From the Center for Infectious Medicine,
Karolinska Institutet, Stockholm, Sweden

# INNATE LYMPHOID CELL PLASTICITY AND HETEROGENEITY IN HUMAN TISSUES

Luca Mazzurana


Karolinska Institutet

Stockholm 2021

**Cover Image**: Crossing thoughts. An immunologist's mind during the research process. Graphical representation of machinery, technologies, organs, genes and cells fundamental to the creation of this thesis.

Image designed by Art Director © Luca Monterosso-Wittmann. https://www.lucamw.com/


Published by Karolinska Institutet.
Printed by US-AB

ISBN 978-91-8016-093-3

# Innate lymphoid cell plasticity and heterogeneity in human tissues

## THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Luca Mazzurana**

Public defence: **Friday 26th of February, 2021 at 9:30am**
Lecture Hall 4V, Alfred Nobels allè 8, Huddinge

*Principal Supervisor:*
Jenny Mjösberg, Associate Professor
Karolinska Institutet
Department of Medicine, Huddinge
Center for Infectious Medicine

*Co-supervisor(s):*
Mauro D'Amato, Professor
Monash University
School of Biological Sciences

Niklas Björkström, Associate Professor
Karolinska Institutet
Department of Medicine, Huddinge
Center for Infectious Medicine

Hans-Gustaf Ljunggren, Professor
Karolinska Institutet
Department of Medicine, Huddinge
Center for Infectious Medicine

*Opponent:*
Daniela Finke, Professor
University of Basel
Department of Biomedicine

*Examination Board:*
Susanna Cardell, Professor
University of Gothenburg
Institute of Biomedicine

Helen Kaipe, Associate Professor
Karolinska Institutet
Department of Laboratory Medicine

Mikael Benson, Professor
Linköping University
Department of Biomedical and Clinical Sciences

To my family

*“Con metodo, con costanza, a modo mio”*

M. Zambotti

# ABSTRACT

The human immune system is a vital mechanism that protects the host from outside threats. The two main branches of this system, consisting of innate and adaptive immunity, aid the host when dangers of both imminent and protracted nature occur. Innate lymphoid cells (ILC) play a key role in innate immunity and are classified into distinct groups based on their function, transcriptional profile and development. Their discovery is young, with much research needed to understand all aspects of their phenotype, genotype, behavior in homeostasis and disease.

This thesis describes the breakthroughs my collaborators and I were able to reach during my five years of doctoral studies, in order to better understand the biology and physiology of ILC. In the first part of this thesis I describe the history and classification of ILC, the efforts being done so far by the field to understand their development, and the possible combinations of changes in ILC status, known as plasticity or trans-differentiation properties of ILC under particular environments or stimuli. As **Paper II**'s results are based on a cohort of Inflammatory Bowel Disease (IBD) samples, and in **Paper I** we analyze ILC in intestinal biopsies from IBD patients**,** I also outline the main characteristics representing this disorder while focusing on the role of ILC in IBD.

Next, after defining the aim of my studies, I describe how the data was obtained, as a big part of my work consisted in learning how to handle methods and technologies such as flow cytometry, fluorescence-activated cell sorting (FACS) and single cell RNA sequencing, among others.

Finally, a discussion of the results is aimed at highlighting the major breakthroughs achieved. In **Paper I**, we set out to better understand the function of the Ikaros family of transcription factors in ILC, focusing our efforts on IKZF3 (encoding Aiolos) and its role in ILC trans-differentiation via a drug-induced silencing approach with the immune-modulatory agent lenalidomide. In **Paper II**, a large cohort of IBD samples was analyzed in order to find disturbances in peripheral blood ILC protein expression. We were able to uncover differences in several activation proteins in the IBD cohort, when compared to a like-sized cohort of samples from healthy controls. In **Paper III**, a big effort was put into implementing Smart-seq2 RNA sequencing technology to a large number of ILC from a variety of tissues. This allowed us to better understand the heterogeneity of ILC in the circulation, secondary lymphoid and mucosal tissues. We generated a large dataset that will require time to be exploited in full, constituting a roadmap for future studies aimed at understanding human ILC biology and function.

In summary, the work presented in this thesis provides findings and datasets that have the potential to advance the ILC field.

# LIST OF SCIENTIFIC PAPERS

I. **Luca Mazzurana**, Marianne Forkel, Anna Rao, Aline Van Acker, Efthymia Kokkinou, Tamaki Ichiya, Sven Almer, Charlotte Höög, Daniella Friberg and Jenny Mjösberg. Suppression of Aiolos and Ikaros expression by lenalidomide reduces human ILC3-ILC1/NK cell transdifferentiation. *European Journal of Immunology*. 2019. 49:1344-1355. Copyright Wiley-VCH Verlag Gmbh & Co. KGaA, Weinheim . Reproduced with permission, license number 4960230662656.

II. **Luca Mazzurana**, Ferdinando Bonfiglio, Marianne Forkel, Mauro D'Amato, Jonas Halfvarson and Jenny Mjösberg. Chron's disease is associated with activation of circulating innate lymphoid cells. *Inflammatory Bowel Diseases*. 2020. izaa316. Copyright © Oxford University Press. Reproduced with permission, license number 4967751102509.

III. **Luca Mazzurana**, Paulo Czarnewski, Viktor Jonsson, Leif Wigge, Markus Ringnér, Teresa C Williams, Avinash Ravindran, Åsa K Björklund, Jesper Säfholm, Gunnar Nilsson, Sven-Erik Dahlén, Ann-Charlotte Orre, Mamdoh Al-Ameri, Charlotte Höög, Charlotte Hedin, Sylwester Szczegielnak, Sven Almer and Jenny Mjösberg. Tissue-specific transcriptional imprinting and heterogeneity in human innate lymphoid cells revealed by full-length single-cell RNA-sequencing. *Cell Research*. 2021.

# SCIENTIFIC PAPERS NOT INCLUDED IN THE THESIS

IV. Yugo Ando*, **Luca Mazzurana***, Marianne Forkel, Kazuichi Okazaki, Mamiko Aoi, Peter T Schmidt, Jenny Mjösberg and Francesca Bresso. Downregulation of MicroRNA-21 in colonic $CD3^+$ T cells in UC remission. *Inflammatory Bowel Diseases*. 2016. 22:2788-2793.

V. Jenny Mjösberg and **Luca Mazzurana**. ILC-poiesis: Making tissue ILC from blood. *Immunity*. 2017. 46:344-346. Preview.

VI. Jovana Maric, Avinash Ravindran, **Luca Mazzurana**, Åsa K Björklund, Aline Van Acker, Anna Rao, Daniella Friberg, Sven-Erik Dáhlen, Akos Heinemann, Viktoria Konya and Jenny Mjösberg. Prostaglandin $E_2$ suppresses human group 2 innate lymphoid cell function. *The Journal of Allergy and Clinical Immunology*. 2018. 141:1761-1773.

VII. **Luca Mazzurana**, Anna Rao, Aline Van Acker and Jenny Mjösberg. The roles of innate lymphoid cells in the human immune system. *Seminars in Immunopathology*. 2018. 40:407-119. Review.

VIII. Avinash Ravindran, Elin Rönnberg, Joakim S Dahlin, **Luca Mazzurana**, Jesper Säfholm, Ann-Charlotte Orre, Mamdoh Al-Ameri, Peter Peachell, Mikael Adner, Sven-Erik Dahlén, Jenny Mjösberg and Gunnar Nilsson. An optimized protocol for the isolation and functional analysis of human lung mast cells. *Frontiers of Immunology*. 2018. 9:2193.

IX. Jovana Maric, Avinash Ravindran, **Luca Mazzurana**, Aline Van Acker, Anna Rao, Efthymia Kokkinou, Maria Ekoff, Dominique Thomas, Alexander Fauland, Gunnar Nilsson, Craig E Wheelock, Sven-Erik Dáhlen, Nerea Ferreirós, Gerd Geisslinger, Daniella Friberg, Akos Heinemann, Viktoria Konya and Jenny Mjösberg. Cytokine-induced endogenous production of prostaglandin $D_2$ is essential for human group 2 innate lymphoid cell activation. *The Journal of Allergy and Clinical Immunology*. 2019. 143:2202-2214.

X. Anna Rao, Otto Strauss, Efthymia Kokkinou, Mélanie Bruchard, Kumar P Tripathi, Heinrich Schlums, Anna Carrasco, **Luca Mazzurana**, Viktoria Konya, Eduardo J Villablanca, Niklas K Björkström, Ulrik Lindforss, Hergen Spits and Jenny Mjösberg. Cytokines regulate the antigen-presenting characteristics of human circulating and tissue-resident instestinal ILCs. *Nature Communications*. 2020. 11:2049.

* equal contribution

# CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AHR | Aryl hydrocarbon receptor |
| APC | Antigen presenting cell |
| AREG | Amphiregulin |
| CD | Cluster of differentiation |
| CHILP | Common helper innate lymphoid progenitor |
| CILP | Common innate lymphoid progenitor |
| CLP | Common lymphoid progenitor |
| CyTOF | Cytometry by time of flight |
| DC | Dendritic cell |
| DE | Differential expression |
| DM | Diffusion map |
| dNTP | Deoxyribose nucleoside triphosphate |
| eQTL | Expression quantitative trait loci |
| FACS | Fluorescence-activated cell sorting |
| GI | Gastrointestinal |
| GTeX | Genotype-Tissue Expression project |
| GvHD | Graft-versus-host disease |
| GWAS | Genome-wide association study |
| HC | Healthy control |
| HSC | Hematopoietic stem cell |
| HSCT | Hematopoietic stem cell transfer |
| IBD | Inflammatory bowel disease |

| | |
|---|---|
| ieILC1 | Intraepithelial ILC1 |
| IFN-γ | Interferon-y |
| iILC2 | Inflammatory ILC2 |
| IKZF3 | IKAROS family zinc finger 3 |
| IL | Interleukin |
| ILC | Innate lymphoid cells |
| ILCp | ILC precursor |
| LTi | Lymphoid tissue inducer |
| LTip | Lymphoid tissue inducer progenitor |
| MMLV | Moloney murine leukemia virus |
| MSP | Minimum spanning tree |
| MZ | Marginal zone |
| NCAM | Neural cell adhesion molecule |
| NCR | Natural cytotoxicity receptor |
| NF- kβ | Nuclear factor k β |
| NGS | Next generation sequencing |
| nILC2 | Natural ILC2 |
| NK cell | Natural killer cell |
| NKp | Natural killer cell precursor |
| NSAID | Non-steroidal anti-inflammatory drug |
| RNA-seq | RNA sequencing |
| RORγt | Retinoid-related orphan receptor yt |

| | |
|---|---|
| scRNA-seq | Single-cell RNA sequencing |
| Smart | Switch mechanism at the 5’ end of RNA templates |
| SNP | Single nucleotide polymorphism |
| SS2 | Smart-seq2 |
| t-SNE | t-distributed stochastic neighbor embedding analysis |
| Th | T helper |
| TSO | Template switching oligonucleotides |
| UC | Ulcerative Colitis |
| UMAP | Uniform manifold approximation and projection |
| UMI | Unique molecular identifier |
| WGCNA | Weighted correlation network analysis |

# 1 INTRODUCTION

Throughout life, our body is constantly under attack, and more often than not we are unaware of this fact. We ought to be grateful to our defense system, that protects us from a variety of threats, including bacteria, viruses, fungi, protozoa, worms and other pathogens. We are able to exploit this ancient mechanism thanks to evolution, which in its broadest sense, took one billion years to develop [1] via a very slow process, from phagocytosis in protozoa to the more complex immune system of mammals [2]. In this thesis, I will discuss my research on several cogs in this vital machinery called the innate immune system, comprised of a variety of cells ready to defend us from foreign attacks within minutes. One important piece of this early defense involves innate lymphoid cells (ILC), the focus of my research.

## 1.1 INNATE LYMPHOID CELLS

Human innate lymphoid cells (ILCs) are a type of lymphocyte lacking rearranged antigen specific receptors, often regarded as the innate counterpart of T cells [34]. ILCs provide a rapid immune response upon threat, regulating inflammation by quickly releasing cytokines, before T cells take over. So far, they have been reported as residing in many tissues (i.e. adipose tissue [5], tonsils [6], gut [7], thymus [8], skin [9], liver [10], lung [11], uterus [12], bone marrow [13] and nasal polyps [14]) and , at low frequency, in circulation [15].

Their discovery spans over five decades [16], with most findings made during the last 10 years [17]. In the early 1970's, a series of papers and reports [18] describing "natural" cytotoxic reactivity to cancer in humans [19] and mouse [20] were published, coining the cells responsible for this activity as "N" cells [20]. In 1975, at Karolinska Institutet, similar results were reported [21], changing the name of the cells to natural killer (NK) cells. After these discoveries, the next breakthrough in the innate lymphocyte field was made in the late 90's, with the discovery of lymphoid tissue inducer cells [22]. It took an additional 10 years for the field to realize that the innate lymphoid cell family was more heterogeneous than once thought, with a series of discoveries ranging from defining transcription factor differences among subsets [23], to innate type 2 cytokine producers [24,25,26] and finally the identification of ILC progenitors [27,28], providing proof of NK cells being developmentally different than the rest of ILCs. During the current decade, a number of novel discoveries were made in the field, shedding more light to the innate lymphoid cell biology, with a proposal for uniform nomenclature in 2013 [29]. Recently, smaller novel ILC subsets are continuously being discovered [6], but ILCs are now classified into five major subsets (**Figure 1**), based on their developmental trajectories [17].

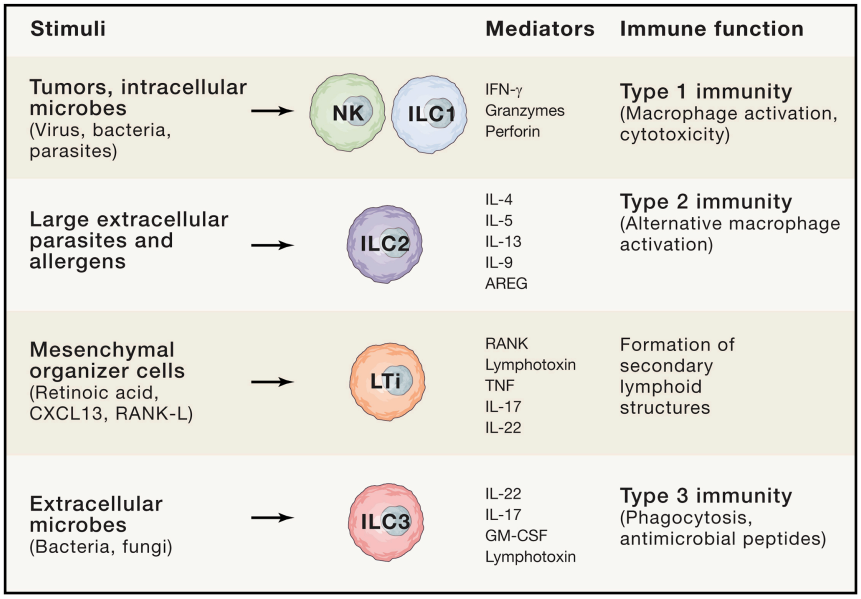


**Figure 1**. Classification and immune function of ILC. Reproduced with permission from Vivier et al. 2018. *Cell* [17].

