develop into dynamic regimes that remain qualitatively unchanged given a finite range of parameters, e.g., aging. Do such dynamic regimes exist in the immune system? How could one induce transition among the different dynamical regimes? With the systems immunology approach, there is a slowly growing number of studies that have started to respond to these. For example, variations in immune systems of healthy individuals have shown that the immune cell frequencies of young, identical twins (< 20 years) correlated strongly, while older twins (> 60 years) did not, suggesting a cumulative influence from environmental factors over time that

shape immune system [79]. Similarly, immune cell frequencies of older individuals as a group are more heterogeneous than younger individuals. Human cytomegalovirus infections make an immune phenotype look older irrespective of the actual age of the individual in the immune space [80]. Also, from a theoretical point of view, Agliari et al advised a model describing different immune regimes for B and T-cell interactions explaining the ability of the immune system to respond simultaneously to multiple distinct antigens by combining techniques from statistical mechanics and graph theory [32].

Finally, algorithms for dimensionality reduction are used for visualization purposes or downstream analysis. They capture different aspects of the dataset and differ on how they preserve the structure of the raw data. The following are commonly used for dimensionality reduction of mass cytometry single-cell data. Principal Component Analysis (**PCA**) [81] captures $40 - 50\%$ of the variance on the top three principal components. Since it assumes a linear distance between cells, it misrepresents cells existing on non-linear distance. On the other hand, t-distributed Stochastic Neighbor Embedding (**t-SNE**) [73, 82] provides an alternative. t-distribution focuses on preserving local neighborhoods and usually is successful at keeping non-linear relationships between markers better than linear approximation methods such as **PCA**. In contrast, **t-SNE** does not maintain the global structure of the data, causing some limitations in its interpretation.

2.5 From reductionist to holistic view of immunology

Current advances in technologies provide high-throughput measurements and analysis of multiple parameters simultaneously. With this in mind, researchers have started to ask questions pertaining not to a single cell type or a single protein but to many cell populations and serum proteins simultaneously aiming to understand collective behavior and patterns of the human immune components. This allowed tackling questions like inter-individual differences dictated by genetic diversity and environmental exposures in humans, which were otherwise unattainable using conventional methods [83]. In a single time point, perturbations are applied to the immune system, and simultaneous measurements of initial and resulting states are performed to quantify immune response [84]. Over time or in multiple time points, temporal aspects of the complexity of immune system behaviors can be captured. Moreover, humans are subject to several perturbations throughout their lifetimes and infections. This has led researchers to explore an individual's infection history.

These technologies enabled interrogating the immune system in the systems-level setting. A couple of studies have shown a variation of the immune cell composition in the blood of individuals, which measures the disparities between individuals' responses to the same immunological stimulation. Changes in the cell populations' phenotypes are compared against the baseline in these cell populations over time, within an individual, and or between individuals. Human immune cell frequencies and serum protein levels are initially found to be stable over a period of time for a given individual and result in highly variable between healthy individuals, at least in healthy adults [85–87].

The variation within individuals is due to heritable factors and an unexpected dominance of non-heritable factors influences from infections, vaccines, nutritional factors, and the microbiome, explaining most of the variation for 77% of all immune cell frequencies, functions, and serum protein measurements [79]. Similarly, Patin et al. have demonstrated that heritable factors influenced the parameters of innate cells more than adaptive cells. The environmental exposures analyzed included smoking, age, and latent infection with cytomegalovirus, and those explained more of the variance in adaptive than innate cells [88]. Adaptive changes in the immune system induced by environmental influences are explained by increased epigenetic variations with age [89].

These are few examples exemplifying the use of a holistic approach in understanding immune system phenotype and function. This paradigm shift focusing on human studies is believed to extend our understanding of the immune system that could be more easily translated to the clinics [16, 83].

Chapter 3

Development of the human immune system

Heterogeneous immune phenotypes in adults are widely accepted and are highly explained as an accumulation of non-genetic factors' influence. Little is known on the identity of these factors and when they exert this influence. As a start, one should note the phenotypical and functional differences between the fetal and adult immune systems. The former emerges from a relatively sterile environment, in utero, while being prepared to fight against enormous pathogens (bacteria, fungi, and virus) directly after birth [90, 91]. Given the differences in exposure to antigens and environmental stimuli for infants and adults, examining disease susceptibility both at the early life and later in life demands an understanding of the heritability of immune responses and the variability of the responses [92, 93]. Understanding the early life immune response could lead to identify, prevent, and treat early life infectious diseases.

3.1 Immune cell composition of blood throughout development in early life

Fetal hematopoiesis takes place in the yolk sac and moves to the liver and other tissues during gestation, where both lymphoid and myeloid cells arise [94]. The earliest tissue-resident macrophages appeared at gestational week four, and the most initial T cells identified at ten weeks of gestation [95]. The progenitors resided in the fetal liver at 6 to 22 weeks of gestation, and the production of the cell moved to the bone marrow until birth [94, 95]. Our knowledge of the ontogeny of the fetal immune system comes from studies in mice [96].

The study of immune system development in human's fetus is naturally

proven challenging to carry on, and more is known about development from birth to adulthood. With high dimensional single-cell technologies, researchers have studied counts and frequencies of white blood cells after birth. In many studies, some common patterns are observed and suggest a robust developmental process. For example, *i*) white blood cell count decreases from birth until adulthood where it reaches a stable number [97, 98], *ii*) the ratio between naïve and memory T-cell decreases continuously [97–100], and the CD4:CD8 ratio decreases from birth to adulthood [97–99, 101–103].

3.2 Passive immunity by maternal antibodies in human newborns

Antibodies exist in different isotypes such as IgA, IgD, IgE, IgG, and IgM. During pregnancy, maternal antibodies (IgG) are transferred across the placenta to the fetus [104, 105] by the neonatal Fc receptor (FcRn), and other possible unidentified receptors [106]. Several factors, such as IgG subclass [107, 108], antigen specificity[109, 110] and chronic maternal infections [111] have been proposed to influence the rate of IgG transfer. Antibody transfer is believed to occur after 28 weeks of gestation [104, 112]. Therefore, preterm birth is a determinant factor for the transfer of maternal antibodies and may limit the transport of IgG [113–115]. Such a lack of passive immunity has been suggested as a reason for the higher risk of infectious diseases in preterm children. The transferred antibodies provide passive immunity until newborn IgG starts to produce around 15 weeks after birth [116].

Many studies have focused on individual viruses [117–120] by mostly using the Enzyme-linked Immunosorbent Assay (**ELISA**) method, which is limited in throughput and coverage. With recent technological advances, we have started to ask questions like: what are the transferred maternal antiviral antibodies? How large is the repertoire of transferred maternal antiviral antibodies, their specificities, and duration after birth? Answering these questions would imply understanding infectious disease susceptibility, and vaccine scheduling in newborn children.

3.3 Environmental influences on the Immune system

Among factors shaping the immune system include heritable and non-heritable. This latter has been shown to influence the immune system in healthy adults largely [87, 121, 122]. In early life, non-genetic factors accepted to shape the human immune phenotype are identified and include mode of delivery, medication, diet, and infections [123, 124]. There is evidence showing that environmental exposures during the first days of life can predispose for nor-

mal immune development [125, 126].

For that, several studies have shown how these non-genetic factors influence the immune phenotype in newborn children. One of the factors that are well studied is the mode of delivery: vaginal delivery due to the increase in stress hormone levels during labor corresponds to higher levels of several cytokines [127–130]. At the same time, caesarian sections have been related to higher granulocyte and leukocyte in cord blood [127, 128, 131]. Other environmental factors that are associated with the immune phenotype include but not limited to antigen exposure [132, 133] , breastfeeding [134–137], viral infections [134], and maternal smoking [138].

Lastly, the effects of non-genetic factors on the immune system are displayed through the latent modulation of the gut microbiota [139, 140]. The bacteria exposures of the baby at birth initiate the microbiota, which develops and stabilizes years later in life [141]. Similarly, the immune system and the gut microbiota composition are both partially shaped by host genetics [142–144]. Recently, studies have shown a relationship between gut microbiota composition and environmental factors. The microbiota composition is influenced by some perinatal factors such as mode of delivery [145, 146], diet [145–148], and antibiotics [145–148]. There is increasing evidence showing the association not only between the microbiota and the immune phenotype but also the development of immune-related disease [149, 150].

Chapter 4

Aims

The overarching aim of this thesis is to understand the natural development of the human immune system with a focus on studying how environmental factors shape it. To do this, we used systems-level analyses to investigate the immunity from different cohorts consisting of very preterm children to old adults with various environmental exposures. From here, immune system behaviors expressed in immunological observables such as cell frequencies, immune phenotypes, and perturbed genes expressions give insights into how the immune system works.

The specific aims of each study are:

Paper I. To investigate whether differences in immune response could be explained by variation in immune cell compositions across individuals.

Paper II. Assess antiviral antibodies in preterm and term mother-child pairs, their specificities, and duration after birth.

Paper III. Apply Pareto optimality theory to longitudinal blood samples from newborn children profiled by mass cytometry to understand immune phenotypic trade-offs and their mechanisms describing how the immune system adapts to environmental influences early in life.

Paper IV. To study the transcriptional variation of human immune responses to bacterial and viral challenges in Primary Immunodeficiency Disorder (**PID**) patients.

Chapter 5

Methods

5.1 Study designs

The data contained in this thesis are collected from different cohorts. Respective ethical committees have approved the studies protocol and written informed consent has been obtained from all participants except for the used data from a public repository.

5.1.1 Paper I

Paper I is based on data of 1575 healthy individuals across multiple cohorts: sampled 1) at the Stanford University clinical trials unit, 2) in the United Kingdom [151], and 3) in Belgium [85]. They consist of men and women between 8 and 89 years of age. Among the participants were twin subjects. All three cohorts measured the frequencies of various immune cell populations in samples of peripheral blood mononuclear cells. Moreover, a diverse set of functional responses (168 signaling responses to cytokine stimulations and vaccination to five influenza strains) was measured in the Stanford cohort.

5.1.2 Paper II and III

Paper II & III are based on the TELLUS cohort of human newborns and their respective parents' blood samples collected longitudinally at the Karolinska University Hospital at delivery (from the umbilical cord), and during week 1, and 40 (from peripheral blood). The cohort consists of very preterm children (< 30 weeks of gestational at birth) and term children ($>= 37$ weeks of gestational age at birth).

In Paper II, we assessed antibodies against 93,904 epitopes from 206 viruses in 32 preterm and 46 term mother-child pairs. We compared detected antibodies in both children groups as well as within mother-child dyads. Also,

we performed multiplex **ELISA** assays to quantify IgG specific to known immunodominant epitopes from five common viruses.

In Paper III, we have analyzed blood samples collected longitudinally from 15 newborns of TELLUS cohort. The cells were analyzed using single-cell mass cytometry resulting in the antibody-detected expression of 38 proteins in thousands of cells. 22 and 14 markers were selected to precisely focus the analysis of B and Cluster of Differentiation (**CD**)4+ T cells, respectively.

5.1.3 Paper IV

Paper IV is based on the generated bacterial and viral-induced immune transcriptional profiles from **PID** cohort. It is an ongoing study and consists of around 20 patients with different immunodeficiency disorders undergoing treatment at the Karolinska University Hospital. We compared the results of **PID** patients with healthy individuals.

5.2 Experimental methods

In this section, I'm introducing the working principles of the main experimental methods used to generate the data presented in this thesis. We have employed methods capturing cellular components behaviors at different length scales.

5.2.1 Mass cytometry

Mass Cytometry or Cytometry by Time Of Flight (CyTOF) is a mass spectrometry-based technique for high-throughput measurements of protein abundances with single-cell level resolution [60, 62, 152]. It is a variation of flow cytometry in which antibodies are labeled with heavy metal isotopes rather than fluorochromes.

Cells in suspension are stained with antibodies bound with a metal isotope and are nebulized one by one (Figure 5.1). The spray is evaporated, and the cells are ionized in argon plasma. The ionized isotopes are passed through an electric field that accelerates the charged ions and allow filtering away naturally abundant ions with a low mass that would be otherwise noise to the signal. The purified isotopes are detected, and a time-of-flight spectrometry quantifies the ion abundance. The ion abundance is directly related to the expression levels of the proteins that are bound to the antibodies. Also, this allows for the combination of up to 40 antibodies without between channels. Mass cytometry provides a highly multiplexed method for profiling single cells. In immunology, it is commonly used for cellular phenotypic profiling

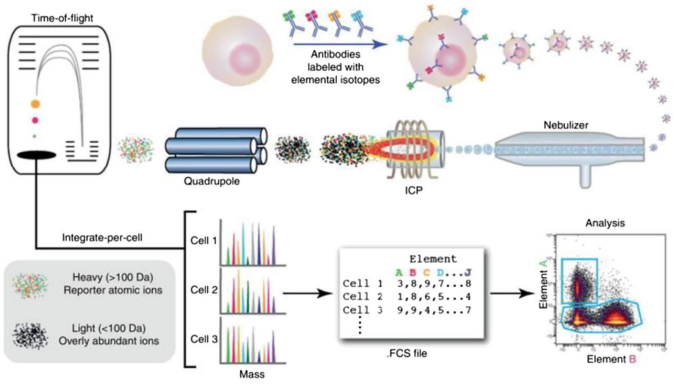

Figure 5.1: Schematic workflow of single-cell analysis in a mass cytometry. Figure reproduced from [152] with permissions from Elsevier (license number: 4782520495083)

and functional analysis. However, mass cytometry as an antibody-based method does not support comprehensive profiling of the cells, but instead, it is limited to a pre-defined set of proteins for which probes such as monoclonal antibodies exist [63].

5.2.2 mRNA sequencing

mRNA sequencing is a Deoxyribonucleic Acid (**DNA**) sequencing-based technology for the high-throughput analysis of the abundance of all mRNA transcripts existing inside cells. It enables identifying expressed genes and quantifying their abundances, which are indicative of the phenotype and functions of cells.

To get mRNA content in a population of cells, cell membranes are broken in a process called cell lysis, and the cellular contents are released and stabilized. From the captured mRNA transcripts, gene expression to these transcripts can be inferred. Given the current sequencing technology, the extracted mRNA is converted into complementary Deoxyribonucleic Acid (**cDNA**) and is linearly amplified through polymerase chain reaction to ensure a sufficient amount of material for a robust sequencing. The resulting cDNA materials are subjected to DNA sequencing. Subsequently, reads are aligned to the human reference transcriptome, and the number of reads from each gene determines the abundances of mRNA molecules.

5.2.3 VirScan

VirScan is a phage display method that enables human virome-wide of immune responses at the epitope level [69]. VirScan assay is based on the immunoprecipitation of antibody-targeted bacteriophages from a library of

around 94,000 56-amino acid peptides representing the entire proteomes of 1000 viral strains from all 206 viruses known to infect human cells.

Initially, the phage library and plasma containing antibodies are incubated together, and antibodies bind to the peptides expressed on the surface of the bacteriophages. Then, antibody-phage complexes are precipitated to magnetic beads. Furthermore, the antibody-bead complexes are washed in distilled water, and heating at a high temperature lyses the phages. The region coding for the clonally expressed peptide is amplified through polymerase chain reaction run and subjected to massive parallel DNA sequencing. To identify the targeted peptides, the reads are mapped to the reference viral library, and the read counts for each peptide.

5.3 Computational methods

5.3.1 Multivariate features selection

Most of the technological development in recent year aims at measuring multiple parameters in a given sample. These could be protein, gene expressions or immune cell frequencies.
Suppose that we are interested in understanding how an individual's immune system would respond to a stimulus. Given a cohort of N size, we will collect functional or transcriptional responses, $\{y_1, \ldots, y_N\}$, to stimuli. At the same time, we will measure immune cell population frequencies or gene expressions in all individuals in the cohort, X. Then, from here, we need to describe the inevitable interindividual variation in the response, y. We can assume that only a fraction of the variables or predictors in X explain the variation in the response, y (renormalization theory).

To this end, Partial Least Squares (**PLS**) [153] regression is used to relate the immune cell compositions or gene expressions of individuals with functional responses to stimuli. Linear model assumes that the response and predictors are related through the functional form

$$\hat{y} = X\beta \tag{5.1}$$

Where, y is a vector, $\hat{y}$ is the predicted response, X is a matrix of predictors, and β is the coefficient vector for functional response. PLS finds a set of linear combinations of cells populations that are strongly correlated with the functional response. PLS carry out feature selection as a group by maximizing the covariance between X and y.

$$\max_{\beta} cov(X, y) \tag{5.2}$$

Indeed, a set of collective variables is chosen so that individuals with similar coordinates in the space drawn by these collective variables respond similarly to stimuli.

PLS is a "coarse graining" method that allows investigating the behavior of immune system as we zoom out and examine the system at different length scales.

5.3.2 Archetype analysis

As discussed above, the immune system is inherently regulated. It mounts an immune response against infection and maintains self-tolerance to host tissues through checkpoints [154, 155]. This is observed on the well-regulated structure of the immune phenotype space from single-cell data that results in a specific geometrical shape. In addition, single cells can exist in discrete groups and/or in a continuum when cells are under cell differentiation process.

Archetype analysis is an unsupervised approach based on an evolutionary tradeoffs theory to learn the geometrical shape of proteins or gene expression space [156]. The theory suggests that cells that need to perform multiple tasks, their performance levels cannot be concurrently optimized for all tasks, and hence the competition between them affects phenotype selection. As a result, cells grow and differentiate themselves with respect to the environment through a precise trade-off. To this end, the concept of Pareto optimality has been used [156–159]. As a result, cells are arranged in convex geometrical objects or polytopes like line, triangle, tetrahedron, etc. The vertices, called archetypes, of these structures are the extreme protein expression profiles optimal for each task.

Identifying archetypes. How do we find these geometrical structures? We identify first the archetypes of a polytope best-fitting convex hull of the data in protein expression space by using Principal Convex Hull Analysis (**PCHA**) algorithm [160]. The goal of the algorithm is to find the best-fit polytope whose vertices, archetypes, are on the convex hull of the data. To do this, the algorithm needs to meet two conditions.

First, the vertices are expressed as a weighted average (convex combination) of the data points. This is to make sure that the archetypes are the representatives of the protein expressions space. Suppose that we are given N cells with M measured protein surface markers $\{x_1, ..., x_N\}$, where $x_i \in \mathbb{R}^M$.

Each archetype can be expressed as a convex combination of all cells,

$$z_j = \sum_{i=1}^{N} c_{ij} x_i \tag{5.3}$$

where c_{ij} are the convex coefficient, $0 \leq c_{ij} \leq 1$, $\sum_{i=1}^{N} c_{ij} = 1$ and $j = 1, ..., k$ for k archetypes. For a set of archetypes, their representation could be written as follows

$$Z = XC \tag{5.4}$$

where $X \in \mathbb{R}^{M \times N}$ is the data point matrix, $C \in \mathbb{R}^{N \times k}$ is the coefficient matrix and $Z \in \mathbb{R}^{M \times k}$ is the matrix of archetypes.

Second, the archetypes need to be on or near the convex hull of the data. Approximating the data points by a weighted average of the archetypes does this. Given k archetypes $\{z_1, ..., z_k\}$, where $k_j \in \mathbb{R}^M$, each cell is expressed as

$$x_i = \sum_{j=1}^{k} s_{ji} z_j \tag{5.5}$$

where s_{ij} are the convex coefficient, $0 \leq s_{ji} \leq 1$, and $\sum_{i=1}^{N} s_{ji} = 1$. For all cells, $\hat{X} = ZS = XCS$, where $S \in \mathbb{R}^{k \times N}$ is the matrix of s_{ji} coefficients.

The algorithm is constrained to optimize both criteria. Given that it aims to approximate the data points from the archetypes, the optimization problem consists of minimizing the difference between original and is solved using a projected gradient descent on a constrained space:

$$\arg\min_{c,s} \|X - XCS\|^2 \tag{5.6}$$

From here, the number of archetypes that describes the data is determined. To do this, PCHA computed the explained variance [159, 161, 162] for each number of archetypes $k = 2, ..., k_{max}$ as follows

$$EV = \frac{1}{N} \sum_{N=1}^{N} (1 - \frac{|p_n - s_n|}{|p_n|}) \tag{5.7}$$

Where p_n is the n^{th} data point from N points and s_n is the closest point to p_n in the found polytope. Therefore, an optimal number of archetype k^* is identified when increasing number of archetypes does not significantly increase EV. As a result, optimized set of archetypes of a polytope that best encompasses the data is obtained using the Sisal [163] algorithm, which is less sensitive to outliers.

Characterization of a continuum of cell states in the presence of multi-tasks inside a polytope. After finding a polytope that best fit our data (convex regions), we need to explore the concavity region in the middle of the polytope [164].
Multiple tasks are characterized by performance functions of cells working together to provide a function within a differentiated cell type [164]. The overall performance was defined as a sum of contributions of individuals cells, $j = 1, \ldots, N$ to each of tasks, $i = 1, \ldots, k$. The vector M_j is the protein expression of cell j. The performance of cell j in task i, $P_i(M_j)$, is partially defined by its protein expression profile. In addition, with longitudinal data, a continuum of cell state can arise when tasks have spatiotemporal performance gradients. The movement of cells in the protein expression space over time, especially in newborn children, confirms this.
To incorporate this gradient effect in our definition of performance of cell j, a mutlitplicative factor $\phi_i(t_j)$ is added for each task i.

Therefore, the total performance is an increasing function of the sum of all cells performance in the k tasks and can be written as:

$$F = f(S_1, ..., S_k) \tag{5.8}$$

$$S_i = \sum_{j=1}^{N} \phi_i(t_j) P_i(M_j) \tag{5.9}$$

In order to represent the tasks dependence on protein expression gradients, we used linear performance gradients for tasks at archetype 1 and 2, no gradient at archetype 3, for B-cells for example. this movement of cells with respect to archetypes canbe emulated as $\phi_1(t) = x$, $\phi_2(t) = 0.5 - x$, and $\phi_3(t) = 1$. Where x is the spatiotemporal position of cells. Moreover, a simple linear performance function that decays linearly with Euclidean distance d_ij from the archetype is used $P_i = B - d_{ij}$, where B is a constant. The derivative of F with respect to M leads to the solution

$$0 = \frac{dF}{dM} = \sum_{i=1}^{k} \phi_i(t_j) \frac{\partial f}{\partial S_i} \frac{\partial S_i}{\partial M_j} = \sum_{i=1}^{k} \phi_i(t_j) \frac{\partial f}{\partial S_i} \frac{\partial P_i}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial M_j} \tag{5.10}$$

The solution would show that cells lie along a continuum range of expression in protein expression space inside polytope, whose vertices are the maxima of the performance M^*. The performance gradients $\phi(t)$ describe cell specialization pattern to tasks that suggests a pseudo-time trajectories in the developmental path from single-cells data.

Archetype analysis enables studying changes to the phenotypic space induced by perturbations in the system. I used this method to understand human newborns lymphocyte development.

5.3.3 Serological data analysis

Virscan

Experimentally, the input consists of three types: the phage-only or 'input, the beads-alone or background, and the sample. The latter is treated in duplicates for validation purpose. The reads from all inputs are aligned separately to the reference viral genome and counted reads for each epitope. The output reads counts of the 'input' serves as a null model and a zero-inflated generalized Poisson distribution describes it.

For each sample duplicate, the output reads count is fit a Poisson model and solve for the coefficients of the parameters as function the 'input' read count. Based on this comparison, p-values are calculated for significance of each epitope's enrichment. An epitope is called a hit, if the -log(p-value) is greater than a threshold, 2.3. Hits found in the background of the assay were masked in the hits found in sample. Lastly, number of unique hits per viral species makes up a virus score, so-called VirScore.