### 1.1.1 ILC1 and NK cells

ILC1 are commonly defined as IFN-γ producing cells [17], and are dependent on the transcription factor T-bet (encoded by *TBX21*) for their development. In a recent study identifying the RNA transcripts of ILCs at the single cell level [6], other transcription factors have been reported to be differentially expressed by ILC1, including Aiolos (encoded by *IKZF3*), and LEF-1. Their role in ILC1 development is still not clear, and studies to elucidate this are needed since ILC1 are very heterogeneous, and the least characterized ILC subset.

ILC1 are considered as the innate counterparts of T helper 1 (Th1) cells [3], since they mirror their cytokine capabilities and transcription factor profile. ILC1 are generally non-cytotoxic, producing little granzyme B and no perforin [30]. In addition to IFN-γ, ILC1 also produce TNF [31] upon IL-12 and IL-18 stimulus [30], cytokines produced by dendritic (DC) cells and monocytes under inflammatory conditions [32]. ILC1 produce IFN-γ, which inhibits viral replication and activate macrophages, which in turn increase their phagocytic activity and upregulate major histocompatibility complexes for antigen presentation, in order to initiate an adaptive response [33]. In the mouse, intracellular pathogens such and *Toxoplasma gondii* and *Salmonella enterica* were shown to be cleared by IFN-γ releasing ILC1 [27,34].

A subgroup of ILC1 was discovered recently in the intestinal epithelium [35]. Intraepithelial ILC1 express T-bet, but lack the defining ILC marker, CD127. Interestingly, they also express eomesodermin (encoded by *EOMES*) and require NFIL3 (also known as *E4BP4* [36]) for their development.

As ILC1 are defined as lacking key surface markers expressed by other ILC groups (i.e. CRTH2, CKIT, NKp44), and the expression of its defined transcription factor, T-bet, is heterogenous within the group, its existence has been questioned [37,38,39], mainly by one study performing mass cytometry (CyTOF) analysis of 30 surface markers, which failed to identify

ILC1 as a separate cluster through t-distributed stochastic neighbor embedding analysis (*t-SNE*) [40]. In addition to this, single cell RNA sequencing (scRNAseq) [6] and surface protein staining [15] revealed expression of transcripts normally linked to T cells (i.e. CD4, CD5, CD8, CD28 and some CD3 chains) in ILC1, pointing toward the possibility of T cell contamination during ILC isolation/analysis. In contrast to this evidence, pushing ILC1 towards a T cell phenotype via cytokine treatment failed to up regulate surface CD3 or TCR expression [30], the defining markers of T cells. Finding a correct definition and understanding what makes up the ILC1 subset is still currently under research. Efforts in understanding how other transcription factors affect ILC1 development and differentiation might shed some light on this issue.

NK cells, as mentioned above, were the first innate lymphoid cells to be discovered [21], thus, this subset has undergone much scrutiny and is so far the most defined ILC. In addition to T-bet, NK cells depend on the transcription factor eomesodermin. They are generally divided into a $CD56^{bright}$ subset, abundant cytokine producers but only weakly cytotoxic before activation [41] and a $CD56^{dim}$ subset, that is mostly responsible for cytolytic activity and target cell killing [42]. Recently, additional NK subsets have been defined in peripheral tissues [43]. One of the features dividing NK cells from the rest of ILCs is their lack of high expression of the α chain of the IL-7 receptor (CD127) (although $CD56^{bright}$ express low levels), and their IL-7 independence for development.

NK cells are activated through their natural cytotoxicity receptors (NCR), which leads to viral antigen recognition [44] and tumor cell death [45]. In humans, NKp44 (encoded by *NCR2*), NKp46 (encoded by *NCR1*) and NKp30 (encoded by *NCR3*) are expressed on the surface of the cells, while mouse NK cells only express NKp46. IFN-γ is released by NK cells upon NKp30 engagement by dendritic cells, making DCs release IL-12, IL-15 and IL-18 cytokines, thus promoting a positive feedback loop between NK cells and DCs, which triggers an adaptive system activation via Th1 cells and ultimately stopping viral replication and initiating infected cell death [46]. Other receptors on NK cells surface, when activated, can trigger cytotoxic molecule release [47]. Virally infected or tumor aberrant cells upregulate the expression of MHC class 1 proteins with high affinity for the NK cell activating receptor NKG2D, thus initiating target cell killing [48,49]. Perforin molecules create a pore on the target cell that facilitates the entry of granzymes into the target cell's cytoplasma, where they cleave caspases and other molecules, causing cell death [50].

### 1.1.2 ILC2

One of the first proofs of an innate cell with type 2 cytokine capabilities was reported in a study published in 2001, where IL-25-treated $Rag2^{-/-}$ mice were shown to produce IL-13 and IL-5 [51]. Since such mice lacked the machinery to create mature T and B cells, the only way for cells to react to an IL-25 stimulus was through the innate system. Other studies found that similar models lacking T and B cells cleared *Nippostrongylus brasiliens* infection via IL-13 [52,53]. In 2010, three key studies identified innate cells capable of type 2 responses in a variety of tissues in mouse, and they coined such population nuocytes, after the $13^{th}$ letter of the greek alphabet (*nu*), since they wanted to underline their IL-13 releasing capabilities [25,26,24]. Shortly after, the same cells were identified in human tissues and in circulation [14], finally

changing the nomenclature to its current accepted form of ILC2 [29].

ILC2 are dependent on the transcription factor Gata-binding protein (GATA3) [54]. In the mouse, the retinoic acid receptor RAR-related orphan receptor α (RORα) [55] is also needed for ILC2 development, while in human this is still under debate. As with all ILCs, they lack lineage markers and express IL-7Rα, and they are defined by a combined expression of the prostaglandin $D_2$ receptor CRTH2, CD25 and CD161 proteins in humans [14], and CD90, SCA-1, ICOS and KLRG1 in mice [25,26,24]. They produce type 2 cytokines IL-4, IL-5, IL-9 and IL-13, but also IL-6, IL-8 and GM-CSF. In the mouse, amphiregulin (AREG) has also been detected [56,57]. ILC2 are activated by soluble factors like lipid mediators [58,59], cell surface ligands (CD30, ICOS) and cytokines like IL-25 (also called IL-17E), TSLP and IL-33, released by multiple cells. Epithelial and mast cells can produce both IL-25 and IL-33, with eosinophils being IL-25 producers and fibroblasts, dendritic cells, macrophages, osteoblasts and endothelial cells being IL-33 producers. Basophils also release IL-4, which plays a role in maintaining ILC2 proliferation, and can activate ILC2 responses during cutaneous inflammation [60].

Activation of ILC2 by these cytokines causes a type 2 reaction by ILC2s, which in turn change the physiology of nearby cells. IL-4 and IL-13 are known to induce muscle contraction [61], needed in the event of a parasite infection. IL-13 also triggers tissue repair, both directly, by affecting the collagen production and deposition, and indirectly, by elevating mucus secretion by goblet cells, or promoting TGF-β mediated repair [62]. Moreover, during allergic lung inflammation, IL-13 is known to be responsible for inducing $CD40^+$ dendritic cell to migrate towards the lymph nodes, where they can present antigens to naïve T cells and initiate an adaptive reaction, consisting in Th2 cell differentiation [63].

IL-5 recruits and activates eosinophils, which attach to the surface of parasites [64] and release molecules toxic to the worm [65]. AREG is also expressed by ILC2 under IL-33 exposure [57,56]. AREG has important roles in the homeostasis of inflamed tissues and on the integrity of the epithelium under inflammatory conditions [66,67], through its capacity to activate macrophages (M2) that can limit acute inflammation and promote tissue repair through Arginase and IGF-1, an epithelial mitogen [68].

In mice two subgroups of ILC2 have been described. Steady state or natural ILC2s (nILC2), which reside in the lung and respond to IL-33, and inflammatory ILC2s (iILC2), that respond to IL-25 (but not IL-33) in multiple tissues, such as lung, spleen and liver. iILC2s are also detected in circulation, and are found in tissues upon inflammation[69], and, like nILC2s, they were found to clear *Nippostrongylus brasiliensis* parasite infection via IL-13, and protect against *Candida albicans* fungus infection via IL-17.

In human, so far there has not been any reports of iILC2-like cells, although it has been reported that a subgroup of ILC3 cells have the ability to co-express IL-17 [70] and IL-13/IL-5 [71].

### 1.1.3 ILC3 and LTi cells

In the late 2000's, a series of reports, first in mice [72,23] and then in humans [73] showed how NK cell receptor positive innate lymphocytes were able to produce IL-22, naming this particular subset NK22. Later studies were able to distinguish them from NK cells, changing the nomenclature to group 3 ILCs [29]. ILC3 are dependent on the RAR-RORγt transcription factor, maintained by IL-2 and IL-7 cytokines and are activated by IL-1β, and IL-23. Activation leads to the production of IL-22, IL-17A, IL-17F and granulocyte macrophage colony-stimulating factor (GM-CSF).

In an inflammatory context, activated dendritic cells and macrophages are known to release IL-1β and IL-23. The combination of these two cytokines has been shown to work particularly well in terms of IL-22 release by ILC3s [73,74]. IL-22 is known to enhance tissue repair and promote barrier immunity by decreasing bacterial threat. It aids host defense against extracellular pathogens by increasing the proliferation of epithelial cells, to strengthen the epithelial barrier upon pathogen attack. In collaboration with IL-17, it promotes production of defensins that have antimicrobial functions [75], and enhances goblet cells mucus production for extra barrier protection [76]. In mice, IL-22 was shown to be effective in clearing the *C. rodentium* pathogen [77] while ILC3 produced IL-17 provided protection following *Candida albicans* infection [78].

So far, ILC3 represent the most diverse ILC subset. In mice, they can be subdivided into $NKp46^+$ and $NKp46^-$ subsets, while in humans the same subsets are distinguished by the presence or absence of NKp44 [79]. $NKp44^+$ ILC3s are known to produce the most IL-22 and are also marked by the expression of Helios (encoded by *IKZF2*), a member of the Ikaros family of transcription factors. $NKp44^-$ are the predominant IL-17 producers, and in a recent study they were shown to include ILC progenitors (ILCp), able to extravasate from the circulation, home to tissues and start differentiating into different ILC groups [80]. Moreover, a study looking at single cell RNA transcripts in human ILC has revealed additional heterogeneity [6]. It shows the presence of an ILC3 subgroup with antigen presenting capabilities expressing HLA-DR and HLA-DP transcripts, and a subgroup expressing CD62L and CD45RA, both markers of naivety, matching the phenotype of the aforementioned ILCp [6].

Another subset that also expresses RORγt and releases IL-22 is lymphoid tissue inducer cells (LTi), however, at least in mice, they develop from a progenitor distinct from the ILCp, thus, they are regarded as a distinct ILC subset [27,17]. They were first discovered in the late 90's and named after their function in organogenesis of secondary lymphoid structures [22]. They play a role during embryogenesis through the release of TNF-α and lymphotoxin-β [81]. Recently, neuropilin-1 (encoded by *NRP1*, also known as CD304 or BDCA4) was found on human cells with LTi function [82], and also in the transcripts of LTi cells isolated from mouse [83], indicating that NRP1 is a conserved marker defining this ILC group.

## 1.2 DEVELOPMENT OF HUMAN INNATE LYMPHOID CELLS

Extensive studies have been performed in the recent years in order to understand the origin of innate lymphoid cells, especially in mice. ILCs originate from the common lymphoid progenitor (CLP), found in the bone marrow. Expression of multiple transcription factors, like the nuclear factor, IL-3 regulated (NFIL3) [84], the thymocyte selection associated high mobility group box (TOX) [85], the inhibitor of DNA binding 2 (ID2) [86], T-cell factor 1 (TCF1) [87] and the ETS proto-Oncogene 1 (ETS1) [88], lead to the formation of the common innate lymphoid progenitor (CILP). NK cell precursors (NKp) can differentiate from the CILP to give rise to NK cells, through the action of T-bet and eomesodermin [89].

Thanks to GATA binding protein 3 (GATA3) expression, the CILP develops into the common helper innate lymphoid progenitor (CHILP) [90], which is capable of differentiating into either the aforementioned ILCp (via zinc finger protein PLZF expression [91]) [28] or the LTI progenitor (LTIp). LTi cells need RORγt, TOX and ID2 to transition into their mature form, to then perform their organogenesis function.

In humans, the intermediate cellular stages of ILC development are not yet defined completely. A RORγt$^+$CD34$^+$ hematopoietic progenitor cell found in tonsils and the intestinal lamina propria has been linked to ILC3 development but not NK cell development [92], while a similar population of CD34$^+$CD45RA$^+$CD117$^+$ found in secondary lymphoid tissues was shown to be capable of maturing into all ILC subsets including NK cells [93], thus sharing more similarities with the CHILP. A recent study also showed how CD117$^+$ ILCs found in the blood harbor ILCp derived from hematopoietic stem cells, and that they can migrate to tissues to generate all ILC subsets, including NK cells (**Figure 2**) [80]. These studies underline the possible difference between human and mouse ILC biology, since in mice it is currently believed that ILCs are tissue-resident cells generated from local precursors [94]. More studies are needed in order to further elucidate developmental stages, migration and tissue residency of ILCs in both human and mouse.