Statistical model of antibody kinetics

We were interested in determining antibody levels in children and their mothers. Total IgG was measured with an ELISA assay in all longitudinal serum samples. To capture the kinetics of passively acquired maternal antibodies against certain viruses, we used a stochastic model [165]. For each individual, log-transformed levels of antibody, x, at a particular age, a, are described by a normal distribution with age-dependent mean $m(a)$ and standard deviation, s.

$$x \sim N(m(a), s^2) \tag{5.11}$$

The log-transformed levels of maternal antibodies decay linearly with age and with a decay rate, d, from $ln(m_0 + b)$ for low ages, and approaches a detectable level of antibody, b, for higher ages. The mean log-transformed level of maternal antiviral antibodies varies with age

$$m(a) = \ln(m_0 e^{-da} + b) \tag{5.12}$$

We used the maximum likelihood estimation to regress the parameters m, b, d and s. We also computed the half-time, $\ln(2)/d$, and the mean-life, $1/d$, of antibody.

Chapter 6

Results and discussion

In this chapter, the main findings contained in the appended papers are presented and discussed. The first section is based on the attached paper I, and results from understanding inter-individual variation in immune cell composition and their responses to cytokine stimulations. In the second section, which is based on paper II, a global analysis of maternal antibodies to all viruses in newborns is presented. In the third section, an analysis of phenotypic trade-offs and their mechanisms using Pareto optimality (Paper III) is displayed. Finally, a transcriptional method aiming at understanding immune defects in patients with primary immunodeficiencies (Paper IV) is discussed.

6.1 Paper I: Continuous immunotypes describe human immune variation and predict diverse responses

6.1.1 Introduction

The relative immune cells frequencies in blood in the human population can vary extensively between individuals, and the differences are poorly understood. Surprisingly, these variations do not seem to have a substantial impact on immune fitness.

We hypothesized that for the immune system to exhibit such flexibility, individual cells have to be able to tune their response to stimulation in relation to the levels of other immune cells. This is similar to quorum sensing in bacteria, where collective behaviors like biofilm formation and sporulation emerge from interactions and signals between bacteria.

6.1.2 Results

In this paper, the aim was to investigate whether differences in immune response could be explained by variation in immune cell compositions across individuals.
To do this, we analyzed the frequencies of various immune cell populations and their responses in samples of peripheral blood mononuclear cells from 1575 healthy individuals across multiple cohorts.

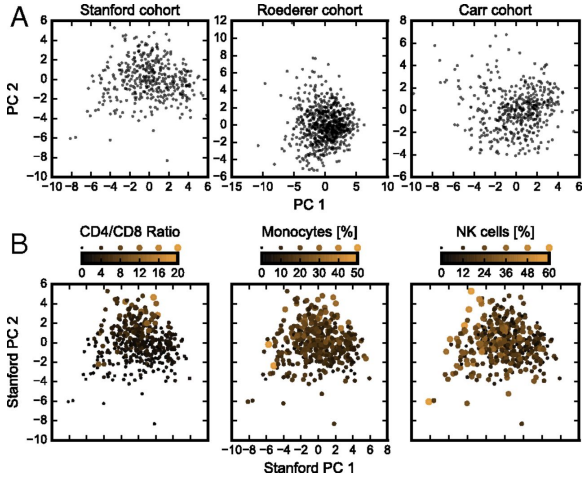

Figure 6.1: Principal component analysis of immune cell composition. a) Projection of high dimensional immune cell composition onto the two principal components for Stanford, Reorderer, and Carr cohorts. b) Displaying individuals with extreme values of specific immune cell populations on the top two PCs of global immune cell composition data from Stanford cohort.

The main finding of the study is that the inter-individual differences in immune cell composition are continuously distributed rather than clustering in discrete groups with similar immune cell composition (Figure 6.1a). Next, we show that individuals with extreme outlier values of specific cell type frequencies are not outliers with respect to their collective immune cell composition (Figure 6.1b).

Similarly, individuals with extreme values (considered outliers) in individual immune cell populations (CD4/CD8 ratio, Monocytes, and NK cells) are not outliers using a collective description via PCs (Figure 6.2). Thus, to define individuals with outlier phenotypes, collective states of their immune cell composition can be helpful.

To this end, a conceptual framework for the analysis of measured frequencies of various immune cell populations in healthy humans and their responses to diverse cytokine stimulation is devised. We sought to know a small set of linear combinations of specific immune cell frequencies that could predict individual immune responses to stimuli. To this end, PLS method identified

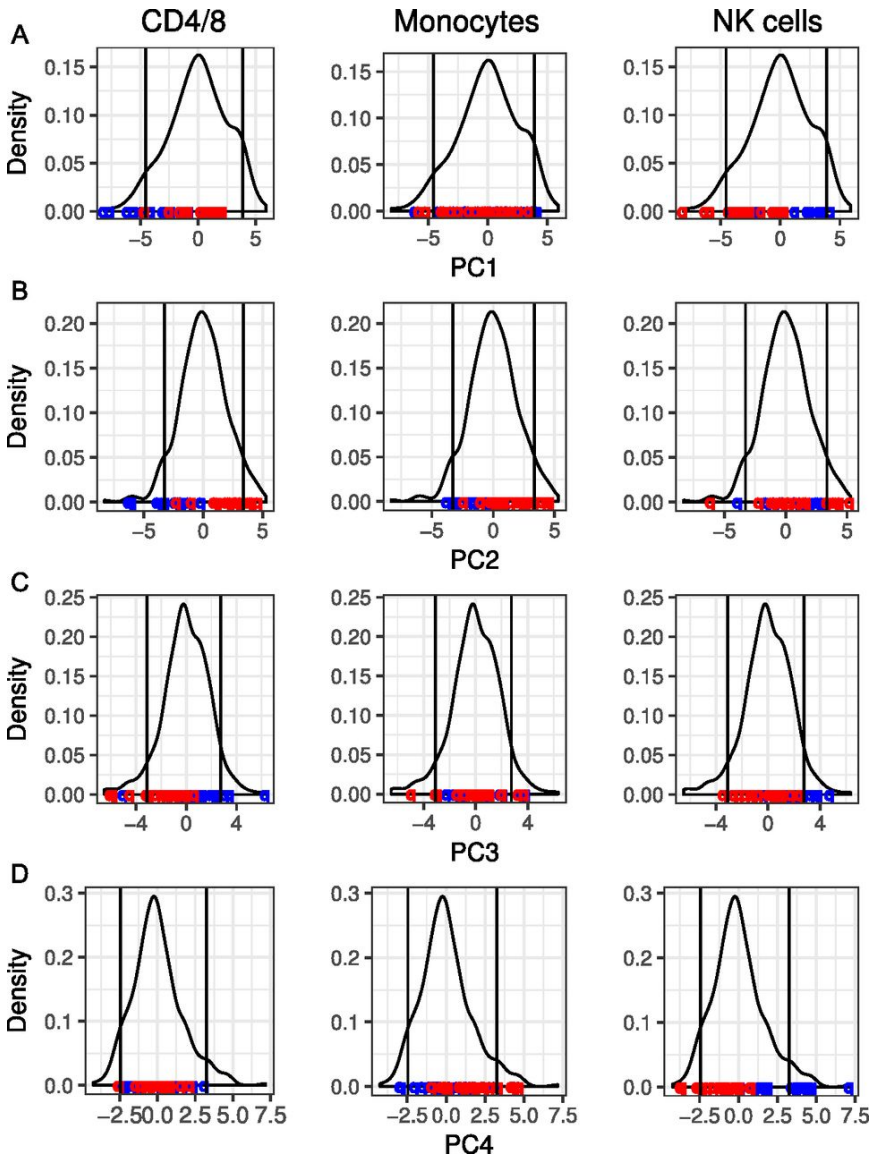

Figure 6.2: Overlaying the top (red) and bottom (blue) 5% of the CD4/CD8 ratio, monocytes, and NK cells onto the distributions of PC scores across the Stanford cohort.

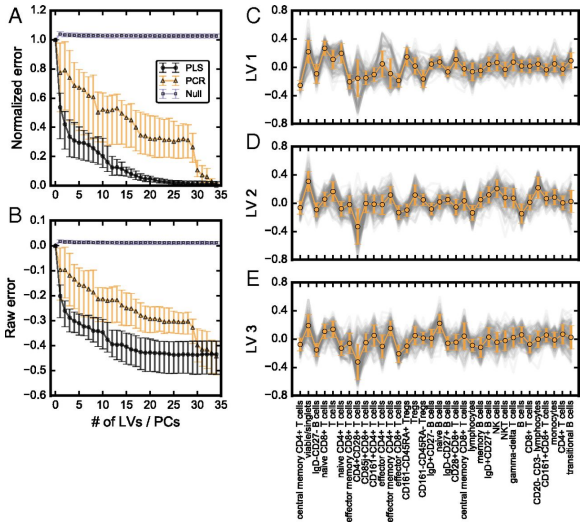

Figure 6.3: Linear regression model analysis. Plotting model error as function of number of variables (e.g latent variables and PCs) represent learning curves for PLS and Principal Component regression (**PCR**) models: with a) normalized, and b) raw error. c-d) Displaying the mean and interval of signatures of the top three latent variables across all responses.

a set of three variables of individual immune cell frequencies that robustly predict diverse functional responses in healthy individuals (Figure 6.3). This set of aggregate immunological variables defines an individual's immunotype.

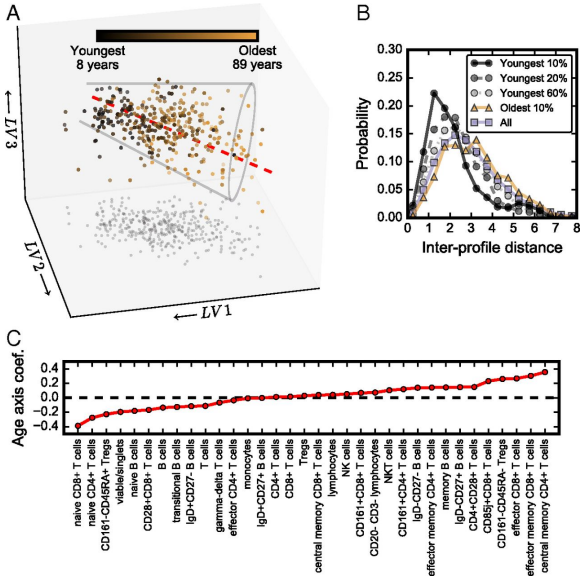

Figure 6.4: Projection of individuals in the immunospace defined by the top three latent variables. The color code reprents gradient of age, and the red line is an immunological age axis. b) Distribution of interindividual Euclidean distances in the immunospace spaned by the first 3 LVs for different discrete age groups. c)Immunological age axis defined by regression coefficient of immune cell populations.

To better understand human immune variation, we analyzed the distribution of individuals' immunotypes displayed in the space defined by the top three PLS components, see figure 6.4. We find that even though inter-individual differences in specific cell population frequencies can be significant, unrelated individuals of younger age have more homogeneous immunotypes than older individuals. Increasing heterogeneity of immunotypes in old individuals shows the diversification of immune cell composition with age. It indicates the effects of diverse cumulative environmental influences on an individual's immune system over time.

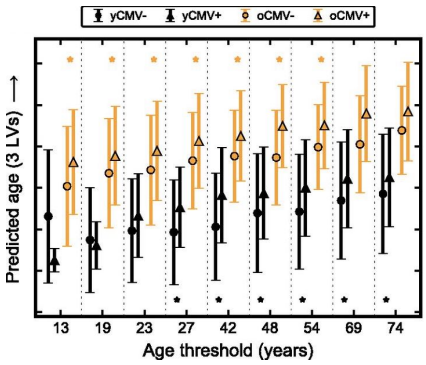

Figure 6.5: Distributions of positions of individuals to the line of age axis are compared between HCMV seronegative and seropositive individuals for young and old.

Finally, we sought to understand how Human Cytomegalovirus (**CMV**) serostatus shapes the human immune system. To do this, we projected individuals' immune cell composition onto the age axis, a line that best correlates with age in space defined by top three PLS components. By analyzing the distribution of individual's cell populations projected onto the line (Figure 6.4a), we found that the human **CMV** seropositivity was represented by a change in immunotype. Across age groups, the distribution of cytomegalovirus seropositive individuals is shifted upward along the age axis (Figure 6.5), and displayed immunotype similar of older individuals.

6.2 Paper II: The repertoire of maternal anti-viral antibodies in human newborns

6.2.1 Introduction

Maternal antiviral Antibodies (**mavAb**) provide neonates with primary protection against viral infections early in life. However, little is known about the global repertoire of transferred **mavAb** IgG, its variation in children, the epitopes targeted, and its duration after birth remains poorly defined. To fill this gap, we have collected samples from very preterm and term newborn

children at birth and during weeks 1, 4, and 12 of life.

In this study, we assess maternal antibodies against 93,904 epitopes from 206 viruses (except zika and corona COVID-19 viruses) known to infect human cells in 78 mother-child dyads.

6.2.2 Results

In this paper, we assessed maternal antiviral antibodies in preterm and term mother-child pairs using VirScan [69] method. The method output is defined as a number of peptide hits that do not share epitopes, called VirScore.

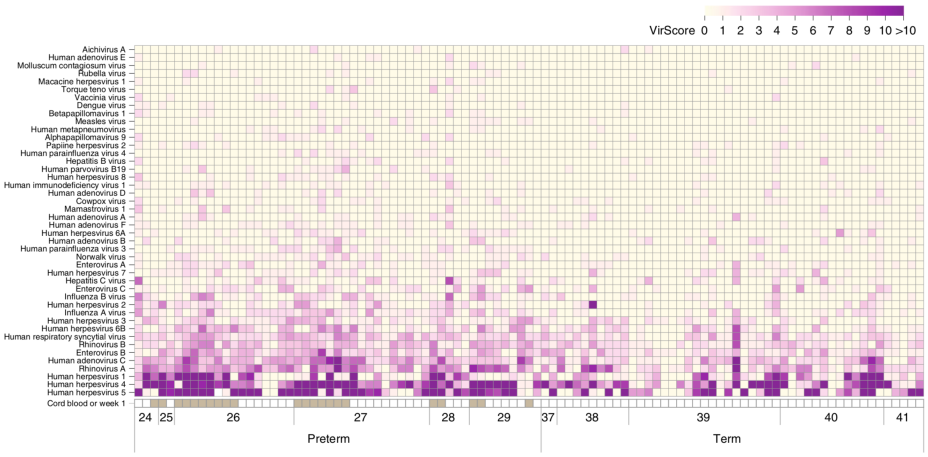

Figure 6.6: The repertoire of maternal antiviral antibodies. Top 44 viruses most commonly targeted by maternal antibodies in preterm (<30 gestational week) and term (>37 gestational week) children. The color code box at the bottom indicates time point, either at birth or day 2-3, when the first available sample for each child was collected.

We investigated the global repertoires of maternal antibodies to viruses in samples collected at birth or first sample taken in the first week of life in term and preterm children. Surprisingly, we found that term and preterm children received similar repertoire of maternal IgG. This suggests that the transfer of maternal antibodies happen before 24 weeks of gestation. The most commonly target antibodies in newborn children is similar to the parents, and they are **CMV**, Herpes Simplex Virus (**HSV**), Epstein-Barr virus (**EBV**), rhinovirus A, and adenovirus C, see Figure 6.6.

To gain a deeper insight into the determinant of protective immunity, we investigated antibody specificity on an epitope level. First, we found a linear relationship between the number of epitopes targeted by maternal antibodies and the genome size (Figure 6.7a), suggesting that virus with large

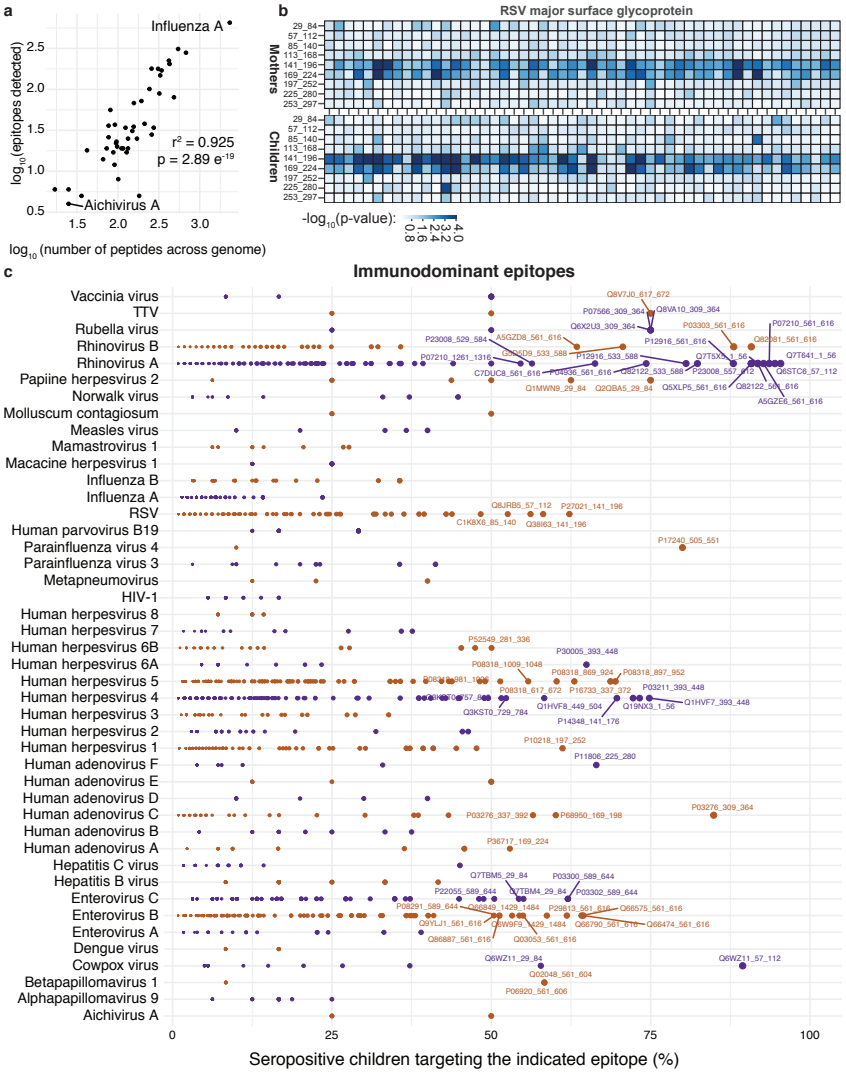

Figure 6.7: a) Viral genome size linearly correlates with the number of epitopes targeted by maternal antibodies. **b)** IgG antibodies in mother (top) and children (bottom) more often targets similar peptides of RS-virus major surface glycoprotein G. **c)** All epitopes targeted in seropositive newborn children show immunodominant epitopes ($> 50\%$ of seropositive)

genome size have many epitopes that are recognized by antibodies. Second, we wanted to know if there exists epitopes that are preferably targeted during an immune response in the repertoire of maternal IgG, called immunodominant epitopes. For example, in both mother and children, antibodies targeted more often an epitope located in 141-224 amino acid region of Respiratory Syncytial (**RS**)-virus major surface glycoprotein G (Figure 6.7b). Lastly, we displayed the epitopes as a percentage of seropositive children in all viruses commonly targeted by maternal antibodies in newborn children (Figure 6.7c). The number immunodominant epitopes (epitopes targeted by $> 50\%$ of seropositive children) varied from virus to virus. These epitopes have implication in developing vaccine.

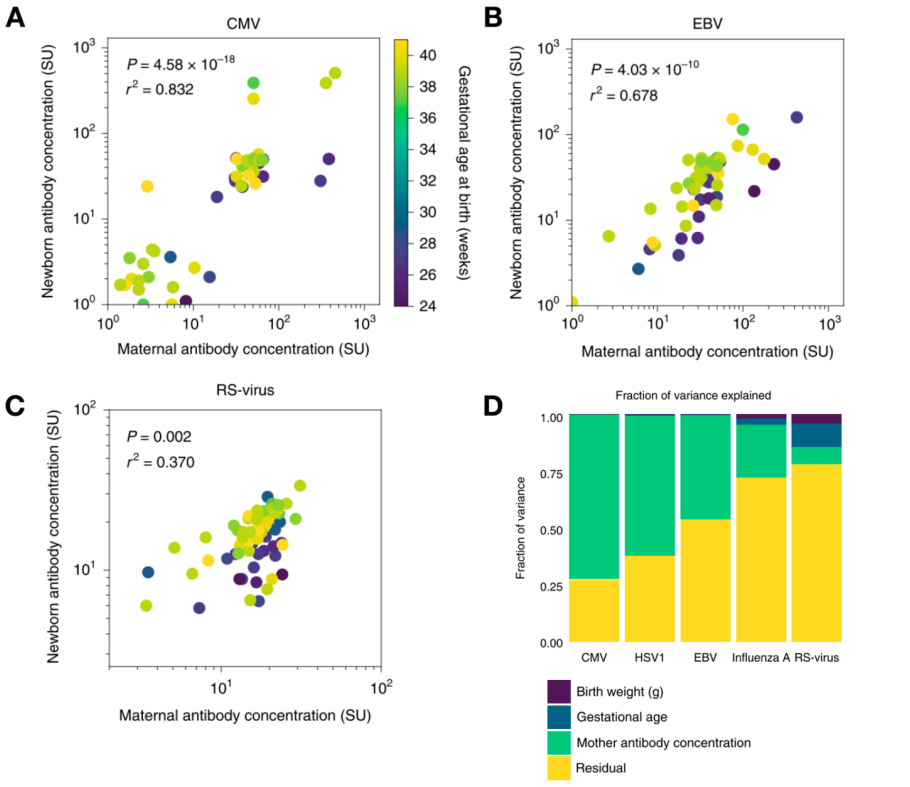

Figure 6.8: Factors determining maternal antibody concentrations. Pearson's correlation between antibody concentrations profiled by ELISA in mothers and their children **a)** CMV, **b)** EBV, and **c)** RS-virus. **d)** Variance in IgG newborn IgG concentration is explained by birth weight, gestational age, maternal antibody concentration, as well as the unknown or residuals.

Although a correlation of 0.5 is found between the concentration of antibodies and VirScan VirScores output, we quantified antibody levels in CMV, EBV, and RS-virus using enzyme-linked immunosorbent (ELISA) assay. Term children have antibodies with higher IgG concentrations than preterm children

(Figure 6.7a-c). Maternal IgG concentration was the most determinant factor of IgG concentration in newborns for CMV, and EBV; while gestational age influenced the IgG concentration for RS-virus (Figure 6.7d). The differences in determinant factors suggest that placental antibody transfer follows different rules.

	CMV		EBV		RSV		HSV-1		IA	
	Pre	Term	Pre	Term	Pre	Term	Pre	Term	Pre	Term
$t_{1/2}$	47.8	48	54.9	84.5	66.3	81.6	51.8	74.3	65.4	78
t	67.4	69.3	79.1	121.9	95.3	117.8	74.7	107.2	94.5	112.6

Table 6.1: Antibody life span (halftime, $t_{1/2}(days)$, and lifetime, $t(days)$)

We studied the duration of conferred protection by maternal antibodies in children. Maternal antibodies last 2-3.5 months (Table 6.1) on average and premature infants loose their antibodies much faster than full term children due to their lower concentrations transferred at the time of birth. Given that it has been reported that maternal antibodies interfere with vaccines, this finding could be helpful in vaccine scheduling in newborn children.

6.3 Paper III: Analyses of phenotypic trade-offs to understand human lymphocyte development

6.3.1 Introduction

In multicellular organisms like the immune system, individual cells face a dilemma. No cell can be optimally suited for all possible tasks, and therefore cells specialize to perform specific tasks. For example, in the human immune system, cytotoxic lymphocyte kills virus-infected cells, and B-lymphocytes produce antibodies, etc. That way, a range of phenotypic variants found in a system can be explained regarding of trade-offs between specialist phenotypes and more generalist phenotypes. There is a need for describing such phenotypic trade-offs from high dimensional single-cell data.

We have applied the Pareto optimality theory (162) to longitudinal blood samples from newborn children analyzed by mass cytometry to describe how the immune system adapts to environmental influences early in life.

6.3.2 Results

In this paper, the goal was to understand immune phenotypic trade-offs and their mechanisms describing how the immune system adapts to environmental influences early in life. To do this, we applied Pareto optimality theory

to learn geometrical shapes of B and CD4+ T cells protein expression space.