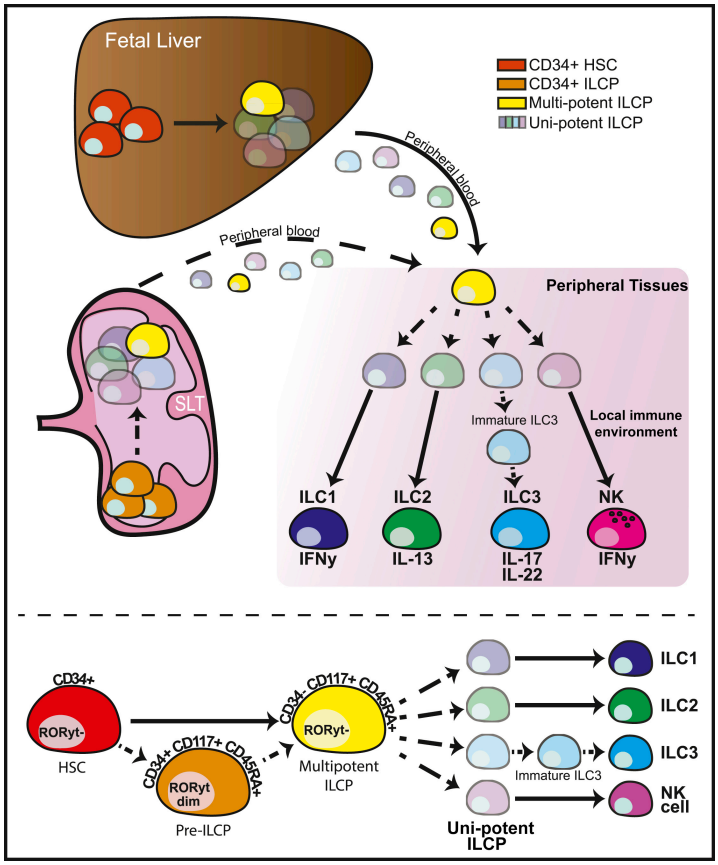


**Figure 2**. Model for differentiation, tissue distribution and migration of ILCp. HSC: hematopoetic stem cells. ILCp: ILC precursors. (Top) $CD34^+$ HSC and $CD34^+$ ILCp are present in fetal liver and other secondary lymphoid organs. These cells mature into $CD34^-$ multi-potent ILCp that migrate to the peripheral blood to then seed tissues and mature into ILC. (Bottom) $CD34^+$ HSC give rise to multipotent ILCp directly or through a $CD34^+RORγt^{dim}$ intermediate step. Reproduced from Mjösberg J., Mazzurana L. *Immunity*. 2018

## 1.3 PLASTICITY IN INNATE LYMPHOID CELLS

Human ILCs are naturally very plastic; they have the capacity to change their phenotype and function upon cytokine stimulus, thereby adapting to environmental stimuli. In this section I will give an overview of such trans-differentiation capacity. **Figure 3** gives a graphical overview of the plastic combinations ILC can achieve.

*ILC1-ILC3 plasticity*

In the context of inflammatory bowel disease (IBD), tissues in the gut can be extremely inflamed, to the point that surgery is the only possible solution to make the patient feel better, by removing inflamed sections of the intestine. In this highly inflamed environment, studies have shown that the ILC frequencies do not change, but the proportion of ILC1 in the lamina propria of Crohn's disease patients does, increasing substantially [30]. In such an environment, $CD14^+$ DCs are found to be releasing high amounts of IL-12, which binds to cells expressing IL-12 receptors, including ILC3. When stimulated with IL-12, ILC3 are found to downregulate their RORγt expression and to upregulate T-bet expression, the TF expressed mainly by type 1 cytokine producer cells. They also activate the cellular machinery necessary

for IFN-γ production. In culture, it has been shown that after 4 days of ILC3 stimulation with IL-12, the cells change their phenotype, losing CD117 and NKp44, and taking on an ILC1 (or ex-ILC3-like) phenotype [95]. *In vitro*, this process is reversible by stimulating "ex-ILC3s" with cytokines promoting RORγt expression, like IL-23 and IL-1β. This increases RORγt expression while decreasing T-bet, and re-activates IL-22 production, while shutting down IFN-γ release. This also works on freshly isolated ILC1 [95].

We explored the transcriptional mechanisms underpinning ILC3-ILC1 plasticity in **Paper I**, with tonsillar ILC3 transdifferentiating towards ILC1-like cells under IL-12 IL-1β stimulus *in vitro*.

*ILC2-ILC1 plasticity*

As is the case with ILC3-ILC1 differentiation, ILC2-ILC1 differentiation also occurs in an inflammatory environment. Inflamed lung provides the right environment for this, especially in the case of chronic obstructive pulmonary disease (COPD), a condition characterized by chronic inflammation [96]. ILC2 stimulated with IL-12 and IL-1β upregulate T-bet and become $CD117^-$ and lose CRTH2, acquiring an ILC1 (or "ex-ILC2") phenotype [97,98,99].

A recent study showed how IL-1β primes ILC2 for this process, mainly by increasing the expression of IL-12 receptor (IL12RB2). Changing the chromatin landscape also opens the locus required for IFN-γ transcription [100]. IL-4 is shown to halt or reverse this effect, differentiating the "ex-ILC2" to the original ILC2 phenotype [97].

*ILC2-ILC3 plasticity*

In human, recent studies report the transdifferentiation of nasal polyp ILC2 into IL-17 producing ILC3-like cells. This process is triggered when cells are under the influence of TGF-β, IL-23 and IL-1β stimuli, cytokines released in the microenvironment upon microbacteria infection, like *Staphylococcus Aureus* and *Pseudonomas aeruginosa.* This plastic behavior of ILC2 explains the accumulation of ILC3 in the nasal polyps of patient suffering from cystic fibrosis [70].

In mice studies showing how iILC2 found in lungs and expressing intermediate amounts of RORγt can lose IL-13 production and adopt IL-17 producing capabilities upon *C.albicans* infection, or stimulated *in vitro* by a IL-2, IL-7, IL-1β, IL-23 and TGF-β cocktail. [69]

*NK cell-ILC1 plasticity*

Recent studies on mice have explored the possibility of NK cell differentiation to ILC1. It was shown that both in vivo and in vitro NK cells could differentiate into an ILC1-like phenotype by TGF- β stimulus [101]. This pathway is controlled by SMAD4, a signal transducer that facilitates signaling pathway of TGF- β family cytokines. NK cells lacking SMAD4 lost their ability to produce cytotoxic molecules, thus losing their tumor killing capabilities [102]. The plasticity between these two cell subsets is still to be addressed in humans.

*ILC3-NK plasticity*

In human, one report described the ability of tonsillar ILC3 to transdifferentiate into NK cells. This transition is promoted by IL-12 and IL-15 stimuli, and triggers CD94, Eomes and T-bet upregulation. Cytokine release is also triggered, with TNF-α, IFN-γ, granzymes and perforin release, efficiently being able to target and kill K-562 cell lines [103]. This report was in part confirmed by results in **Paper I**, with NK cell like phenotype obtained upon IL-12 and IL-1β driven transdifferantiation of ILC3 towards ILC1/NK-like cellular status.

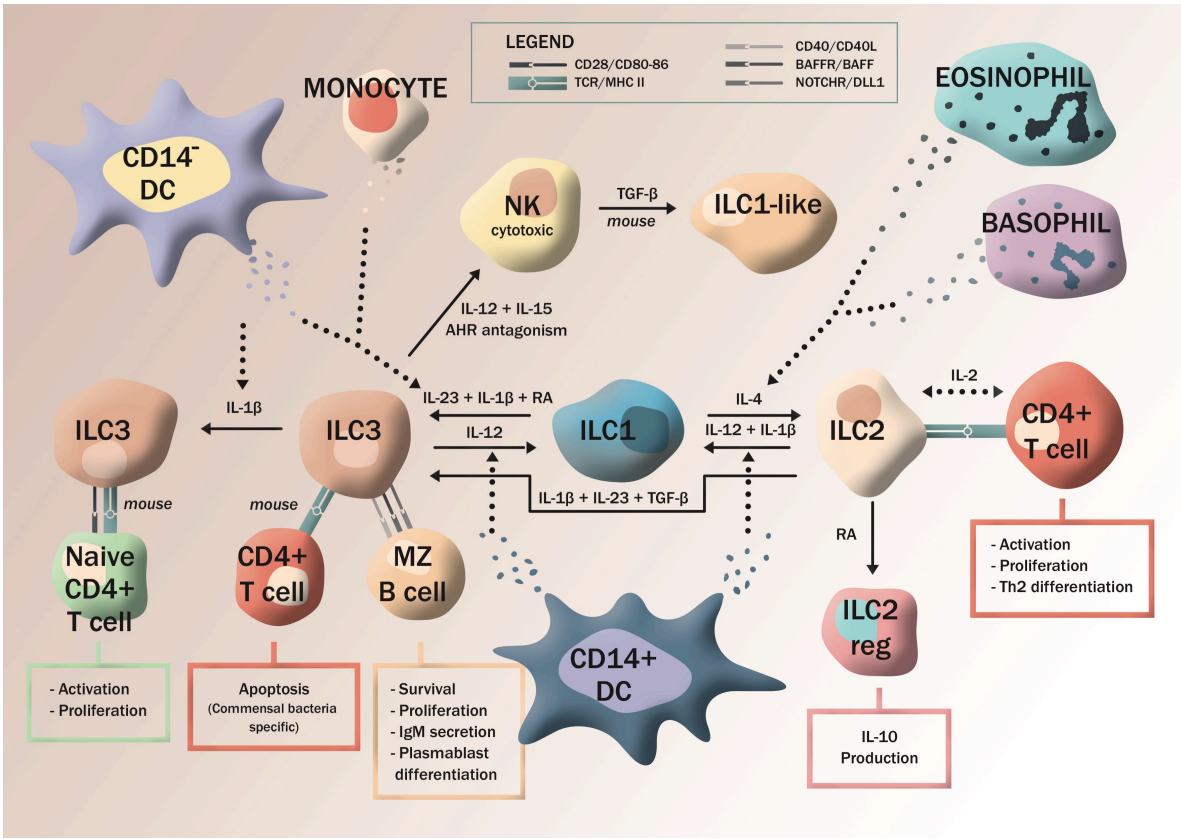


**Figure 3**. A graphical overview of human ILC plasticity and interaction with other cells. MZ: marginal zone; DC: dendritic cell; Human ILC respond to a variety of cytokines in their microenvironment released by neighboring cells, which trigger ILC trans-differentiation. ILC can interact with several other cell types influencing their phenotype and function. Reproduced from Mazzurana et. al. 2018. *Semin Immunopathol* [104]

# 2 INFLAMMATORY BOWEL DISEASE

Inflammatory bowel disease (IBD) is a complex disorder resulting from the interaction of genetic and environmental factors, priming the gut immune response to malfunction [105]. In **Paper I** we studied intestinal biopsies from IBD patients, while in **Paper II** we analyzed a large cohort of blood samples from IBD patients and compared cell type frequencies to samples from healthy blood. Here, I describe the pathogenesis of this disease and depict what role ILC play in IBD.

What characterizes IBD is a chronic inflammation of areas of the gastrointestinal (GI) tract, and can manifest itself as Ulcerative Colitis (UC) or Crohn's disease (CD). A non-patchy inflammation limited to the large intestine, the colon and the rectum, characterizes UC. The inflammation in UC tends to affect the mucosal and sub mucosa layers of the colon, rich in lymphocytes. A patchy inflammation, that may affect any part of the GI tract, and even extra-intestinal sites, characterizes CD. In **Paper II**, the majority of the CD cohort exhibited ileocolonic inflammation, with the rest of the cohort having inflammation restricted to either the ileum or the colon. In contrast to UC, the inflammation in CD can be so severe as to be transmural, affecting all layers of the GI wall [105]. The impaired barrier function leads to a translocation of commensal bacteria from the lumen to the submucosa, exacerbating inflammation [106]. This inflammation is caused by the influx of activated cytokine-producing macrophages and neutrophils to the area, together with eicosanoids, free radicals and proteolytic enzymes [107], leading to the common symptoms of IBD, namely abdominal pain, rectal blood loss, diarrhea and weight loss [107,108]. Regulatory mechanisms to contain this acute immune response fail to function properly, aggravating the state from acute to chronic. Another common characteristic of IBD is its relapsing-remitting course, especially in young adults, contributing heavily to the reduction in the quality of life of individuals affected by it [109,110].

The cause for IBD has been linked to a variety of factors, one of which is genetics. Genome-wide associated studies (GWAS) have analyzed the genetic information of large IBD cohorts and compared that to the one of healthy cohorts [111,112,113,114,115]. Many differences in genomic composition have thus been linked to this disease, with several publications connecting the *in silico* results from GWAS to mechanistical confirmation *in vivo* or *in vitro* [116,117].

Environmental factors are another quite broad feature that has been linked to IBD [118]. For CD, factors like cigarette smoking [119,120], having had an appendectomy [121], dietary animal protein [122], NSAID [123] or antibiotic use [124], psychosocial stress [125], depression [126] and oral contraceptive use [127], have all been associated with an increased risk of contracting this disease. Dietary fiber [128] and Vitamin D [129] have been connected with a decreased risk. For UC, consumption of dietary arachidonic [130] and linoleic acid [131], NSAID use [123], psychosocial stress[132], depression [126] and postmenopausal hormone use [133], have been linked with a higher risk, while cigarette smoking [134] and early appendectomy (before the age of 20) [135] have both curiously been associated with a decreased risk.

As the immune system is directly or indirectly affected by both genetics and environmental factors, it is only logical to assume that the composition and function of immune cells is altered during IBD, as I will discuss in the next section, which focuses on ILC.

## 2.1 INNATE LYMPHOID CELLS IN IBD

As innate cells represent the first line of defense when the host is attacked, they play an important role in managing the invading pathogen, by sensing microbial antigens and organizing a first response through cytokine release. Antigen presenting cells later present the processed microbial antigen to the adaptive immune cells, which take over the immune response. In IBD, commensal bacteria leak to the submucosa, effectively initiating an immune response.

Several reports have shown NK cells to be present at higher frequencies in IBD patients in the colonic lamina propria compared to controls [136,137]. NK cells are known to produce IFN-γ when activated, a cytokine very efficient at recruiting macrophages (IFN-γ was originally called macrophage-activating factor), increasing their antimicrobial mechanisms and antigen presentation [138]. Upon IFN-γ activation macrophages release IL-12, priming naïve $CD4^+$ T cells for differentiation to one of their activated states, Th1 cells, which are in turn capable of releasing ulterior IFN-γ [139,140]. This vicious circle is exacerbated by the maturation of dendritic cells, also releasing IL-12 to assist Th1 trans-differentiation [141].

As mentioned in the chapter section about ILC plasticity, another innate lymphoid cell that is affected by IL-12 is ILC3. In humans, ILC3 are broadly divided into $NKp44^-$ and $NKp44^+$, with the negative subgroup being more naïve and the positive subgroup being IL-22 producers, and the major ILC3 subgroup present in a healthy intestine [142]. In IBD, $NKp44^+$ ILC3 have been found to trans-differentiate into an ILC1-like status, gaining the ability to secrete IFN-γ [95]. Validating this finding, IBD disease severity has been shown to be associated with an increased presence of ILC1 [143]. Two ILC1 subgroups have been found to be enriched in the inflamed intestine, one restricted to the lamina propria [30] and one enriched in the epithelial tissue (ieILC1) [35]. What distinguishes them is the presence of CD103 on ieILC1, which also lack CD127, with the latter marker present in lamina propria ILC1. ieILC1 are more similar to NK cells given the presence of CD94 and intracellular expression of *EOMES*, a transcription factor found in high amounts in NK cells. Moreover, ieILC1 release granzymes and perforin [35]. Both subgroups produce IFN-γ when activated. As IBD is characterized by a chronic immune response, the presence of an protracted elevated concentration of IFN-γ in the intestine, originating from ILC1, trans-differentiated $NKp44^+$ ILC, $CD4^+$ Th1 and NK cells, was found to negatively affect the intercellular tight junctions between intestinal epithelial cells, damaging the function of the epithelial barrier and effectively paving the path for pathogens to invade [144,145].