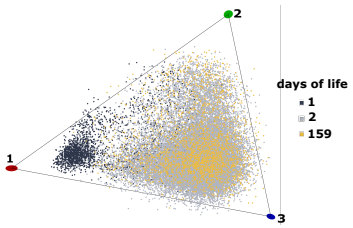

Figure 6.9: B cell protein expressions plotted in the top two principal components space are bounded by a triangle.

	arch.1	arch. 2	arch. 3
markers		CD9	CD39
		CD38	CD22
		CD24	CD44
		CD29	IgD
		CD20	CD99
		CD7	CD7
		CD49d	HLA-DR
		CD27	CD19
		CD45RA	CD45RA
		CD5	CD29
		HLA-ABC	HLA-ABC
			CD20
			CD49d
cell subsets		Naïve	IgD+ memory
time point	Day 1	Day 159	

Table 6.2: Enriched markers, cell subsets, and time point at each archetype in triangle B cells

We wanted to understand whether cells formed well-separated clusters of cells with similar protein expression profiles. The single-cell data showed an elongated cloud of phenotypes with expression gradients. Cell phenotypes of the first available sample in each child were different from phenotypes of samples from later time points. This indicates that there are changing phenotypes of individual cell populations over time. Then, we sought to understand whether cells fall in a low-dimensional polytope. Towards this end, we fitted the data to polyhedral with 3-7 vertices and assessed their statistical significance. We found that neonates B and CD4+ T cells protein expression spaces are best represented by a triangle (figure 6.9) and a non-linear 4D simplex (figure 6.10), respectively.

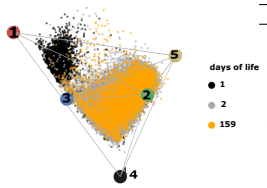

Figure 6.10: CD4+ T cell protein expressions projected in the low dimensional space spanned by top three PCs are best represented by a non-linear 4D simplex.

	arch.1	arch. 2	arch. 3	arch. 4	arch. 5
markers			CD29	CD38	CD45RA
			CD161	CD5	CD38
			CD44	CD45	CD44
			HLA-DR	CD3e	CD4
			CD5		CD29
			CD4		CD29
			CD45		CD5
					CD127
					CD27
					CD9
					CD7
					CD45
					CD3e
cell subsets	Naïve	Naïve	central	effector	Naïve
time point	Day 1	Day 2	day1/159	day 2	day 1/159

Table 6.3: Enriched markers, cell subsets, and time point at each archetype in pentahedron CD4+ T cells

The vertices, called archetypes, correspond to extreme protein expression profiles optimal for specific tasks. To infer tasks or functions, we tested the enrichment of markers as function of Euclidean distance from each archetype, and retained markers that are maximally expressed in the cells closest to each

archetype. Archetypes correspond to major cell subset with a set of markers, except archetype with closest cells that display homogeneous protein expressions (Table 6.2&6.3).

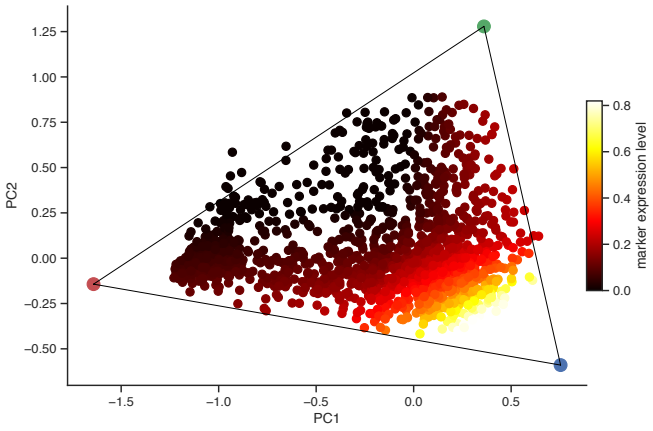

Figure 6.11: A continuum pattern of adaptive protein expressions describe specialization of cells to tasks in a triangle B cells.

Next, we sought to know the adaptive protein-expression continuum and how it corresponds to the specialization of cells to tasks overtime. We found that B cells lie along a continuum range of expression in protein expression space inside the triangle. Specifically, 1D continuum patterns of adaptive protein-expression describe specialization of cells to tasks and infer pseudotime trajectories in the developmental path of B cells. This suggests that there are many temporal paths with different ordering to tasks, rather than one temporal path to differentiation.

6.4 Paper IV: A transcriptional method to understand functional defects in patients with primary immunodeficiencies

6.4.1 Introduction

PID is a heterogeneous group of disorders that may lead to faulty innate and adaptive immune system. In many PID patients, the immune response is insufficient to eliminate specific pathogens, which leads to an increased susceptibility to infections and autoimmunity in this group of patients. A challenge remains in the diagnosis and treatment of PIDs due to variable symptoms dependent on the underlying immune system defects.

We developed a transcriptional method based on whole blood stimulated with Bacillus Calmette-Guérin (**BCG**),Lipopolysaccharidea (**LPS**), and Influenza A virus (**IAV**) to better understand functional defects in 11 PID patients. mRNA sequencing technique analyzes the induced transcriptional programs.

6.4.2 Results

In this work, we aimed to study transcriptional variation of human immune responses to bacterial and viral challenges in primary immunodeficiency disorder (PID) patients. We particularly wanted to compare transcriptional

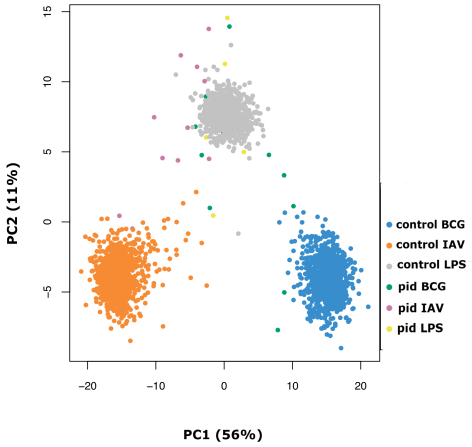

Figure 6.12: Transcriptional responses to microbial antigens are different between PID patients and similar in healthy controls. We used PCA to project high dimensional immune gene expression profiles defined as a ratio between stimulated and non-stimulated states.

responses to microbial pathogens between healthy individuals and PID patients. We asked whether there is an overlap of transcriptional responses in PID patients and healthy controls. Immune stimulations induced transcriptional responses that clustered in discrete groups of healthy individuals with a similar response to a specific stimulant antigen and formed unclear clusters for PID patients irrespective of the antigen. We conclude that transcriptional responses to specific microbial antigens are different in PID patients and similar in healthy individuals.

Given that the transcriptional responses to specific antigen are sparsely distributed among PID patients, we asked whether there is a collective set of genes that could predict the variation of responses among patients. Towards this end, Partial Least Square Discriminant Analysis (**PLS-DA**) defined a set of linear combinations of the genes known as PLS components, termed

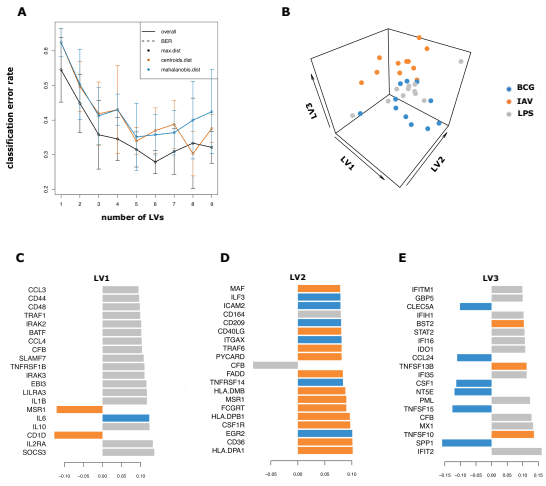

Figure 6.13: Three key combinations of gene expression profiles predict variation of transcriptional responses to microbial challenges in PIDs patients. A) Classification error rate suggests an optimal number of predictive components to be 3. B) PIDs patients plotted in the low dimensional space defined by the first three LVs. c-e) Top 20 genes contributing the most to the first top 3 LVs, color coded by a stimulation type with the maximal expression levels.

Latent Variable (**LV**). By projecting individual PID patients in the low-dimensional LVs, patients with similar response to specific immune stimulation form clear clusters.

Finally, we compared each PID patient with healthy controls to identify gene variants associated with transcriptional responses to prototypical stimulations. Given the heterogeneity of identified gene variants across patients, we selected top 40 genes contributing to each of the top 3 LVs for visualization purpose (Figure 6.14). This allows assessing and understanding functional defects in PID patients.

6.5 Discussions

In this thesis, a systems approach is used to characterize human immune system behaviors as probed in different experiments with a focus on elucidating environmental influences. How closer are we in providing a conclusive description of organized behaviors observed in the human immune system?

To perform its protective function and preventing immunopathology, the immune system involves coordinated action of many components of the system. The challenge here is to select a simple representation for the elementary in-

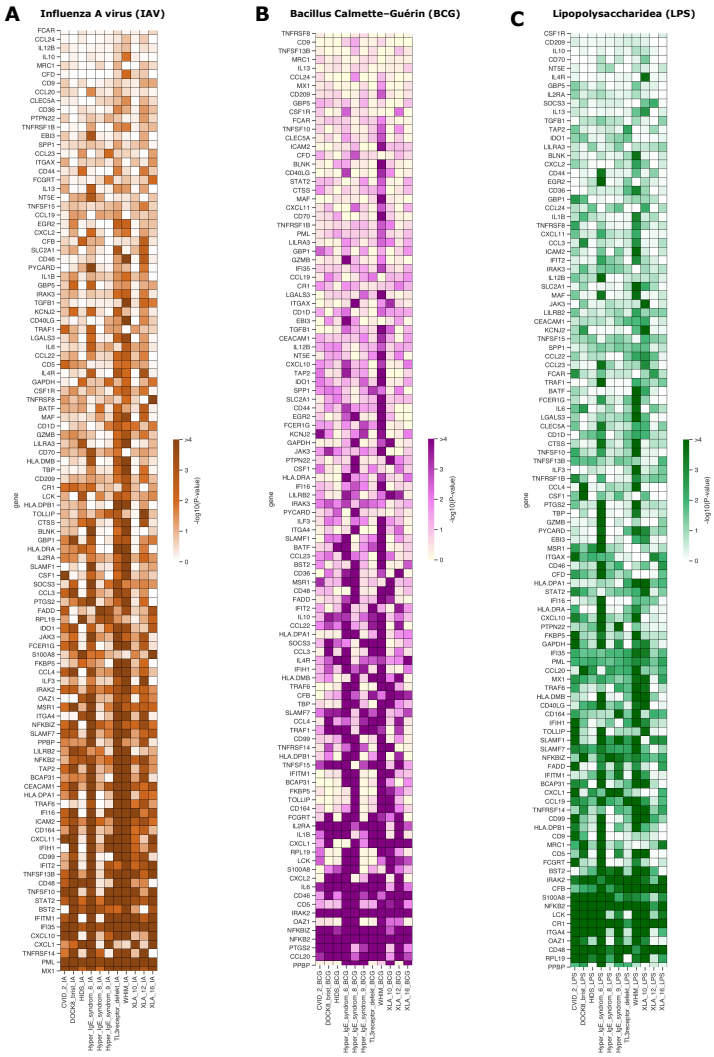

Figure 6.14: Comparison of gene expression profiles between healthy controls and PID patients. The statistical test shows the statistical significance levels of enrichment by -log10(Pvalue).

teractions that could determine the obtained immunological observables. For example, to mount an immune response, cooperativity of multiple immune cell populations provides inhibitory and stimulatory feedback. The use of such simple model of the immune system alleviates our lack of knowledge of most of the rules for immune cells interdependence. Indeed, such missing knowledge could be found in the literature. However, in many cases, studies are performed under different experimental conditions or answer different questions, hence a need to standardize assays [166, 167]. As a result, mathematical methods that establish which components are more relevant than others in giving rise to a given immune system behavior are used.

Our findings presented in this thesis have proven that immunology can benefit from a global view of the immune system. We have shown that a set of collective individual immune cell population can predict a healthy individual's immune response to stimulation. Also, we have identified a set of linear combination of gene expression profiles that predict variation of transcriptional responses to microbial stimulants in PID patients.

A possible criticism of the results we present here is that our results depend on measurements done on cohorts of particular individuals. Given the differences in genetics, environmental exposures, lifestyle, and immune phenotype in human populations worldwide, it would be interesting to carry out similar studies on independent cohorts from different geographical locations with similar or additional parameters measurement. For example, assessing maternal antiviral antibodies in children born in Africa or Asia could validate further our findings.

Chapter 7

Conclusions

In this section, I'm outlining the new results obtained during the course of this thesis.

- The human immune variation is continuous, and not described by discrete groups of individuals with similar immune cell populations. A set of aggregate immune cell population frequencies can define an individual's immunotype, and robustly predict diverse functional responses to cytokine stimulations. Although inter-individual variations in specific cell population frequencies can be large, older individuals have, by far, more heterogeneous immunotypes than younger individuals.

- The global repertoire of maternal antibodies target about 5-10 different viruses, and the transferred antibodies mirrors those found in the mothers. To our surprise, the repertoires of maternal antibodies are similar between very preterm children born before 30 weeks gestation and term children born after 37 weeks gestation. Also, the functional capability of antibodies was comparable until around three months of age. However, the concentrations of antibodies at birth are lower in preterm than in term children. The conferred protection by maternal antibodies lasts 2-3.5 months on average, and premature infants loose their antibodies much faster than full term children due to their lower concentrations transferred at the time of birth.

- We learned geometrical shapes of protein expression space of developing human newborn lymphocyte in early life. Single B cells are arranged in a triangle, while CD4+ T cells are best represented by pentahedron. The vertices of these structures are extreme protein expression profiles optimal for tasks and correspond to major cell subsets. Cells lie along a continuum of expression inside polytope. In triangle B cells, a 1D continuum of states describes cell specialization pattern to tasks and suggests pseudo-time trajectories in the developmental path

of newborn B cells.

- The variation of transcriptional responses to microbial stimulants is large among PID patients, and low in healthy individuals. Our results show that three combinations of a collective set of gene expression profiles can explain the differences in transcriptional responses to stimuli from PID patients. We identified gene variants associated with the differences in transcriptional responses to microbial stimulants between PID patients and healthy individuals allowing understanding immune functional defects in patients.

Acknowledgment

To many people and now limited to the following few, I'm sincerely grateful.

To **Petter Brodin**, my supervisor, for believing in me and providing me with an opportunity of getting involved in an interdisciplinary field of bioinformatics with a focus on systems immunology. When I joined the lab, I was a Physicist without any knowledge of immunology or analysis techniques used in the field. Your fearlessness and tireless way of working continue to inspire me. I will always be indebted to him for teaching me how to think like a research scientist while answering to a biological question. Thank you for your many suggestions, constant support, stimulating scientific conversations, and your guidance during my studies.

To **Jeff**, my co-supervisor, for contributing to my learning journey by providing me single-cell mRNA data and for the fruitful meetings we have had. Thank you!
To **Lukas**, my mentor, for his invaluable feedback on my work, and suggestions on how to carry out a bioinformatics research.

To **Dirk**, **Jane**, **Björn** for accepting to taking part in my public defense, **Mauno** for acting as an opponent as well as **Kanth** for agreeing to chair my defense.

To current and former members of BrodinLab at ScilifeLab: **Kanth**, **Jaromir**, **Yang**, **Jun**, **Constantin**, and **Anna Karin**, without your collective efforts in recruiting participants, sampling, transferring samples, storing, processing and generating the data, this thesis would not exist. To **Christian**, a long life collabo for generating most of the data presented in this thesis, thank you! To the new Ph.D. students **Tan**&**Lucie**, you both will surely do great and enjoy your journey of becoming researchers. To **Vijay**, **Nadia**, and **Qiu**, master students in the group. To **Camila**, for reading parts of this thesis, thank you! To you, all, thank you for your thoughtful conversations that we have shared. You have made my stay comfortable all these years and memorable.

To everyone in the **Lehtiö lab** and the **Ola Larsson labs**, for being good neighbors, sharing lunches and cakes on multiple occasions as well as for the fun I have had when our floor football team participated in the KI tournament.

To KI administration, **Anne Rasikari** and **Lillemor Melander**, for assisting me in my early days at KI, and KBH staff especially **Andrea Merker** for making sure that all the paperwork is in order in preparation for my public defense.

To the **National Graduate School in Medical Bioinformatics** for providing me with a platform to learn advanced bioinformatics, to present my research, to network, and get feedback.

To collaborators from hospital **Kajsa**, **Ewa**, and **Anna** for sharing your newborn cohort with us; to many others from abroad with whom we published together, I have learned a lot from you. To **Jean**, **Korem**, and **Adler**, members of Uri Alon lab for answering so many of my questions on archetype analysis. Special thank you to all parents and children who generously provided materials contained in this thesis.

To the amazing friends with whom I have spent time with playing volleyball after work: **Felix**, **Magnus**, **Martin**, **Aneta**, **Ewa**, **Magda**, **Mat**, **Michael**, **Ameya**, **Stefania**, **Simon**, **Claudia**, **Roque**, **Sanjar**, **Sofia**, **Sophia**, **Matilda**, **Kirill**, **Andreea**, **Vinicius**, **Nitin**, **Eric**, and everyone who plays volleyball at Electrolux. Thank you for organizing, hosting, partying, and winning so many matches all these years. Also, congratulations to our junior team in formation! **Aneta** will be a great coach of Team Drive junior!

Ku **babyeyi** banjye banyibarutse, banshyigikiye mu myigire yanjye, kandi bagakora uko bashoboye kose ngo nige neza kandi heza. To all my siblings for their love, encouragement, and support.

To my beautiful wife **Elise**, whose courage and determination never cease to amaze me, this thesis owes as much to her nurturing as it does to my effort. To our daughter **Amia** who brings joy and peace in our daily life.

Bibliography

[1] Gordon S. Phagocytosis: The Legacy of Metchnikoff. Cell. 2016;166(5):1065 – 1068.

[2] Travis J. On the Origin of The Immune System. Science (New York, NY). 2009 May;324(5927):580–582.

[3] Cooper MD. The early history of B cells. Nature reviews Immunology. 2015 Mar;15(3):191–197.

[4] Haskins, K, Kubo, R, White, J, Pigeon, M, Kappler, J, Marrack, P. The major histocompatibility complex-restricted antigen receptor on T cells. I. Isolation with a monoclonal antibody. Journal of Experimental Medicine. 1983 Apr;157(4):1149–1169.

[5] Burnet FM. A modification of Jerne's theory of antibody production using the concept of clonal selection. CA: a cancer journal for clinicians. 1976 Mar;26(2):119–121.

[6] Hozumi N, Tonegawa S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proceedings of the National Academy of Sciences. 1976 Oct;73(10):3628–3632.

[7] SNELL GD, HIGGINS GF. Alleles at the histocompatibility-2 locus in the mouse as determined by tumor transplantation. Genetics. 1951 May;36(3):306–310.

[8] Gurtner GC, Werner S, Barrandon Y, Longaker MT. Wound repair and regeneration. Nature. 2008 May;453(7193):314–321.

[9] Sorge RE, Mapplebeck JCS, Rosen S, Beggs S, Taves S, Alexander JK, et al. Different immune cells mediate mechanical pain hypersensitivity in male and female mice. Nature Neuroscience. 2015 Aug;18(8):1081–1083.

[10] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar;144(5):646–674.

[11] Rankin LC, Artis D. Beyond Host Defense: Emerging Functions of the Immune System in Regulating Complex Tissue Physiology. Cell. 2018 Apr;173(3):554–567.

[12] Steinman RM, Cohn ZA. IDENTIFICATION OF A NOVEL CELL TYPE IN PERIPHERAL LYMPHOID ORGANS OF MICE: I. MORPHOLOGY, QUANTITATION, TISSUE DISTRIBUTION. Journal of Experimental Medicine. 1973 May;137(5):1142–1162.

[13] Kiessling R, Klein E, Pross H, Wigzell H. „Natural" killer cells in the mouse. II. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Characteristics of the killer cell. European journal of immunology. 1975 Feb;5(2):117–121.

[14] Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. Journal of immunology (Baltimore, Md : 1950). 1986 Apr;136(7):2348–2357.

[15] Jerne NK. Towards a network theory of the immune system. Annales d'immunologie. 1974 Jan;125C(1-2):373–389.

[16] Davis MM, Brodin P. Rebooting Human Immunology. Annual Review of Immunology. 2018 Apr;36:843–864.

[17] Tauber A. The biological notion of self and non-self. 2002;.

[18] Tauber AI. The immune self: theory or metaphor? Immunology Today. 1994 Mar;15(3):134–136.

[19] Perelson AS, Weisbuch G. Theoretical and experimental insights into immunology; 2013.

[20] Perelson AS, Weisbuch G. Immunology for physicists. Reviews of Modern Physics. 1997 Oct;69(4):1219–1268.

[21] Cohen IR. Tending Adam's Garden: evolving the cognitive immune self; 2000.

[22] Cohen IR. The cognitive paradigm and the immunological homunculus. Immunology Today. 1992 Jan;13(12):490–494.

[23] Cohen IR. Discrimination and dialogue in the immune system. Seminars in immunology. 2000 Jun;12(3):215–219.

[24] Janeway CA, Travers P, Walport M, Science MSNYG, 2017. Immunobiology: the immune system in health and disease. 2005;.

[25] Harvill ET. Cultivating Our "Frienemies": Viewing Immunity as Microbiome Management. mBio. 2013 May;4(2):1.

[26] McFall-Ngai M. Care for the community. Nature. 2007 Jan;445(7124):153–153.

[27] Eberl G. A new vision of immunity: homeostasis of the superorganism. Mucosal Immunology. 2010 Sep;3(5):450–460.

[28] Rock KL, Latz E, Ontiveros F, Kono H. The Sterile Inflammatory Response. dxdoiorg. 2010 Mar;28(1):321–342.

[29] Marques RE, Marques PE, Guabiraba R, Teixeira MM. Exploring the Homeostatic and Sensory Roles of the Immune System. Frontiers in Immunology. 2016 Mar;7(2):e27.

[30] Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K, Jarrossay D, et al. Social network architecture of human immune cells unveiled by quantitative proteomics. Nature immunology. 2017 May;18(5):583–593.

[31] Merbl Y, Itzchak R, Vider-Shalit T, Louzoun Y, Quintana FJ, Vadai E, et al. A Systems Immunology Approach to the Host-Tumor Interaction: Large-Scale Patterns of Natural Autoantibodies Distinguish Healthy and Tumor-Bearing Mice. PLOS ONE. 2009 Jun;4(6):e6053.

[32] Agliari E, Annibale A, Barra A, Coolen ACC, Tantari D. Immune networks: multi-tasking capabilities at medium load. Journal of Physics A: Mathematical and Theoretical. 2013 jul;46(33):335101.

[33] Madi A, Hecht I, Bransburg-Zabary S, Merbl Y, Pick A, Zucker-Toledano M, et al. Organization of the autoantibody repertoire in healthy newborns and adults revealed by system level informatics of antigen microarray data. Proceedings of the National Academy of Sciences. 2009 Aug;106(34):14484–14489.

[34] Coffelt SB, Kersten K, Doornebal CW, Weiden J, Vrijland K, Hau CS, et al. IL-17-producing γ δ T cells and neutrophils conspire to promote breast cancer metastasis. Nature. 2015 Jun;522(7556):345–348.

[35] Martins LCA, Rocha NP, Torres KCL, Dos Santos RR, França GS, de Moraes EN, et al. Disease-specific expression of the serotonin-receptor 5-HT(2C) in natural killer cells in Alzheimer's dementia. Journal of neuroimmunology. 2012 Oct;251(1-2):73–79.

[36] Barbosa IG, Machado-Vieira R, Soares JC, Teixeira AL. The immunology of bipolar disorder. Neuroimmunomodulation. 2014;21(2-3):117–122.

[37] Germain RN, Schwartzberg PL. The human condition: an immunological perspective. Nature immunology. 2011 May;12(5):369–372.

[38] Jenq RR, van den Brink MRM. Allogeneic haematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. Nature reviews Cancer. 2010 Mar;10(3):213–221.

[39] Maloney DG, Grillo-López AJ, White CA, Bodkin D, Schilder RJ, Neidhart JA, et al. IDEC-C2B8 (Rituximab) anti-CD20 monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma. Blood. 1997 Sep;90(6):2188–2195.