The role ILC2 have in IBD is unclear. There are a couple of studies showing their enrichment among both CD patients [146], and UC patients [142], although at very small frequencies. Thanks to their ability to release IL-5 and IL-13, they are very efficient at clearing helminth infections and mediate allergic asthma. As they also respond to IL-33 with amphiregulin (AREG) production, a mouse study showed how ILC2 can have a tissue-protective role in the intestine [147]. The cytokine IL-25 also activates them, triggering an IL-13 response thought to elicit goblet and tuft cells regeneration after an helminth infection in mouse [148]. Given the very small number of ILC2 found in inflamed IBD samples in human, it is difficult to pinpoint and describe their role and if the same mechanisms are as valid as in the mouse model.

All in all, ILC play an important role in the pathogenesis of IBD, especially due to their IFN-γ production capabilities in an inflamed environment. As they can sense and respond to cytokines in the environment they should be taken into consideration when developing novel therapeutic agents for IBD.

# 3 AIMS

The work included in this thesis aimed to gain insight into several aspects of ILC biology in homeostasis, an *in vitro* inflammatory environment and in disease.

Specifically, the three different studies aimed to:

- **Paper I**: Determine the involvement of the transcription factor Aiolos (encoded by IKZF3) in ILC3/1 transdifferentiation, and to assess its role in ILC1 function through gene silencing.
- **Paper II**: Analyze peripheral ILC disturbance in a cohort of IBD samples compared to samples at homeostasis, focusing on proteins encoded by IBD risk genes as assessed in published GWAS.
- **Paper III**: Transcriptionally profile ILCs using single-cell RNA sequencing via Smart-seq2 technology, in order to delineate heterogeneity, identify potential marker proteins for ILCs and compare transcriptional profiles between peripheral blood, mucosal and lymphoid tissues.

# 4 METHODOLOGICAL CONSIDERATIONS

A variety of techniques have been used during my research, both *in vitro, ex vivo* and *in silico*. Here I will focus on flow cytometry, the technological backbone to all papers included in this thesis, and gene expression analysis by single cell RNA sequencing, a key technology used in **Paper III**. Moreover, as all methods and technologies were applied to precious human samples, I will discuss sample collection as well as ethical considerations.

## 4.1 ETHICAL CONSIDERATIONS

All studies included in this thesis used human material as the primary data source. All studies were performed in accordance with ethical guidelines described in the Declaration of Helsinki. Informed consent was obtained for all studies, from either the patients directly or from legal guardians in the case of children prior to obtaining the sample. Relevant clinical information, i.e. gender, age, medical history, was shared with the researcher in order for more specific data analysis to be performed.

For **Paper I**, we received tonsils, blood and gut biopsies. The tonsils were obtained from patients who suffered from obstructive sleep apnea. We receive the tonsils after the tonsillectomy. Hence, removal of tonsils was part of the clinical procedure and did not impose any further intervention for the patient.

The most ethically delicate tissue used were gut biopsies, also in **Paper I** and **Paper III**. They were retrieved from patients undergoing colonoscopy, consequently prolonging the procedure since doctors needed to first collect biopsies for clinical matters, and then collect tissues for research, thus lengthening the colonoscopy procedure by several minutes. Also, before performing the colonoscopy, blood was obtained from the patients for clinical matters and research. This process was regarded as safe and of little pain to the patients.

In **Paper III** lung tissue was collected from patients undergoing surgical lung resection, mainly due to cancer. Non-affected tissue was used for the project. No extra tissue was surgically removed in order to accommodate our study requirements; hence participation in the study did not entail further intervention.

The additional risks endured by the patient is considered reasonable and justifiable by the gain of new research being performed on their specific diseases, hopefully leading to novel and more efficient medications and treatments.

The studies included in this thesis were approved by the respective regional ethic committees as per the following permit number: Dnr 2016/1415-32, Dnr 2017/2404-32, Dnr 2017/1659-32, Dnr 2016/128-32 (**Paper I** and **III**), Dnr 2018/1819-31/1 (**Paper III**) and Dnr 2010/313 (**Paper II**).

## 4.2 FLOW CYTOMETRY

As the name of this particular method suggests (cyto = cell, metry = measurement), a flow cytometer measures single cells in a fluidic system. The basic principle consists in pointing a laser beam to the flowing stream of cells, and through the help of strategically placed detectors, disturbances in the beam of light are measured when each cell passes through the beam, giving information about the cell's size and granularity. This basic information can already give the immunologist an idea of what kind of cells are being analyzed, given that lymphocytes, monocytes and neutrophils are easily distinguishable by size and granularity.

Modern cytometers are able to gather multiple data points on each cell, by being equipped with special detectors measuring fluorescence. Fluorescent compounds within or on top of a cell's membrane can be excited when hit by a laser beam, emitting light that can be measured by the detectors. There are a wide variety of commercially available fluorochromes that are conjugated to antibodies specific to a surface antigen or intracellular proteins. The only limitation to how many different fluorochromes can be used in one experiment is the number of laser excitation sources available in the cytometer. The main cytometers used to gather the data presented in this thesis were equipped with four lasers (blue, red, yellow-green and UV), enabling detection of up to 18 colors.

In addition to the capabilities of flow cytometers to gather data on the size, granularity, presence or absence of a protein and its level of expression, fluorescence-activated cell sorting (FACS) offers the ability to sort live cells into a specific tube or the well of a plate based on the cytometry data. FACS enables the researcher to isolate a specific cell of interest from a sample, without harming the cell. The use of FACS technology is key in ILC research, as ILC are a rare lymphocyte population that often needs enrichment in order for the researcher to generate meaningful data. FACS was used to isolate ILC mainly from tonsils in **Paper I**, while in **Paper III** a considerable effort was made to set up ILC single cell sorting in 384 well plates from a variety of tissues.

Data acquired from flow cytometers was analyzed with the FlowJo version 9 or 10 software and SPICE version 5 or 6. Unbiased analysis (i.e. not gate-based) was done through R software (various versions), with stochastic neighbor embedding analysis (SNE) or uniform manifold approximation and projection (UMAP) methodology being implemented in order to reduce the multidimensionality of the data [149].

## 4.3 GENE EXPRESSION ANALYSIS

During the past two decades, the way researchers approached gene expression analysis has varied substantially. In **Paper I**, we used one of the earlier methods to assess gene expression differences, namely microarrays. This technology is based on the hybridization of synthetic oligo probes to target strands of cDNA, resulting in fluoroscence. By scanning the chip using high-throughput screening, the fluorescence is quantified, giving the researcher reliable data

on the levels of gene expression of the sample analyzed. Although microarrays are relatively inexpensive and very successful in analyzing large amounts of biological samples in a limited amount of time, they also come with limitations. These limitations vary, from high background levels due to cross-hybridization [150], to reliance of existing knowledge about the genome sequence. In order to overcome such limitations, RNA sequencing was developed.

RNA sequencing (RNA-Seq) is a method using next-generation sequencing (NGS) technology to measure the presence and the amount of RNA transcripts of a biological sample at a given time. It is a powerful method that replaced microarray analysis in the mid-late '00s, as it made it possible to fully sequence the whole transcriptome while microarrays only offered limited profiling of pre-defined transcripts [151]. RNA-seq measures the average expression level of each gene across a large population of cells, giving the researcher a great tool for comparative transcriptomics (when analyzing samples from the same tissue, or a specific type of cell, in different species). The single-base resolution of this method is a big leap from the microarray standard, which had a resolution limited by the size of the probes in the chip, ranging from several base pairs to 100 base pairs. [152]. Additionally, RNA-seq does not have an upper limit when it comes to the quantification of gene expression level, which is simply equal to the number of transcripts present in the sample for a specific gene. This quantification ranges from a few transcripts up to thousand-folds. The high dynamic range is far more sensitive than microarrays, which have ranges spanning from one-hundredfold to a few hundredfold, making it hard for microarrays to detect genes that have a very low or very high expression [152].

A significant drawback of RNA-seq is its inability to provide insights into the heterogenous nature of gene expression. In other words, it does not tell the researcher the level of gene expression of every cell, since that can be differentially expressed within the pool of cells used as input for RNA-seq, but only shows the average level of expression of such pool. This is important information for a researcher who wants to analyze the heterogeneity of a system, like in the case of developmental studies. In order to overcome these limitations, a method taking into account not pools of cells but single cells was needed.

In the '90s, a number of techniques that analyzed gene expression levels in single cells were invented, such as fluorescent reporter constructs [153], single molecule RNA FISH [154], and later quantitative PCR [155]. Although these techniques lead to important discoveries, they were very low-throughput, analyzed a small number of single cells and a limited number of genes. In 2009, Tang et al developed the first method that allowed to perform RNA-seq on single cells [156], paving the path for the single cell sequencing revolution still ongoing today.

### 4.3.1 Single cell sequencing – Smart-Seq 2 technology

During the last 13 years, single cell throughput has increased substantially (**Figure 4**). In the first single cell RNA study, Tang et al picked each cell under a microscope before lysing it and performing the downstream protocol. The manual process of picking single cells by

micropipettes required incredible work, and the data obtained by a few single cells was quite limited. This limitation slowed the widespread use of this technology considerably. New methods increasing the throughput were developed in 2011, when multiplexing was introduced by the Linnarson lab [157]. It brought the number of single cells analyzed to close to 100, but in 2013 [158] and later in 2014 [159] the use of integrated fluidic circuits and liquid handling robotics made it even easier to capture single cells, triggering an extensive use of scRNA analysis within immunological research. In **Paper III**, through FACS sorting, we were able to isolate thousands of cells. Nowadays, the uprise of automated capture systems in nanoliter-sized aqueous droplets [160], picoliter chambers [161] or microwells have brought the reagent volumes and labour intensity to an even lower threshold, while raising the throughput to the tens of thousands of cells.

The majority of scRNA-seq library preparation protocols share a general workflow, although the chemistry used has changed throughout the years (**Figure 4**). In broad terms, after cell isolation with one of the methods discussed above, the cell is lysed, its mRNA is reverse transcribed, tagged with a barcode or cell identifier, and the cDNA of all the cells is then pooled, amplified and finally sequenced.

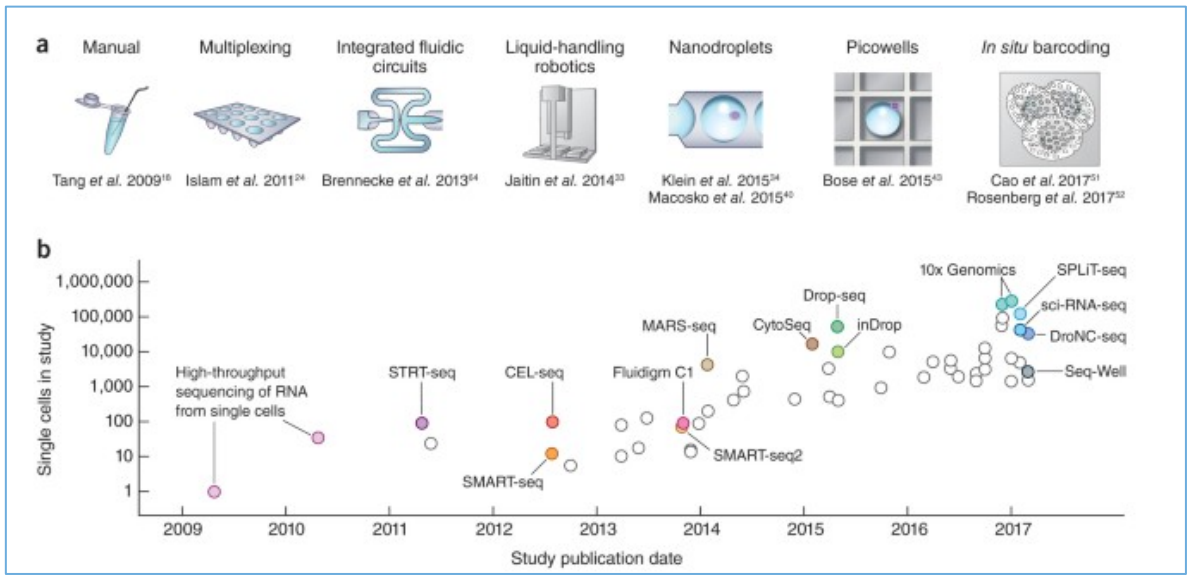


**Figure 4**. **A**: Depiction of several isolation methods for single cell retrieval in a chronological order. **B**: Timeline showing the progress and improvement in number of single cells isolated in studies ranging from 2009 to 2017. The main chemistry protocols used in the most important published studies are overlaid on the timeline. Adapted by permission from Macmillan Publishers Ltd: Springer Nature, Nature Protocols, Svensson v. et al., copyright 2018. [162]

In **Paper III**, our collaborators used Smart-seq2 methodology to prepare libraries, which allowed full-length transcriptome sequencing, something that is not possible in tag-based quantification scRNA library protocols. As the name suggests, full-length sequencing allows a uniform read coverage of the transcripts, while tag-based methods capture either the 3' or 5' end of the transcript. A recent publication comparing the six prominent scRNA library

preparation methods (CEL-seq2/C1 [163], Drop-seq [160], MARS-seq [159], SCRB-seq [164], Smart-seq/C1 [165], Smart-seq2 [166,167]) , showed how Smart-Seq2 ranked highest in detected genes per cell (highest sensitivity), and had the highest accuracy of transcript level quantification [168].

As our goal was to characterize a rare population of cells such as ILC, maximizing the gene detection rate was our primary interest. As Smart-Seq2 was key in achieving that goal, a brief description of the protocol is shown in **Figure 5**. Briefly, the isolated cells are lysed in a buffer containing dNTP and oligo DT primers with a known anchor sequence at their 5' end, which prime reverse transcription of the first strand. For the second strand synthesis, template switching is performed, which requires the terminal transferase activity of Moloney murine leukemia virus (MMLV) reverse transcriptase. MMLV transcriptase adds several cytosines to the 3' end of the first strand to act as a binding partner to the template-switching oligos (TSO) containing a locked nucleic acid (LNA; improves the thermal stability of the TSO), that effectively primes the synthesis of the second strand of cDNA. After reverse transcription, all cDNA strands share the same universal anchor sequence (ISPCR) and can undergo PCR amplification. For conversion into sequencing libraries, the amplicons are simultaneously fragmented by Tn5 transposase, and tagged with adaptor sequences that are then used as binding sites for the concluding enrichment PCR. The final libraries can then be pooled and sequenced [166].

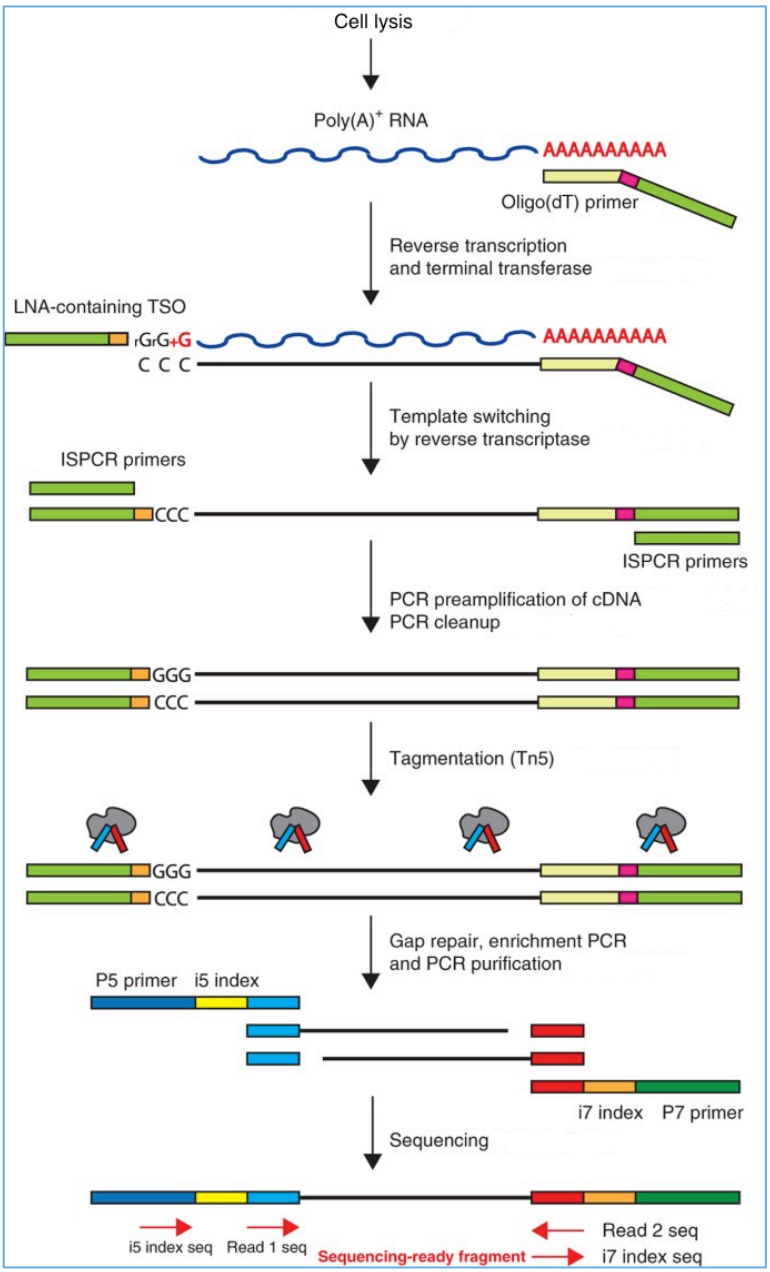


**Figure 5**. Outline of the Smart-seq2 protocol and the corresponding procedure steps. Adapted by permission from Macmillan Publishers Ltd: Springer Nature, Nature Protocols, Full-length RNA-seq from single cells using Smart-seq2. Picelli, S. et al., copyright 2014. [167]

# 5 RESULTS AND DISCUSSION

## 5.1 PAPER I - AIOLOS AND IKAROS AND THEIR ROLE IN ILC3 TO ILC1/NK TRANSDIFFERENTIATION

The work presented in **Paper I** stems from a study previously published by our group [6], which focused on unraveling heterogeneity of human tonsil ILC by scRNA-seq. One transcript that caught our attention was a transcription factor from the Ikaros transcription factor family, called Aiolos, encoded by *IKZF3*.

*The Ikaros family of transcription factors*

The Ikaros family of transcription factors was discovered throughout the '90s by a greek researcher, who named these zinc finger DNA binding proteins after characters from greek mythology. They function as key regulators of lymphocyte specification, expressed in HSC at an early developmental stage [169]. The family is composed of five members, Ikaros (encoded by *IKZF1*), Helios (encoded by *IKZF2*), Aiolos (encoded by *IKZF3*), Eos (encoded by *IKZF4*) and Pegasus (encoded by *IKZF5*). Ikaros is a transcriptional regulator that controls differentiation into the T cell lineage and B cell lineage [170]. Mice homozygous for the deletion of the Ikaros DNA-binding domain lack CLP, and consequently T, B and NK cells[171]. Aiolos was identified as a homolog of Ikaros, expressed in committed lymphoid progenitors, and its expression increases throughout T and B cell development [172]. More recently Aiolos has been linked to Th17 cell differentiation and regulation of NK cell maturation [173,174]. Ikaros and Aiolos interact with each other via heterodimerization. Another factor that can heterodimerize with both Aiolos and Ikaros is Helios, which is detected in the earliest hematopoietic sites of the embryo, in adult HSC and is later present in specific T cell subsets [175]. Helios has also been linked to regulatory T cell differentiation [176,177]. The Helios protein shows a high degree of conservation to the Ikaros protein (73% overall similarity) and the Aiolos protein (67% similarity) [175]. Eos is most highly related to Helios, and it is involved in osteoclast differentiation and muscle development [178,179]. Pegasus is the newest members of the family and limited research has been done to understand its function. [180].

As shown in **figure 6**, Ikaros is expressed more or less homogeneously by all human tonsil ILC, while Helios is mostly found on ILC3. What caught our attention was the high expression of Aiolos in tonsil ILC1. As ILC1 lack a marker that defines them, and *TBX21* (encoding for T-bet) transcripts in tonsils were found to be very low in ILC1 by scRNA-seq analysis, the Aiolos transcription factor seemed like the perfect candidate for a more in-depth analysis.

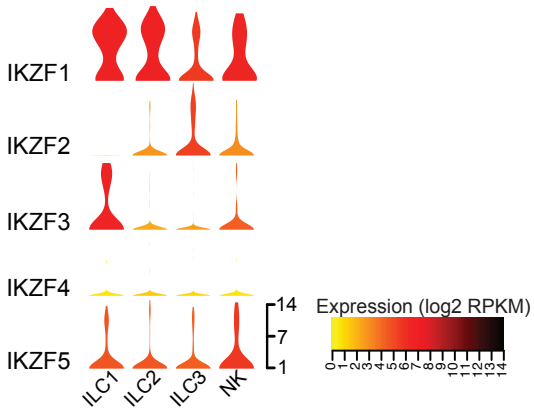


**Figure 6**. Ikaros family transcripts expression in human tonsil ILC. Violin plots show expression distribution across ILC subgroups. Adapted from Mazzurana et al. *EJI* 2019.

We validated the sequencing findings on blood and tonsil tissues on a protein level. Efforts were focused on Ikaros, Aiolos and Helios, as Eos and Pegasus showed a low homogeneous expression in ILC. We could confirm a high level of Ikaros expression by all ILC (NK cells included), expression of Helios in ILC3, with its highest expression in NKp44$^{+}$ ILC3, and Aiolos expression in ILC1 and NK cells. Overall, the TF pattern of blood and tonsils were very similar to each other.

As we targeted the analysis on ILC1, we co-stained T-bet and Eomes, two transcription factors known to be expressed in group 1 ILC, together with Aiolos. We could find a substantial amount of triple positive cells in our analysis of tonsil ILC1, even though every effort was done to exclude NK cell inclusion. The discovery of ILC1 expressing Eomes is in line with other reports showing Eomes$^{+}$T-bet$^{+}$ in blood, gut and liver ILC1 [142,10]. In the liver, the distinction between NK and ILC1 is even more challenging given the presence of a resident NK cell being Eomes$^{-}$T-bet$^{+}$, active in cytokine release and with a reduced cytotoxic capability [181]. The only distinction for ILC1 is its KIR and CD49a expression and the lack of CD127 [181]. The hunt to characterize a *bona fide* ILC1 given their confusing TF profile has been a topic of debate [40,182,38,37]. In **Paper III** we tried to answer this question by performing scRNA-seq and downstream unbiased analysis on a variety of tissues, where we could confirm the presence of an Eomes$^{+}$ ILC1 population at steady state mainly in blood, but also in tonsil, colon and lung (Paper III, Figure 1E and S3). This population was separate and distinct to NK cells and Eomes$^{-}$ ILC1, which could possibly contain *bona fide* ILC1.

In a disease state characterized by intestinal inflammation, ILC1 are known to increase in numbers, as shown in previous studies [30,95]. In **Paper I** (Paper I, Figure 2) we confirmed this finding by analyzing samples from patients affected with IBD, that were highly enriched in ILC1 compared to non-IBD samples or samples from a non-affected area. Aiolos was found to be expressed in ILC1 in both the inflamed and non inflamed samples.

Interestingly, around 20% of tonsil ILC1 were comprised of either Aiolos single positives or Aiolos$^{+}$ Tbet$^{+}$Eomes$^{-}$. Given Aiolos' role in gene regulation, we set out to learn more about its function. As working with ILC1 is extremely difficult given their limited number and their

fragile nature *in vitro*, we performed a series of experiments on tonsil ILC3 to better understand Aiolos' role in ILC3-ILC1 plasticity.

*Aiolos is upregulated during ILC3 to ILC1/NK trans-differentiation*

In **Paper III**, when examining ILC3 from tonsil tissue by scRNA-seq, we could distinguish two distinct clusters within tonsil ILC3 when analyzing the data through dimensionality reduction software (Paper III, Figure 1d, cluster 4 and 9). Through trajectory analysis we could categorize one subgroup as more activated, expressing transcripts such as *NCR2*, and the other cluster more naïve, expressing transcripts such as *SELL*, that we renamed ILC3a and ILC3b respectively. These findings were in line with previous reports on human NKp44$^{+/-}$ ILC3 [183].

In **Paper I**, using an *in vitro* system, we sorted and cultured tonsil ILC3 (comprising of ILC3a and ILC3b) and treated them with either IL-2 + IL-12 + IL-1β cytokine mix or IL-2 + IL-23 + IL-1β, artificially pushing the ILC3 towards an ILC1-like state or towards a more active ILC3 state respectively, as seen in **Figure 7**.

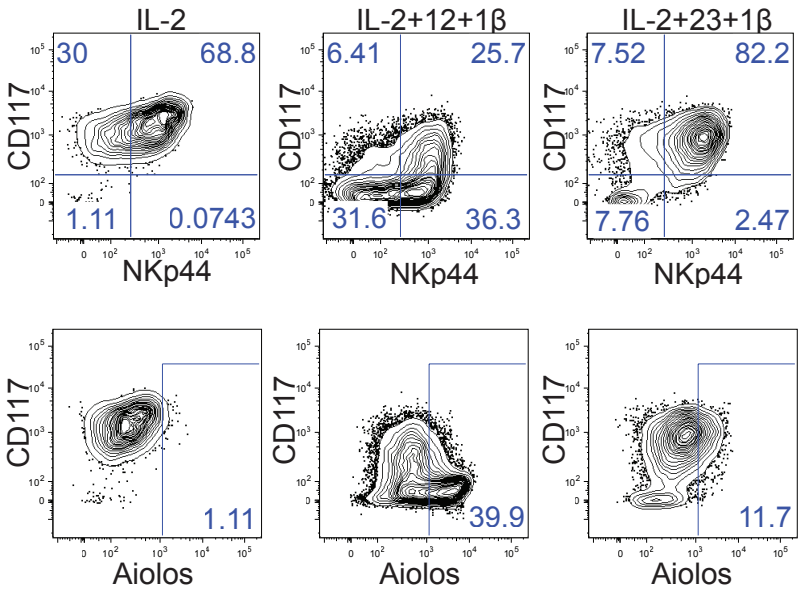


**Figure 7.** Representative plot showing ILC trans-differentiation and consequent Aiolos expression after an *in vitro* 4 to 6 day culture and cytokine treatment. Adapted from Mazzurana et al. *Eur J Imm*. 2019.

Aiolos was highly upregulated in the IL-2 + IL-12 + IL-1β condition compared to the control (IL-2 stimulation) or IL-2 + IL-23 + IL-1β condition. Ikaros and Helios were also homogeneously upregulated, maintaining a high expression in all conditions.

The cells generated by this culture were IFN-γ producers, and comprised of both ILC1 and NK cells, that we distinguished by the expression or lack thereof of the TF Eomes and the inhibitory NK protein NKG2A. Eomes$^-$ NKG2A$^-$ cells expressing Aiolos and/or T-bet, that also expressed CD161, fit with the phenotype of ILC1, while Eomes$^+$ NKG2A$^+$ and Eomes$^+$ NKG2A$^-$ cells fit the phenotype of NK cells.

The presence of NK-like cells after trans-differentiation culture was expected, given various reports of describing similar results. Raykova et al. used the cytokine mix IL-12 + IL-15 to trans-differentiate human ILC3 cells to NK cells or NK-like cells, both at the bulk and at the single cell level via cloning [103]. Even though they did not use IL-1β, they did culture tonsil ILC3 in a system using irradiated allogenic lymphocytes (also called feeder cells) for expansion. This artificial system has been reported to provide a source of IL-1β, together with other cytokines such as IFN-γ, IL-12, TNF-α, IL-6, IL-8 and IL-13, providing an additional inflammatory stimulus to the culture [97]. The NK cells produced in their culture displayed characteristics of early stage 4 NK cells, with the machinery necessary to produce and release cytotoxic molecules like granzymes and perforin [103]. In **Paper I**, both $NKp44^-$ and $NKp44^+$ ILC3 cultured with IL-2 IL-12 and IL-1B were capable of producing granzyme B and a small amount of perforin (data not shown).

Hughes et al. describe how another developmental intermediary of NK cells, known as "stage 3", characterized by a phenotype of $CD34^-CD117^+CD94^-$, could include ILC3 expressing IL1R1 and the aryl hydrocarbon receptor (AHR). AHR is a ligand-dependent transcription factor at the center of much research in the past decade [184,185,186]. It has been linked to the mediation of gut ILC2-ILC3 balance in the murine gut [187], ILC3/ILC1 balance in the human ileum in CD patients [188], Th17 differentiation [189], and Treg development [190]. By inhibiting AHR, they were able to differentiate $IL1R1^{hi}$ IL-22 producing ILC3 into $CD94^+Eomes^+T\text{-}bet^+$ cytolytic mature NK cells [191].

Given that tonsil ILC3 in **Paper I** were sorted as $Lineage^-CD127^+CD117^+CD94^-$, it is possible that some of these immature NK developmental intermediaries found their way in the ILC3 gate, and under an inflammatory stimulus of IL-12 and IL-1β they developed into NK cells. Furthermore, KLRC1 transcripts (encoding for NKG2A) were found in steady state tonsil ILC3 in **Paper III**, although not expressed at the surface. It is possible that the same cells under trans-differentiating stimuli would translate that transcript into protein, equipping the cell with an inhibitory protein.