[40] Calne RY, White DJ, Thiru S, Evans DB, McMaster P, Dunn DC, et al. Cyclosporin A in patients receiving renal allografts from cadaver donors. Lancet (London, England). 1978 Dec;2(8104-5):1323–1327.

[41] Feldmann M. Development of anti-TNF therapy for rheumatoid arthritis. Nature reviews Immunology. 2002 May;2(5):364–371.

[42] Kwon ED, Hurwitz AA, the BFPo, 1997. Manipulation of T cell costimulatory and inhibitory signals for immunotherapy of prostate cancer. National Acad Sciences;.

[43] Hodi FS, Mihm MC, Soiffer RJ, Haluska FG, Butler M, Seiden MV, et al. Biologic activity of cytotoxic T lymphocyte-associated antigen 4 antibody blockade in previously vaccinated metastatic melanoma and ovarian carcinoma patients. Proceedings of the National Academy of Sciences. 2003 Apr;100(8):4712–4717.

[44] Iwai Y, Terawaki S, Honjo T. PD-1 blockade inhibits hematogenous spread of poorly immunogenic tumor cells by enhanced recruitment of effector T cells. International Immunology. 2005 Feb;17(2):133–144.

[45] Leach DR, Krummel MF, Allison JP. Enhancement of antitumor immunity by CTLA-4 blockade. Science (New York, NY). 1996 Mar;271(5256):1734–1736.

[46] Chaiwatanatorn K, Lee N, Grigg A, Filshie R, Firkin F. Delayed-onset neutropenia associated with rituximab therapy. British journal of haematology. 2003 Jun;121(6):913–918.

[47] Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, et al. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. The New England journal of medicine. 2006 Sep;355(10):1018–1028.

[48] Doiron B, Hu W, DeFronzo RA. Beta Cell Formation in vivo Through Cellular Networking, Integration and Processing (CNIP) in Wild Type

Adult Mice. Current pharmaceutical biotechnology. 2016;17(4):376–388.

[49] Quintana-Murci L, Alcaïs A, Abel L, Casanova JL. Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. Nature immunology. 2007 Nov;8(11):1165–1171.

[50] Mestas J, Hughes CCW. Of mice and not men: differences between mouse and human immunology. Journal of immunology (Baltimore, Md : 1950). 2004 Mar;172(5):2731–2738.

[51] Davis MM. A prescription for human immunology. Immunity. 2008 Dec;29(6):835–838.

[52] Ernst PB, Carvunis AR. Of mice, men and immunity: a case for evolutionary systems biology. Nature immunology. 2018 May;19(5):421–425.

[53] Lauffenburger DA. Cell signaling pathways as control modules: complexity for simplicity? Proceedings of the National Academy of Sciences. 2000 May;97(10):5031–5033.

[54] Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999 Dec;402(6761):C47–C52.

[55] Brodin P, Valentini D, Uhlin M, Mattsson J, Zumla A, Maeurer MJ. Systems level immune response analysis and personalized medicine. Expert review of clinical immunology. 2014 Jan;9(4):307–317.

[56] Frei AP, Bava FA, Zunder ER, Hsieh EWY, Chen SY, Nolan GP, et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. nature methods. 2016 Mar;13(3):269–275.

[57] Picot J, Guerin CL, Le Van Kim C, Boulanger CM. Flow cytometry: retrospective, fundamentals and recent instrumentation. Cytotechnology. 2012 Jan;64(2):109–130.

[58] Perfetto SP, Chattopadhyay PK, Roederer M. Seventeen-colour flow cytometry: unravelling the immune system. Nature reviews Immunology. 2004 Aug;4(8):648–655.

[59] Roederer M, De Rosa S, Gerstein R, Anderson M, Bigos M, Stovel R, et al. 8 Color, 10-parameter flow cytometry to elucidate complex leukocyte heterogeneity. Cytometry. 1997 Dec;29(4):328–339.

[60] Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. Analytical chemistry. 2009 Jul;81(16):6813–6822.

[61] Horowitz A, Strauss-Albee DM, Leipold M, Kubo J, Nemat-Gorgani N, Dogan OC, et al. Genetic and Environmental Determinants of Human NK Cell Diversity Revealed by Mass Cytometry. Science translational medicine. 2013 Oct;5(208):208ra145–208ra145.

[62] 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 (New York, NY). 2011 May;332(6030):687–696.

[63] Olsen LR, Leipold MD, Pedersen CB, Maecker HT. The anatomy of single cell mass cytometry data. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2019 Feb;95(2):156–172.

[64] Chaussabel D. Assessment of immune status using blood transcriptomics and potential implications for global health. Seminars in immunology. 2015 Feb;27(1):58–66.

[65] Gaublomme JT, Yosef N, Lee Y, Gertner RS, Yang LV, Wu C, et al. Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. Cell. 2015 Dec;163(6):1400–1412.

[66] Han A, Glanville J, Hansmann L, Davis MM. Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nature biotechnology. 2014 Jul;32(7):684–692.

[67] Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. T cell fate and clonality inference from single-cell transcriptomes. nature methods. 2016 Apr;13(4):329–332.

[68] Lipkin WI. A Vision for Investigating the Microbiology of Health and Disease. The Journal of Infectious Diseases. 2015 Jul;212 Suppl 1:S26–30.

[69] Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, et al. Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. Science (New York, NY). 2015 Jun;348(6239):aaa0698–aaa0698.

[70] Zdeborová L, Krzakala F. Statistical physics of inference: thresholds and algorithms. Advances in Physics. 2016 Aug;65(5):453–552.

[71] Qiu P, Simonds EF, Bendall SC, Gibbs KD, Bruggner RV, Linderman MD, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nature biotechnology. 2011 Oct;29(10):886–891.

[72] Levine JH, Simonds EF, Bendall SC, Davis KL, Amir EaD, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell. 2015 Jul;162(1):184–197.

[73] Shekhar K, Brodin P, Davis MM, Chakraborty AK. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). Proceedings of the National Academy of Sciences. 2014 Jan;111(1):202–207.

[74] 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.

[75] Setty M, Tadmor MD, Reich-Zeliger S, Angel O, Salame TM, Kathail P, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature biotechnology. 2016 Jun;34(6):637–645.

[76] Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics (Oxford, England). 2015 Sep;31(18):2989–2998.

[77] Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature biotechnology. 2014 Apr;32(4):381–386.

[78] Bendall SC, Davis KL, Amir EAD, Tadmor MD, Simonds EF, Chen TJ, et al. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. Cell. 2014 Apr;157(3):714–725.

[79] Brodin P, Jojic V, Gao T, Bhattacharya S, Angel CJL, Furman D, et al. Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences. Cell. 2015 Jan;160(1-2):37–47.

[80] Kaczorowski KJ, Shekhar K, Nkulikiyimfura D, Dekker CL, Maecker H, Davis MM, et al. Continuous immunotypes describe human immune variation and predict diverse responses. Proceedings of the National Academy of Sciences. 2017 Jul;114(30):E6097–E6106.

[81] Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM. Cytometry by Time-of-Flight Shows Combinatorial Cytokine Expression and Virus-Specific Cell Niches within a Continuum of CD8+ T Cell Phenotypes. Immunity. 2012 Jan;36(1):142–152.

[82] 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.

[83] Brodin P, Davis MM. Human immune system variation. Nature reviews Immunology. 2017 Jan;17(1):21–29.

[84] Chuang HY, Hofree M, Ideker T. A Decade of Systems Biology. dxdoiorg. 2010 Oct;26(1):721–744.

[85] Carr EJ, Dooley J, Garcia-Perez JE, Lagou V, Lee JC, Wouters C, et al. The cellular composition of the human immune system is shaped by age and cohabitation. Nature immunology. 2016 Apr;17(4):461–468.

[86] Tsang JS, Schwartzberg PL, Kotliarov Y, Biancotto A, Xie Z, Germain RN, et al. Global Analyses of Human Immune Variation Reveal Baseline Predictors of Postvaccination Responses. Cell. 2014 Apr;157(2):499–513.

[87] Shen-Orr SS, Furman D, Kidd BA, Hadad F, Lovelace P, Huang YW, et al. Defective Signaling in the JAK-STAT Pathway Tracks with Chronic Inflammation and Cardiovascular Risk in Aging Humans. Cell systems. 2016 Oct;3(4):374–384.e4.

[88] Patin, Etienne, Hasan, Milena, Bergstedt, Jacob, Rouilly, Vincent, Libri, Valentina, Urrutia, Alejandra, et al. Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. Nature immunology. 2018 Mar;19(3):302–314.

[89] Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, et al. Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. Cell. 2018 May;173(6):1385–1397.e14.

[90] Steiniger B, Ulfig N, Risse M, Barth PJ. Fetal and early post-natal development of the human spleen: from primordial arterial B cell lobules to a non-segmented organ. Histochemistry and cell biology. 2007 Sep;128(3):205–215.

[91] Frascoli M, Coniglio L, Witt R, Jeanty C, Fleck-Derderian S, Myers DE, et al. Alloreactive fetal T cells promote uterine contractility in preterm labor via IFN-γ and TNF-α. Science translational medicine. 2018 Apr;10(438):eaan2263.

[92] Li N, van Unen V, Guo N, Abdelaal T, Somarakis A, Eggermont J, et al. Early-Life Compartmentalization of Immune Cells in Human Fe-

tal Tissues Revealed by High-Dimensional Mass Cytometry. Frontiers in Immunology. 2019;10:1932.

[93] Schaffert S, Khatri P. Early life immunity in the era of systems biology: understanding development and disease. Genome medicine. 2018 Nov;10(1):88.

[94] Zhang X, Zhivaki D, Lo-Man R. Unique aspects of the perinatal immune system. Nature reviews Immunology. 2017 Aug;17(8):495–507.

[95] Palmer AC. Nutritionally mediated programming of the developing immune system. Adv Nutr 2: 377–395; 2011.

[96] Costa G, Kouskoff V, Lacaud G. Origin of blood cells and HSC production in the embryo. Trends in immunology. 2012 May;33(5):215–223.

[97] Shearer WT, Rosenblatt HM, Gelman RS, Oyomopito R, Plaeger S, Stiehm ER, et al. Lymphocyte subsets in healthy children from birth through 18 years of age: the Pediatric AIDS Clinical Trials Group P1009 study. The Journal of allergy and clinical immunology. 2003 Nov;112(5):973–980.

[98] Tosato F, Bucciol G, Pantano G, Putti MC, Sanzari MC, Basso G, et al. Lymphocytes subsets reference values in childhood. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2015 Jan;87(1):81–85.

[99] van Gent R, van Tilburg CM, Nibbelke EE, Otto SA, Gaiser JF, Janssens-Korpela PL, et al. Refined characterization and reference values of the pediatric T- and B-cell compartments. Clinical immunology (Orlando, Fla). 2009 Oct;133(1):95–107.

[100] Fragiadakis GK, Baca QJ, Gherardini PF, Ganio EA, Gaudilliere DK, Tingle M, et al. Mapping the Fetomaternal Peripheral Immune System at Term Pregnancy. Journal of immunology (Baltimore, Md : 1950). 2016 Dec;197(11):4482–4492.

[101] Comans-Bitter WM, de Groot R, van den Beemd R, Neijens HJ, Hop WC, Groeneveld K, et al. Immunophenotyping of blood lymphocytes in childhood. Reference values for lymphocyte subpopulations. The Journal of pediatrics. 1997 Mar;130(3):388–393.

[102] Kent A, Scorrer T, Pollard AJ, Snape MD, Clarke P, Few K, et al. Lymphocyte subpopulations in premature infants: an observational study. Archives of disease in childhood Fetal and neonatal edition. 2016 Nov;101(6):F546–F551.

[103] Prabhu SB, Rathore DK, Nair D, Chaudhary A, Raza S, Kanodia P, et al. Comparison of Human Neonatal and Adult Blood Leukocyte Subset Composition Phenotypes. PLOS ONE. 2016 Sep;11(9):e0162242.

[104] Malek A, Sager R, Kuhn P, Nicolaides KH, Schneider H. Evolution of maternofetal transport of immunoglobulins during human pregnancy. American journal of reproductive immunology (New York, NY : 1989). 1996 Nov;36(5):248–255.