*Aiolos and Ikaros silencing decreases ILC3 to ILC1/NK transdifferentiation and increases ILC3 activity*

As both Aiolos and Ikaros were highly upregulated in the transdifferentiation culture, we sought out to determine how the silencing of these two TF would affect the phenotype and function of the cells in culture. For this purpose, we treated the cells with Lenalidomide.

Lenalidomide is an immunomodulatory anticancer drug developed in the '90s. It is an analogue of Thalidomide, a drug that gained notoriety in the '60s after being used as a treatment of pregnancy related morning sickness, causing deformations in the fetus [192]. Lenalidomide's mode of action involves the selective targeting of both Aiolos and Ikaros for degradation, by triggering their ubiquitination via the CRBN-CRL4 ubiquitin ligase [193,194]. It works particularly well in treating multiple myeloma (MM), where both Aiolos and Ikaros play an important role in the disease. Both TF are regulators of NK- κB and indirectly

downregulate the TF MYC and the interferon regulatory factor 4 (IRF4), causing a lower cell cycle in the tumor cell together with an increased apoptosis [195].

Lenalidomide treatment of the ILC3 in culture under an IL-12 + IL-1β trans-differentiating stimulus significantly lowered Aiolos and Ikaros expression. T-bet followed in their path, suggesting that these three TFs might be co-regulated. A confirmation that T-bet and Aiolos cooperate in shutting down the ILC3 program was shown in a recent publication by Cella et al. via a chromatin analysis approach, where they described Aiolos and T-bet collaborating in the repression of regulatory elements such as IL-22 and RORγt active in ILC3. Moreover, our microarray on lenalidomide treated ILC3 cultures showed the upregulation of several ILC3 transcripts, such as *RORC*, *IL-22* and *NRP1*, going in the same direction as Cella et al.'s findings. Interestingly, Helios increased substantially, showing the inversely proportional nature between the three members of the Ikaros family, as was also seen in studies in mice, with IKZF3$^{-/-}$ mice expressing higher levels of IKZF2 transcript [174].

As graphically summarized in **Figure 8**, in **Paper I** we could successfully sort and culture tonsil ILC3, trans-differentiate them into ILC1/NK cells capable of releasing copious amounts of IFN-γ. Upon lenalidomide treatment, Aiolos, Ikaros and T-bet TFs were downregulated. This process triggered the upregulation of RORC and Helios, significantly limiting the ILC3 to ILC1/NK cell trans-differentiation and increasing the proportion of IL-22 releasing ILC3, following the decrease of the ILC1 program.

As ILC trans-differentiation takes place in conditions such as IBD, characterized by high and chronic inflammation in tissues rich of ILC3, more research is needed in regard to targeting such cells with compounds sharing lenalidomide's ability to downregulate key TFs, in order to limit trans-differentiation towards inflammatory cells and instead increase tissue protecting cytokines release such as IL-22.

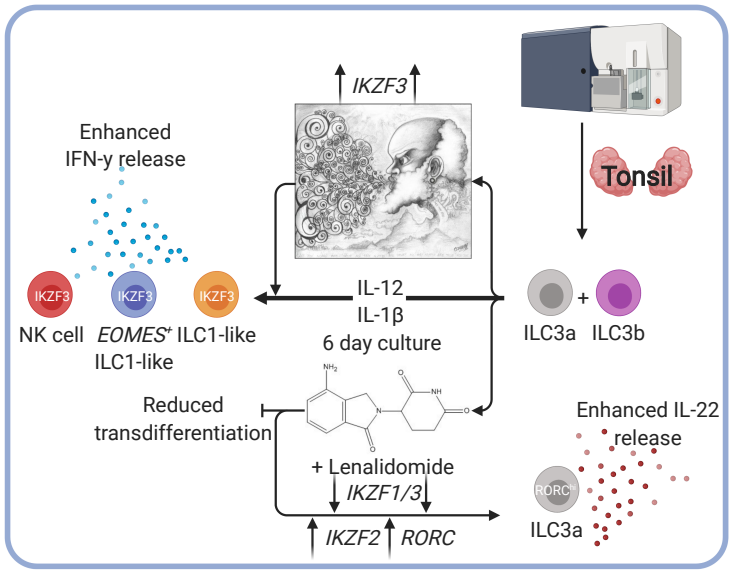


**Figure 8**. A graphical summary of the effect of Lenalidomide treatment to tonsil ILC3 undergoing trans-differentiation towards ILC1/NK cells under IL-12 ad IL-1β influence.

## 5.2 PAPER II - CIRCULATORY ILC IN IBD

When it comes to ILC in IBD, for obvious reasons the intestinal tract has always been the organ of primary focus, as the condition manifests itself primarily in that site. Several studies looked at ILC to identify disturbances in their composition at mucosal sites, most often by looking at patient's samples in the form of punch biopsies, obtained through endoscopic examinations, or resection material after surgical intervention. By focusing on the main organ of interest one leaves out how ILC in other tissues respond to the disorder, with only a few reports analyzing ILC in non-GI tissues, often lacking the power to reach statistical significance in their findings [142,196,197].

In **Paper II** we set out to perform a thorough analysis of ILC in peripheral blood in a large Swedish cohort of peripheral blood samples composed of 53 CD patients and 43 UC patients. To understand IBD specific disturbances in ILC composition, we compared the results to a cohort of 45 age and sex-matched healthy blood donors (HC).

*Exploiting IBD GWAS studies for the selection of ILC-IBD-associated markers*

We adopted a flow cytometry approach to identify possible protein expression differences in the cohort. In order to do that, an ad-hoc panel was created with standard ILC markers and markers for proteins known to be associated with IBD. Given the limited number of channels available on the FACS machine (at the time limited to 16 colors), we adopted an approach consisting in matching GWAS IBD risk genes with markers known to be present in ILC (at least as transcripts).

As briefly introduced in a previous chapter focusing on IBD, GWAS studies take genetic information of a very large cohort of patients, and compare that to the one of healthy controls. The goal of GWAS is to examine common genetic variants in the cohorts to find if any variants are associated with the disorder studied. The comparison of the genomes can lead to the discovery of single nucleotide polymorphism (SNP) associated with the disease of interest. Several GWAS studies analyzing data from IBD cohorts were successful in linking several genetic loci, dictated by interesting SNP locations, to the disease [111,112,113,114,115].

In order to deduce which genes made the most sense to include in the final flow cytometry panel, we extrapolated all genes included in the 201 loci associated with IBD described in the two main GWAS studies on IBD [111,112]. Each IBD risk-associated locus comprised of an average of seven genes per locus, with some loci encompassing only a few genes while some others including dozens of genes, given the linkage disequilibrium patterns in the human genome and the ~250kb span of each risk locus [198]. The consequent list of 1502 risk genes was then compared to the 757 differentially expressed genes in ILC stemming from our study looking at tonsil ILC via scRNA sequencing technology [6]. The resulting 83 matching genes were filtered through the Genotype-Tissue Expression project (GTeX) portal, a comprehensive public database that enables the researcher to study human tissue-specific gene expression and regulation and its relationship to genetic variations [199]. Through GTeX, we could filter the 83 candidate genes, with 20 of them resulting as residing in expression

quantitative trait loci (eQTL), meaning that the genetic variance in the IBD risk loci containing such genes can explain variation in gene expression level [200]. The 20 genes in eQTL were then filtered for level of expression, by setting a threshold to the reads per kilobase of transcript per million mapped reads (RPKM) of tonsil ILC transcripts, although our cohort and GTeX analysis was based on blood. Finally, only a dozen genes survived all steps of this stringent filtering, with only a few having a reliable antibody available for purchase. After this intricate selection process (simplified in **Figure 9**), two genes made the cut: SLAMF1 and HLA-DR.

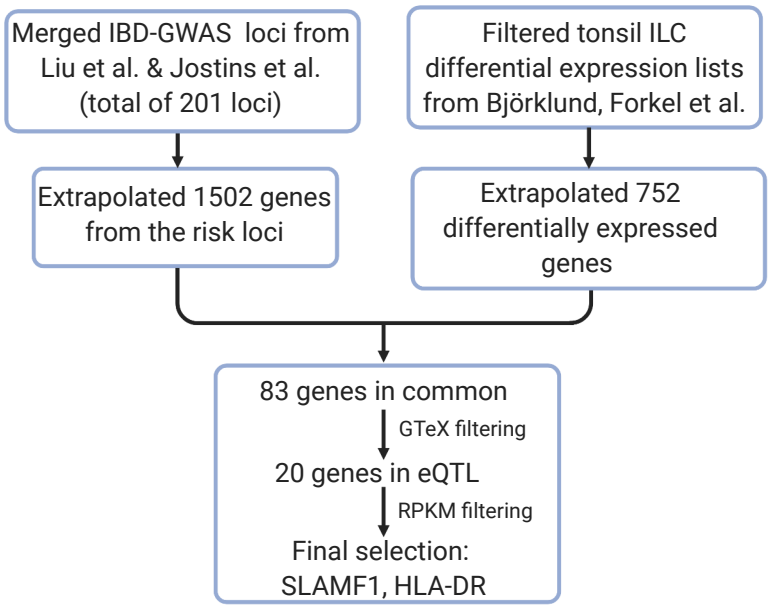


**Figure 9**. Rationale behind the IBD-associated marker selection for the flow cytometry panel.

*Characterization of circulating ILC through multi-color flow cytometry*

A classic and well-established ILC gating strategy was used to investigate ILC frequencies, with all ILC being Lineage$^-$CD127$^+$, and further subdivided into CD117$^+$ ILCp, CRTH2$^+$ ILC2 and CD117$^-$CRTH2$^-$ ILC1. Curiously, we could not find any differences in peripheral blood ILC frequencies between patients and control, as previously reported in similar studies [142]. In **Paper I**, we detected ILC1 enrichment in the inflamed gut biopsies of IBD patients, compared to controls, confirming previous reports [30,142]. Hence, it seems that blood ILC frequencies are not affected by the condition, although a recent study by Creyns et al. detected higher circulating ILC1 and ILC2 frequencies, and a lower ILCp frequency on a similar sized cohort of IBD patients and healthy donors to the one presented here [197]. The two cohorts are hard to compare given the patient's treatment status, with Creyns et al. focusing on patients on biologics (Ustekinumab, anti-TNF and Vedulizumab), while ours having only the minority of patients on biologic agents. Moreover, the inflammatory status of the IBD cohorts might differ, as our cohort had a lower average disease activity level for the UC group, as reflected by the Mayo score, and possibly even in the CD group, although the activity indexes adopted in the two papers differ.

As CD is a disease characterized by systemic inflammation, we were successful in discovering a number of activation markers expressed in ILCp from CD patients compared to HC, namely CD56 and NKp44, with HLA-DR additionally being expressed in ILC1 and ILC2. The increased frequency of expression of CD56 and NKp44 proteins was correlated in the CD cohort, hinting to a possible co-regulation. CD56 is a neural cell adhesion molecule encoded by the NCAM1 gene, and is one of the hallmark proteins present on the surface of NK cells, mediating their cytotoxicity and denoting their level of maturity given its level of expression [201,41]. HLA-DR, as shown in a recent study looking at human peripheral blood ILC, is upregulated during inflammation, being NF- kB dependent [202]. NKp44, encoded by NCR2, is an activating protein found on NK cells, belonging to the natural cytotoxicity receptor (NCR) family. In tissue, ILC3 are known to have a heterogeneous expression of NKp44, as we show in **Paper I** (Paper I, Figure 3A - tonsil) and **Paper III** (Paper III, Figure 3g – NCR2 transcripts in tonsil and gut ILC3 at steady state). In blood, at steady state there is no sign of NKp44 protein expression, as shown in the HC cohort (Paper II, Figure 2,4 and 5B). Its upregulation in circulatory ILCp has been reported on a smaller CD cohort (with two of seven CD samples having and enrichment in $NKp44^+$ ILCp) and in leukemia patients, where NKp44 expression was associated with a lower risk of graft-versus-host disease (GvHD) following hematopoietic stem cell transfer (HSCT) [203,196]. Creyns et al. could detect an enrichment of $NKp44^+$ ILCp in CD patients undergoing biologic treatments, especially in the anti-TNF treated subjects, and to a lesser extent in ustekinumab-treated subjects. In our cohort, none of the CD patients with an enriched circulating $NKp44^+$ ILCp was treated with biologics. Interestingly, in both active and remission UC patients, this population of cells was absent (Paper II, Figure 3,4 and 5B), as also reported by Creyns et al., making this finding, within IBD, CD specific. As this population of cells has the ability to produce IL-22, a cytokine known for its tissue repairing qualities and surely useful in a CD context, it would be interesting to find out what triggers their enrichment in circulation of CD patients. More research on this topic is needed.

Lim et al. recently described how blood $CD117^+CD45RA^+$ ILC represent a precursor population of ILC, that can then migrate to tissues and mature into their active form [80]. Given these findings, the marker used to distinguish naïve ILCp in **Paper II** was CD45RA, and the difference in their frequency ($CD45RA^+$ ILCp) was highest between CD and UC, as the UC cohort was mainly composed of patients in remission (36 patients with and average Mayo score of $0.6 \pm 0.7$) with only four in an active state (Mayo score $4 \pm 0.8$). With a higher level of activation in the CD samples, it was interesting, although somewhat expected, to see the frequency of naïve ILCp decrease, as it is most likely that ILCp matured into $NKp44^{+/-}CD56^+$ ILC3. To confirm the switch of naïve $CD45RA^+$ ILCp, a negative correlation to activated subsets such as $CD56^+$ ILC3 and ILC1 frequencies was described in CD patients (Paper II, Figure 3 and 5C).

The IBD-associated marker SLAMF1, found to be present in the ILC2 compartment, had an increased expression within both CD and UC cohorts as compared to HC (Paper II, Figure 2,4 and 5D). When correlating patient clinical information, it was observed that $SLAMF1^+$

ILC2 frequencies in CD patients with active disease were inversely correlated to HBI score (Paper II, Figure 6B). This correlation hints to a possible involvement of this population of cells with reduced disease activity among active CD patients. When compared to their SLAMF1$^-$ ILC2 counterpart, SLAMF1$^+$ ILC2 expressed a higher amount of CRTH2 and CD161, with GATA3 being at the limit of significance (p= 0.056). In order to evaluate the activation level of the SLAMF1$^+$ population, we stained for the activation markers CD25 and CD69, which were not differentially expressed when compared to SLAMF1$^-$ ILC2. Upon ILC2 activating cytokine stimulation, a mix of IL-2, IL-25, IL-33 and TSLP, we could detect IL-13 release from SLAMF1$^+$ ILC2, although the level of release was similar to their counterpart SLAMF1$^-$ ILC2. In **Paper III**, the *SLAMF1* transcript was found in ILC2 from blood, tonsil and lung tissue (Paper III, Figure 4b). Weighted correlation network analysis (WGCNA), used to find clusters of highly correlated genes, placed *SLAMF1* in the same cluster as *IL13*, *IL17RB*, *IL1RL1* and *PPARG*, suggesting a high degree of activation for the cells expressing it. Further research is needed to shed light on the role of this Slam family receptor in ILC, as it could serve as a potential therapeutic target in IBD, given its relationship with a mature and active ILC2 subpopulation of cells.

*Characterization of circulating ILC through dimensionality reduction*

Multi-color flow cytometry is a great tool to research immune cells, given its ability to detect multiple markers for each cell analyzed. The data produced with such technology is thus multi-dimensional, and analyzing it with a classical gating strategy can lead to inaccurate conclusions, since all gates set by the analyst are biased by personal judgment on how strict or loose a gate should be for each specific marker. Because of this issue, we adopted a dimensionality reduction approach to analyze the dataset, in order to verify that our results based on statistics on the frequency of cells from biased gating strategies were not flawed. Although this approach is not completely unbiased, since we selected Lineage$^-$CD45$^+$CD127$^+$ cells by manual gating before feeding the data to the algorithm, it is unbiased when it comes to the downstream analysis of all markers of interest.

The data was projected into lower-dimensional (2D) space, which made it easier to analyze and visualize. The non-linear dimensionality reduction technique called t-distributed stochastic neighbor embedding (t-SNE), a machine learning algorithm focusing on retaining the structure of neighbor points, was adopted [204]. The use of a residual plot display readily portrayed differences between all combinations of CD/UC and HC cohorts. The differences in cell frequencies between cohorts were well represented in the t-SNE plots and kernel density based residual plots, confirming the statistical differences of cell frequencies taken from biased gating.

In summary (**Figure 10**), in **Paper II** we analyzed a large cohort of IBD and HC blood samples via multi-color flow cytometry and downstream dimensionality reduction algorithm. We were able to identify cellular activation especially in the CD cohort, with an enrichment of CD56$^+$ and NKp44$^+$ ILCp and HLADR$^+$ ILC1 and ILC2, compared to steady state frequencies in HC. SLAMF1, a marker selected associated with IBD as shown in GWAS,

was upregulated in ILC2 in IBD compared to HC. SLAMF1$^{+}$ILC2 characterization showed increased cell maturation features compared to SLAMF1$^{-}$ILC2, given the higher CRTH2, CD161 and GATA3 expression. More research is needed to explore the role of circulatory and resident SLAMF1$^{+}$ILC2.

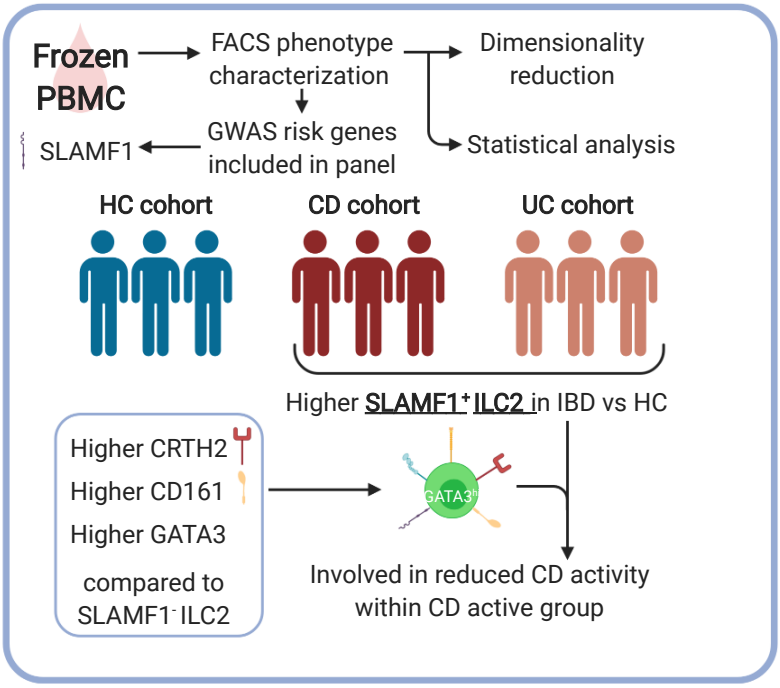


**Figure 10**. Graphical summary of **Paper II** approach and main results.

## 5.3 PAPER III - DISENTANGLING ILC HETEROGENEITY VIA SCRNA TECHNOLOGY

Over the last five years, several publications adopted scRNA-seq technology to investigate the heterogeneity of human ILC in an unbiased way, although these efforts were limited to either a single specific subset of ILC [205], or to a specific single tissue [206,207,6].

The work presented in **Paper III** focuses on exploiting the power of FACS, full length scRNA-seq, dimensionality reduction, correlation analysis and a tight collaboration with the clinics, to thoroughly investigate the cellular status of ILC in four different human tissues, blood, tonsil, lung and colon, at steady state.

Well-established cell isolation protocols were adopted for blood, lung and colon tissues in order to maximize the number of cells retrieved and the quality of the RNA post cell lysis. Cells were sorted by FACS into 384-well plates according to surface marker definition of ILC/NK and $CD4^+$ T cells, with surface marker expression level, identity and well placement of every cell recorded via index sorting (Paper III, Figure S1 – total of 4622 sorted cells). RNA quality was validated, with all plates later processed with smart-seq2 technology and sequenced by our collaborators. Upon retrieval of blood, lung and colon sequences, the tonsil data (648 cells) from Björklund, Forkel et al. was integrated to the dataset and a total of 5270 cells were analyzed.

*Dimensionality reduction and clustering analysis reveals ILC tissue and subset transcriptional imprinting*

In contrast to the dimensionality reduction method adopted in **Paper II** (t-SNE), in **Paper III** we implemented unifold manifold approximation and projection (UMAP) to visualize our dataset in low dimensional space [208]. Compared to similar methods, UMAP provides faster run times, scales well with larger datasets and produces a meaningful organization of cell clusters, preserving the local and global structure in the data [209]. By labeling the resulting UMAP with tissue and celltype metadata from FACS, we were able to visually perceive the subset and tissue specific transcriptional imprinting of ILC (Paper III, Figure 1b-c). In order to identify distinct cell populations and cell states, we adopted the Louvain algorithm, as it was shown to scale well with large datasets [210]. Louvain clustering is an unsupervised graph-based method whose purpose is to find communities (i.e. clusters) of nodes (i.e. cells) with higher probability of being connected to each other than to members of other groups in the network. No matter the clustering method adopted, the process is not completely unbiased, since at least one parameter has to be set by the user. In our case, to optimize the clustering we set a modularity resolution parameter, which divided the quality controlled data (2956 cells) into 20 different clusters (Paper III, Figure 1d). By comparing index data from FACS to each cluster, we observed a high agreement between protein FACS phenotype and the clusters subdivision (Paper III, Figure S3a), validating the resolution parameter set for the clustering step. Clusters were annotated based on known transcription profiles of human ILC [17]. We were able to distinguish single ILC1 and ILC2 clusters for blood, tonsil and lung tissues (annotated as B_ILC1, T_ILC1 and L_ILC1). No colon ILC1 or ILC2 clusters were

detected, as colon at steady state is known to have very little cells belonging to those subgroups of ILC [30,142]. ILC3 cluster number varied, with tonsils and lung having two each (both clusters for tonsils and lung were initially merged and annotated as T_ILC3, and L_ILC3, respectively) and colon having four (initially merged and annotated as C_ILC3). A single cluster was identified for blood progenitors (annotated as B_nILC) and tonsil, lung and colon NK cells (annotated as T_NK, L_NK and C_NK). Two clusters with cells from multiple tissues were discovered, one reminiscent of ILC1 in all tissues, with *EOMES* transcription factor as shared transcript, placed in between ILC1 and NK cell cluster in the UMAP space (annotated as *EOMES*$^{+}$ILC1), and a second cluster containing cells from tonsil, lung and colon, displaying a lower amount of total transcripts and naive properties, termed nILC (**Figure 11**).

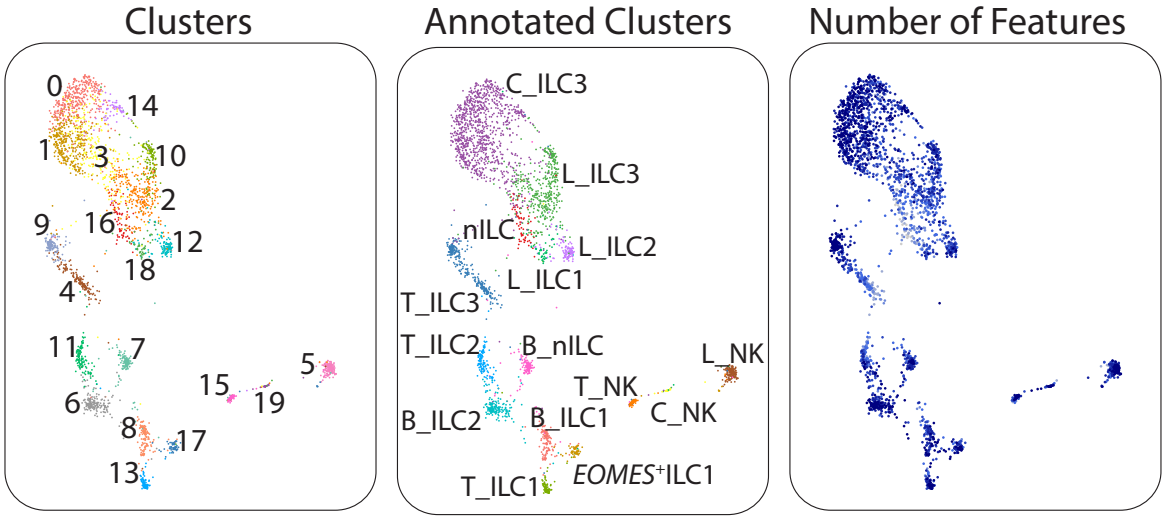


**Figure 11**. UMAP displaying clustered dataset, annotated clusters and number of features per cell (color intensity proportional to the cellular number of features). The number of features displayed for cluster 16/n_ILC are quite lower than for the rest of the clusters, given their naïve nature.

*Gene co-expression analysis identifies ILC transcriptional networks*

When trying to understand how a complicated machine engine works, does one try to count the bolts and nuts in the entire structure, or try to grasp the concept on how the system functions?

We asked ourselves this question when trying to unravel the hidden systems within the ILC transcriptomes. To answer it, instead of only focusing on the analysis of single genes obtained by differential expression (DE) analysis, we integrated DE to weighted gene co-expression network analysis (WGCNA), in order to understand how genes interact in the various ILC cells states discovered via dimensionality reduction and clustering. WGCNA calculates the interaction patterns between co-expressed genes, and places them into modules (i.e. networks of co-expressed genes). We manually tweaked the number of modules in order not to merge distinct expression networks together, setting the module number to 100. Each

module was scored based on its contribution to the explanation of metadata (e.g tissue type correlation, cell type correlation, etc).

This approach allowed us to distinguish the circulatory modules in the blood (rich in migration-associated transcripts such as *SELL*, *S1PR1*, *ITGB2*, *ICAM3*) to the resident transcriptome in colon and lung (expressing *CD69* transcripts, among others), with tonsil displaying features linked to both migration and tissue residency (Paper III, Figure 2d). Some of the key transcripts were validated on the protein level on blood, colon and tonsil tissues (CD62L, CD18, CD69, and NKp44) (Paper III, Figure 2e-g), confirming our findings *in silico*.

Four main modules specified shared characteristics with tonsil, lung and colon ILC3. Several transcripts were shared between lung and colon ILC3, with *CSF2*, *CXCL8*, *BHLHE40* and several *HLADR*/*DP*/*DQ* transcripts reflecting the active status of the cells. The prototypic ILC3 genes *RORC*, *IL1R1*, *KIT* were highest in a module mostly associated with colon ILC3. Tonsil and colon ILC3 gene networks showed similarities, with the presence of *IL23R*, *NRP1*, *TOX*, *NCR2*, *TNFSF13B* and *LTB4R* transcripts, the latter displaying a possible role of ILC3 in leukotriene metabolism, also present in protein form in colon ILC3 (Paper III, Figure 3c).

One leading module linked ILC2 from blood, tonsil and lung tissue, by displaying the classical transcripts defining ILC2, like *GATA3*, *MAF*, *PTGDR2*, *HPGD* and *HPGDS*. A second module was non-specifically associated with ILC2, displaying ties with ILC1 as well, and presenting transcripts such as *IL10RA*, *PTGER2* and *ANXA1*. Interestingly, the third module associated with ILC2 revealed a gene-network present in lung ILC2, specifically lacking *PTGDR2* expression (encoding the CRTH2 protein) and presenting activation markers like *IL13*, *IL1RL1*, *IL17RB* and *PPARG*, hinting to the possibility of activated ILC2 losing their prototypical marker (CRTH2) when active. We confirmed this hypothesis by sorting and exposing blood ILC2 to an activating cytokine cocktail, which indeed downregulated the expression of surface CRTH2 protein, also hinting to the possibility for blood ILC2 to migrate to lung and downregulate CRTH2 under alarmin influence, as suggested in mouse models [211] and a human allergen provocation study [212]. Another transcript highly expressed in this module and specific to active ILC2 in the lung was *SLAMF1*, a marker we already stumbled upon in **Paper II** that is present in activated blood ILC2. In tissues, SLAMF1 can possibly replace CRTH2 as an ILC2 marker, given its high specificity to this ILC subgroup (Paper III, Figure 4b). This remains to be explored.

The core ILC1 module was rich in T cell related transcripts, such as *CD3D*/*E*/*G*, *CD4*, *CD5* and *CD6*, confirming previous reports [15,39]. One other module showed the close association of ILC1, in particular *EOMES*$^{+}$ ILC1, to NK cells, by presenting transcripts linked to cytotoxicity, such as *GZMM*, *CD8A* and *CTSW*, together with **Paper I**'s favorite transcription factor, *IKZF3* (encoding Aiolos).

*Predicted VDJ recombination analysis disentangles blood ILC1 heterogeneity*

Given the numerous T cell related transcripts in blood ILC1 (Paper III, Figure 6b, core ILC1 module), we analyzed the blood dataset for V(D)J recombination via MIXCR [213], having $CD4^+$ T cells from the same blood donors sorted in parallel as a population of comparison. As expected, $CD4^+$ T cells presented mostly predicted AB chain rearrangement, with ILC2 and B_nILC having limited predicted rearrangement. On the other hand, B_ILC1 and *EOMES*$^+$ILC1 exhibited a combination of α/β, γ/δ and single rearranged chains, confirming the speculations on these cells being by some degree associated with T cells (Paper III, Figure 7a and S7a).