[105] Simister N. Placental transport of immunoglobulin G. Vaccine. 2003 Jul;21(24):3365–3369.

[106] Roopenian DC, Akilesh S. FcRn: the neonatal Fc receptor comes of age. Nature reviews Immunology. 2007 Sep;7(9):715–725.

[107] Stapleton NM, Einarsdóttir HK, Stemerding AM, Vidarsson G. The multiple facets of FcRn in immunity. Immunological Reviews. 2015 Nov;268(1):253–268.

[108] Vidarsson G, Dekkers G, Rispens T. IgG Subclasses and Allotypes: From Structure to Effector Functions. Frontiers in Immunology. 2014 Oct;5(16):1.

[109] de Voer RM, van der Klis FRM, Nooitgedagt JE, Versteegh FGA, van Huisseling JCM, van Rooijen DM, et al. Seroprevalence and placental transportation of maternal antibodies specific for Neisseria meningitidis serogroup C, Haemophilus influenzae type B, diphtheria, tetanus, and pertussis. Clinical Infectious Diseases. 2009 Jul;49(1):58–64.

[110] Leineweber B, Grote V, Schaad B, Heininger U. TRANSPLACENTALLY ACQUIRED IMMUNOGLOBULIN G ANTIBODIES AGAINST MEASLES, MUMPS, RUBELLA AND VARICELLA-ZOSTER VIRUS IN PRETERM AND FULL TERM NEWBORNS. The Pediatric Infectious Disease Journal. 2004 Apr;23(4):361.

[111] Palmeira P, Quinello C, Silveira-Lessa AL, Zago CA, Carneiro-Sampaio M. IgG placental transfer in healthy and pathological pregnancies. Clinical and Developmental Immunology. 2012;2012:985646.

[112] Firan M, Bawdon R, Radu C, Ober RJ, Eaken D, Antohe F, et al. The MHC class I-related receptor, FcRn, plays an essential role in the maternofetal transfer of gamma-globulin in humans. International Immunology. 2001 Aug;13(8):993–1002.

[113] Heininger U, Riffelmann M, Leineweber B, Wirsing von Koenig CH. MATERNALLY DERIVED ANTIBODIES AGAINST BORDETELLA PERTUSSIS ANTIGENS PERTUSSIS TOXIN AND

FILAMENTOUS HEMAGGLUTININ IN PRETERM AND FULL TERM NEWBORNS. The Pediatric Infectious Disease Journal. 2009 May;28(5):443–445.

[114] van den Berg JP, Westerbeek EAM, van der Klis FRM, Berbers GAM, van Elburg RM. Transplacental transport of IgG antibodies to preterm infants: A review of the literature. Early human development. 2011 Feb;87(2):67–72.

[115] van den Berg JP, Westerbeek EAM, Smits GP, van der Klis FRM, Berbers GAM, van Elburg RM. Lower Transplacental Antibody Transport for Measles, Mumps, Rubella and Varicella Zoster in Very Preterm Infants. PLOS ONE. 2014 Apr;9(4):e94714.

[116] Conway SP, Dear PR, Smith I. Immunoglobulin profile of the preterm baby. Archives of disease in childhood. 1985 Mar;60(3):208–212.

[117] Malek A, Sager R, Kuhn P, Nicolaides KH, Schneider H. Evolution of Maternofetal Transport of Immunoglobulins During Human Pregnancy. American Journal of Reproductive Immunology. 1996 Nov;36(5):248–255.

[118] Ochola R, Sande C, Fegan G, Scott PD, Medley GF, Cane PA, et al. The Level and Duration of RSV-Specific Maternal IgG in Infants in Kilifi Kenya. PLOS ONE. 2009 Dec;4(12):e8088.

[119] Kliks SC, Wara DW, Landers DV, Levy JA. Features of HIV-1 that could influence maternal-child transmission. JAMA. 1994 Aug;272(6):467–474.

[120] Waaijenborg S, Hahné SJM, Mollema L, Smits GP, Berbers GAM, van der Klis FRM, et al. Waning of maternal antibodies against measles, mumps, rubella, and varicella in communities with contrasting vaccination coverage. The Journal of Infectious Diseases. 2013 Jul;208(1):10–16.

[121] Angerer P, Haghverdi L, Büttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics (Oxford, England). 2016 Apr;32(8):1241–1243.

[122] Yi B, Rykova M, Jäger G, Feuerecker M, Hörl M, Matzel S, et al. Influences of large sets of environmental exposures on immune responses in healthy adult men. Scientific Reports. 2015 Aug;5:13367.

[123] Gollwitzer ES, Marsland BJ. Impact of Early-Life Exposures on Immune Maturation and Susceptibility to Disease. Trends in immunology. 2015 Nov;36(11):684–696.

[124] Goenka A, Kollmann TR. Development of immunity in early life. Journal of Infection. 2015 Jun;71 Suppl 1:S112–20.

[125] Torow N, Hornef MW. The Neonatal Window of Opportunity: Setting the Stage for Life-Long Host-Microbial Interaction and Immune Homeostasis. Journal of immunology (Baltimore, Md : 1950). 2017 Jan;198(2):557–563.

[126] Garand M, Cai B, Kollmann TR. Environment impacts innate immune ontogeny. Innate immunity. 2017 Jan;23(1):3–10.

[127] Almanzar G, Schönlaub J, Hammerer-Lercher A, Koppelstaetter C, Bernhard D, Prelog M. Influence of the delivery modus on subpopulations and replication of lymphocytes in mothers and newborns. Early human development. 2015 Dec;91(12):663–670.

[128] Yektaei-Karin E, Moshfegh A, Lundahl J, Berggren V, Hansson LO, Marchini G. The stress of birth enhances in vitro spontaneous and IL-8-induced neutrophil chemotaxis in the human newborn. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology. 2007 Dec;18(8):643–651.

[129] Zanardo V, Solda G, Trevisanuto D. Elective cesarean section and fetal immune-endocrine response. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics. 2006 Oct;95(1):52–53.

[130] Vogl SE, Worda C, Egarter C, Bieglmayer C, Szekeres T, Huber J, et al. Mode of delivery is associated with maternal and fetal endocrine stress response. BJOG : an international journal of obstetrics and gynaecology. 2006 Apr;113(4):441–445.

[131] Yildiran A, Yurdakul E, Guloglu D, Dogu F, Arsan S, Arikan M, et al. The effect of mode of delivery on T regulatory (Treg) cells of cord blood. Indian journal of pediatrics. 2011 Oct;78(10):1234–1238.

[132] Du Toit G, Sayre PH, Roberts G, Sever ML, Lawson K, Bahnson HT, et al. Effect of Avoidance on Peanut Allergy after Early Peanut Consumption. The New England journal of medicine. 2016 Apr;374(15):1435–1443.

[133] Randomized Trial of Peanut Consumption in Infants at Risk for Peanut Allergy. The New England journal of medicine. 2016 Jul;375(4):398.

[134] van den Heuvel D, Jansen MAE, Nasserinejad K, Dik WA, van Lochem EG, Bakker-Jonges LE, et al. Effects of nongenetic factors on immune cell dynamics in early childhood: The Generation R Study. The Journal of allergy and clinical immunology. 2017 Jun;139(6):1923–1934.e17.

[135] Turfkruyer M, Verhasselt V. Breast milk and its impact on maturation of the neonatal immune system. Current opinion in infectious diseases. 2015 Jun;28(3):199–206.

[136] Winkler B, Aulenbach J, Meyer T, Wiegering A, Eyrich M, Schlegel PG, et al. Formula-feeding is associated with shift towards Th1 cytokines. European journal of nutrition. 2015 Feb;54(1):129–138.

[137] Andersson Y, Hammarström ML, Lönnerdal B, Graverholt G, Fält H, Hernell O. Formula feeding skews immune cell composition toward adaptive immunity compared to breastfeeding. Journal of immunology (Baltimore, Md : 1950). 2009 Oct;183(7):4322–4328.

[138] Almanzar G, Eberle G, Lassacher A, Specht C, Koppelstaetter C, Heinz-Erian P, et al. Maternal cigarette smoking and its effect on neonatal lymphocyte subpopulations and replication. BMC pediatrics. 2013 Apr;13:57.

[139] Lim ES, Wang D, Holtz LR. The Bacterial Microbiome and Virome Milestones of Infant Development. Trends in microbiology. 2016 Oct;24(10):801–810.

[140] Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. Nature medicine. 2016 Jul;22(7):713–722.

[141] Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012 May;486(7402):222–227.

[142] Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, et al. Association of host genome with intestinal microbial composition in a large healthy cohort. Nature genetics. 2016 Nov;48(11):1413–1417.

[143] Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014 Nov;159(4):789–799.

[144] Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. Cell host & microbe. 2016 May;19(5):731–743.

[145] Yassour M, Vatanen T, Siljander H, Hämäläinen AM, Härkönen T, Ryhänen SJ, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Science translational medicine. 2016 Jun;8(343):343ra81–343ra81.

[146] Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Science translational medicine. 2016 Jun;8(343):343ra82–343ra82.

[147] Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. Proceedings of the National Academy of Sciences of the United States of America. 2011 Mar;108 Suppl 1(Supplement 1):4578–4585.

[148] Korpela K, Salonen A, Virta LJ, Kekkonen RA, Forslund K, Bork P, et al. Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. Nature communications. 2016 Jan;7(1):10410.

[149] Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen AM, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell host & microbe. 2015 Feb;17(2):260–273.

[150] Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell. 2016 May;165(4):842–853.

[151] Roederer M, Quaye L, Mangino M, Cell MB, 2015. The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. Elsevier;.

[152] Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK. A deep profiler's guide to cytometry. Trends in immunology. 2012 Jul;33(7):323–332.

[153] Rosipal R, Krämer N. Overview and Recent Advances in Partial Least Squares. In: Subspace, Latent Structure and Feature Selection. Berlin, Heidelberg: Springer, Berlin, Heidelberg; 2006. p. 34–51.

[154] Wei SC, Duffy CR, Allison JP. Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. Cancer discovery. 2018 Sep;8(9):1069–1086.

[155] Sharma P, Allison JP. The future of immune checkpoint therapy. Science (New York, NY). 2015 Apr;348(6230):56–61.

[156] Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, et al. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. Science (New York, NY). 2012 Jun;336(6085):1157–1160.

[157] Oster GF, Wilson EO. Caste and ecology in the social insects; 1978.

[158] Sheftel H, Shoval O, Mayo A, Alon U. The geometry of the Pareto front in biological phenotype space. Ecology and evolution. 2013 Jun;3(6):1471–1483.

[159] Korem Y, Szekely P, Hart Y, Sheftel H, Hausser J, Mayo A, et al. Geometry of the Gene Expression Space of Individual Cells. PLoS computational biology. 2015 Jul;11(7):e1004224.

[160] Mørup M, Neurocomputing LH, 2012. Archetypal analysis for machine learning and data mining. Elsevier;.

[161] Cutler A, Technometrics LB, 1994. Archetypal analysis. Taylor & Francis;.

[162] Thøgersen JC, Mørup M, Damkiær S, Molin S, Jelsbak L. Archetypal analysis of diverse Pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis airways. BMC Bioinformatics. 2013 Dec;14(1):1–15.

[163] Bioucas-Dias JM. A variable splitting augmented Lagrangian approach to linear spectral unmixing. In: 2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). IEEE;. p. 1–4.

[164] Adler M, Korem Kohanim Y, Tendler A, Mayo A, Alon U. Continuum of Gene-Expression Profiles Provides Spatial Division of Labor within a Differentiated Cell Type. Cell systems. 2019 Jan;8(1):43–52.e5.

[165] van den Hof S, Wallinga J, Widdowson MA, Conyn-van Spaendonck MAE. Protecting the vaccinating population in the face of a measles epidemic: assessing the impact of adjusted vaccination schedules. Epidemiology and infection. 2002 Feb;128(1):47–57.

[166] Lithgow GJ, Driscoll M, Phillips P. A long journey to reproducible results. Nature. 2017 Aug;548(7668):387–388.

[167] Hines WC, Su Y, Kuhn I, Polyak K, Bissell MJ. Sorting out the FACS: a devil in the details. Cell reports. 2014 Mar;6(5):779–781.