When merging $CD4^+$ T cells and ILC, UMAP analysis revealed B_ILC1 clustering either with the $CD4^+$T cells (B_ILC1 exhibiting a predicted TRA_B rearrangement) or in a separate cluster (B_ILC1 displaying TRG_D predicted rearrangement) (Paper III, Figure 7b-c). *EOMES*$^+$ILC1 clustered in between $CD4^+$T cells and $CXCR3^+$ Th1 T cells. When running DE analysis on the separate clusters based on predicted chain rearrangements, AB_ILC1 expressed transcripts of TRα and β constant and variable region, together with the *CD4* transcript. GD_ILC1 expressed transcripts of TCRδ and γ constant and variable regions, together with the transcription factor *IKZF2* (encoding for Helios). *EOMES*$^+$ILC1 expressed transcripts for TCRα and β constant and variable regions together with *CD8A* transcripts. Clonal analysis resulted in most cells representing unique clones, with the exception of two ILC1 which clones were shared with $CD4^+$ T cells.

V(D)J recombination analysis sheds some light on the possible origins of B_ILC1, with ILC1 with clonal overlap to $CD4^+$T cells most likely being T cell contaminants and ILC1 with an α/β-TCR rearrangement clustering with $CD4^+$T cells possibly being "failed $CD4^+$ T cells". We speculate that *EOMES*$^+$ILC1 possibly resemble "failed $CD8^+$ T cells" given their transcriptional profile. ILC1 with a δ-TCR rearrangement resembled immature ILC cells, given their naïve transcriptional profile (Paper III, Figure S7e) and that the vast δ-TCR rearrangement was shown to be a feature of immature murine NK cells [214].

In mice, a recent study concluded the possibility for lung ILC2 in homeostasis to be of T cell origin, since these cells display a similar TCR gene rearrangement to γδ T cells. Upon closer inspection, mouse lung ILC2 were shown to undergo TCRγ rearrangement but failed to express it, and ineffectively recombined the TCRδ locus. They propose that cells failing to rearrange their TCR gene, or are incapable to expressing a functional TCR, could cause T cell development to stop, and re-direct those cells to an ILC2 lineage [215].

Given the data presented in **Paper III**, this report raises questions: could the same apply to human ILC1? Considering the multitude of T cell related markers expressed on ILC1 (Paper III, Figure 6b, module 32), it certainly seems that ILC1 are linked to T cells. Thymic and cord blood ILC have been shown to express intracellular *CD3ε* and surface CD5, the former a marker indicating immaturity [216]. As a subset of blood ILC1 also expresses CD5 and has a

predicted TCR rearrangement, it could fit the profile of a T cell with halted development that is re-routed into the ILC lineage. Moreover, CD3ε overexpression has been linked to failed T cell development, when TCR is absent [217] . Given this subset's failed T cells origin, under the right conditions, could it up-regulate a TCR on the surface? Published reports show that a very low percentage of cultured blood ILC1 *in vitro* have even been shown to express TCRα/β at the surface, although after prolonged culture and [15]. Such cells could be T cell contaminants. In Paper III, a few cells sharing clonality with $CD4^+$ T cells from the same donor could be part of this subset, and if cultured for a prolonged period, they might express TCR on the surface. This remains to be explored.

Interestingly, expression of recombination activating enzymes RAG1 and RAG2, known to mediated V(D)J gene recombination, is lacking in ILC (data not shown). In thymocyte development, RAG expression is turned on and off at different maturation steps [218]. It is activated in the first stages to start the rearranging process; if maturing B/T cells do not meet certain requirements (e.g. failure to express antigen receptor) they undergo apoptosis. When taken together, the V(D)J rearrangements results in blood ILC from Paper III and the speculations from recent reports in mice hint to the possibility for maturing thymocytes with down-regulated RAG to escape T cell development, apoptosis, and commit to the ILC lineage, making these cells presumed "failed T cells".

It is also intriguing to see how $RAG^{-/-}$ mice have disturbed ILC frequencies, even though RAG is not expressed in ILC [219,220,221]. Moreover, such mice has been reported to have a challenged NK cell cellular fitness. More studies are needed to understand the functional consequence that RAG deficiency has on ILC, and its impact on cellular fitness [222,223].

Since a subset of blood ILC1 expresses CD4 or CD8 at the surface [15], one wonders how these ILC1 with rearranged TCR, if they are failed T cells, could escape apoptosis while undergoing pro-T/pre-T/double negative and double positive steps in the thymus without a working antigen receptor, escape from the thymus and end up in circulation. A down-regulation of the TCR after thymic education? A bold research approach is needed to answer such questions.

With scRNA technology we can only "predict" the rearrangement of TCR in ILC1, given that the receptors are not actually expressed on the membrane. A possible approach to analyzing an expressed TCR, might involve culturing blood ILC *in vitro* and perform experiments on the resulting TCR expressed on ILC surface.

More research is needed on this topic to better understand this issue.

*Trajectory analysis defining ILC3 transcriptional states*

As shown in earlier chapters and in **Paper I**, ILC are very plastic, and they can adopt dynamic identity changes following the microenvironment or lineage specification during development. Although scRNA-seq methodology allows the researcher to only capture a single picture of the cell status at a given time, transcriptome analysis enables the

reconstruction of cellular trajectories by arranging cellular states along a pseudo-temporal axis, an abstract scale of maturation. In **Paper III**, Slingshot v1.2.0 package was used for trajectory inference [224]. Since some dimensionality reduction methods can create a distorted representation of the original space (e.g UMAP, t-SNE), diffusion maps (DM) embeddings were used, as they more accurately represent gradual identity changes, allowing the identification of differentiation trajectory on scRNA-seq data [225]. A minimum spanning tree (MST) was created on the DM clusters to determine the number of lineages and bifurcation points present in the dataset, using its longest path as a backbone for the pseudotime axis. The naïve cluster (nILC for lung and colon, cluster 4 for tonsil and B_nILC for blood) was set as root, given its transcriptome profile naivety.

Trajectory analysis identified high heterogeneity among ILC3 in the colon, lung and tonsil, inferring potential maturation levels between subclusters, as identified by Louvain clustering (**Figure 12**, colon ILC3a-d, lung ILC3a-b, tonsil ILC3a-b). When comparing clusters through MAST differential expression, we could identify activated subsets (colon ILC3a-b, lung ILC3a-b, tonsil ILC3a), defined by the presence of transcripts such as *XCL1*, *XCL2*, *CXCL2* and *CXCL8*, molecules involved in chemotaxis, *CSF2*, coding for the effector molecule GM-CSF, and *CCL20*, associated with the chemokine CCR6. A naïve-like ILC3 subset was identified in tonsil (ILC3b), and a lymphoid structure related subset in colon (ILC3d). These subsets were defined by transcripts associated with naivety and migration, such as *SELL* and *S1PR1*, respectively.

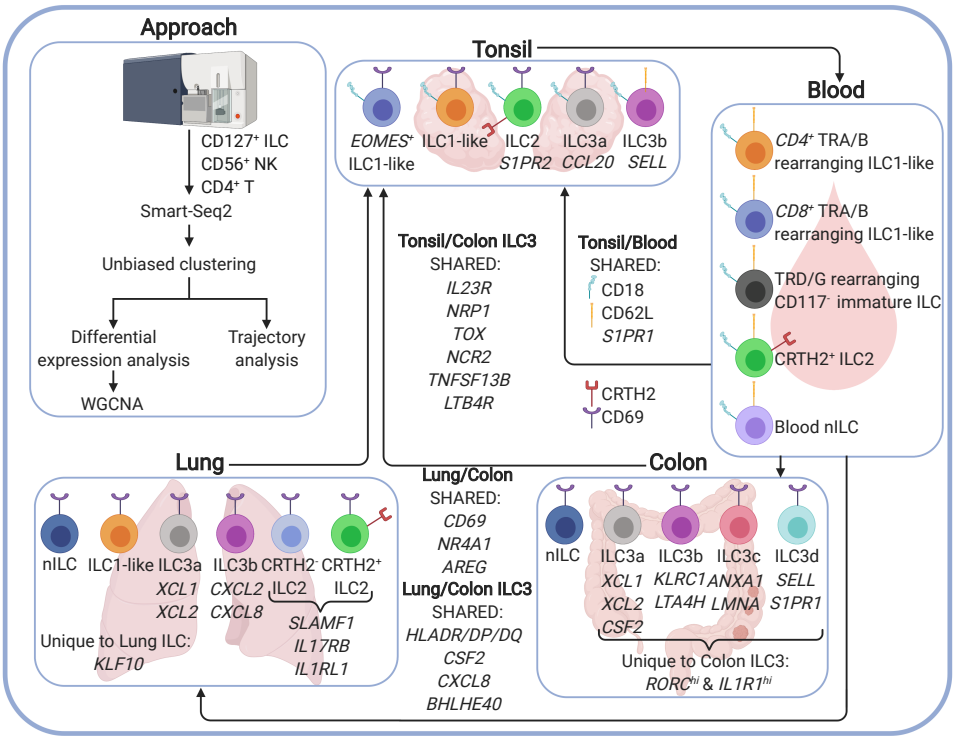


**Figure 12**. Graphical summary of Paper III approach and main findings.

In summary, we analyzed ILC from blood, tonsil, lung and colon tissues, for a total of 2956 cells after quality check. Dimensionality reduction and clustering analysis revealed ILC

heterogeneity and their subset and tissue transcriptional imprinting. Gene co-expression network analysis divided the total diversity of the dataset into 100 networks of gene, which identified ILC transcriptional networks. VDJ recombination analysis was successful in disentangling ILC1 heterogeneity, linking ILC1 to possibly failed T cells, and immature ILC cells. Through trajectory analysis we were able to identify a high degree of transcriptional heterogeneity given the variable activation status of ILC3 subgroups in tonsil, lung and colon.

In conclusion, **Paper III** supported the field's attempt to characterize ILC heterogeneity by offering interesting results and a vast dataset that will lay the foundation to future studies aimed at understanding tissue-specific ILC function.

# 6 CONCLUDING REMARKS

In the past two decades, the ILC field has advanced tremendously, from the discovery of the main ILC subsets to an understanding of their plastic behavior and development. Many questions remain unanswered, especially when it comes to human biology, given the challenge in obtaining tissues from patients for experimental purposes. The findings presented in this thesis, mainly on ILC residing in several human tissues, represent a small but important step forward in the field, and provide publically available datasets for future studies aiming at broadening the understanding of ILC function further. Listed below is a summary of the key findings from each of the three papers:

- In **Paper I** we could show that tonsil ILC3 under inflammatory conditions can trans-differentiate into different flavors of ILC1 and NK cells, building on previous reports. This transdifferation was markerd by the expression of *IKZF3*, encoding Aiolos. By silencing this transcription factor via Lenalidomide, an anticancer drug specifically targeting Ikaros family members, ILC3 to ILC1 differentiation was decreased. This also triggered an increased level of activation in Lenalidomide-treated cultured ILC3, resulting in enhanced IL-22 release. Specific drug-targeting of ILC with an Aiolos silencing drug could provide to be useful in patients with conditions like IBD, where ILC3 to ILC1 transdifferentiation is likely to occur, halting ILC1 release of inflammatory elements contributing to the inflammatory environment in the gut.
- In **Paper II** we characterized ILC in a large cohort of blood samples from patients sufferring from IBD and used blood samples from healthy donors for comparison. We detected an overall activated profile of ILC in CD and/or UC samples, examplified by the presence of NKp44, CD56 and HLA-DR on ILCs, and the lack of markers associated with naivety, such as CD45RA. Of note, we found an increased frequency of $SLAMF1^{+}$ILC2 in CD patients, which was correlated with reduced CD activity in CD active patients. We speculate that this population is an activated mature form of ILC2, given its high CRTH2, CD161 and GATA3 expression and IL-13 production capabilities. More research is needed to define the function of this population further.
- In **Paper III**, through scRNA-seq technology we were able to shed light on ILC heterogeneity in blood, tonsil, gut and lung tissues at homeostasis. Through dimensionality reduction and clustering we were successful in identifying several subpopulations of ILC, some more heterogeneous than others, displaying a range of migratory (blood ILC) and tissue resident (lung and colon ILC) features, with tonsil ILC sharing characteristics of both. Data clustering revealed a naïve ILC subset in the lung and colon tissues, most likely giving rise to more mature subsets, as illustrated by trajectory analysis. V(D)J recombination analysis in blood predicted recombination in ILC1, dividing the cells into several clusters. Through differential expression analysis, ILC to $CD4^{+}$/$CD8^{+}$ T cell and NK cell comparison, we could define such clusters, speculating on the mystery of the likely origin of ILC1, from contaminants, to failed T cells or immature ILCs. Overall, we provide a large dataset

of ILC, T and NK cell transcriptome at full-length, giving the field an opportunity to analyze it further.

## 6.1 FUTURE DIRECTIONS

Although the ILC field has progressed tremendously during my time as a PhD student, many questions remain unanswered, creating an opportunity for researchers to fill in the gaps in ILC-mediated immunity knowledge and understanding.

Possible ideas branching from the projects presented in this thesis could include continuing the investigation on the effects of the Ikaros family of TF on human primary ILC, building on **Paper I**'s results, by looking at single TF roles in ILC development and function. In order to do that efficiently, a stable and reliable gene knockdown/silencing system needs to be researched. Considerable efforts were made in setting up such a system during **Paper I**'s data collection period, but human primary ILC proved to be a tough subset to efficiently transfect. Viral-based transfection could be a way to do that effectively. Validating Aiolos' knockdown effects on ILC3-ILC1/NK trans-differentiation without the parallel silencing of Ikaros (due to Lenalidomide's dual targeting) would be exciting indeed. Based on the results of such an approach, creating a system that specifically targets the Aiolos protein in ILC of IBD patients would represent a nice system to slow down ILC3 to ILC1/NK cells trans-differentiation, effectively dampening inflammation.

Building on **Paper II**, it would be interesting to collect more samples from the IBD and HC cohorts, genotype them and evaluate ILC subsets frequencies via FACS in parallel. The aim would be to assess the determinants of immunologic variance by checking for genotype versus phenotype correlations, in order to quantify the contribution of host genetics to IBD. In order to have enough statistical power for such a project, a cohort in the hundreds of samples would be needed, making this effort considerable for clinicians involved in patient recruitment and sample collection, together with the research group involved in the lab. As FACS technology keeps improving, the detection of >30 parameters parameters is now a reality. Thanks to this technology, the project could investigate several immune subsets, without being restricted to ILC only.

A continuation of **Paper III** could include a comparison between ILC/T/NK cells from the published healthy colon data, to new data of the same subsets in IBD patients. Such a study could possibly reveal novel transcripts differentially expressed in IBD patients. The optimized knockdown system mentioned above could be adopted to efficiently investigate such transcripts and isolate possible mechanisms of protein functions *in vitro*. It would also be of interest to perform V(D)J rearrangement analysis on inflamed gut ILC1, to examine if such ILC1, possibly "ex-ILC3" that have gone through trans-differentiation, have a rearranged TCR.

More generally, future prospectives in the ILC field will most likely rely more on synergetic efforts between bioinformatics and biological research, with cutting-edge technologies providing new ways to tackle biological questions. As single cell sequencing is getting more and more accessible, and considerably cheaper, the amount of cells analyzed will only increase, with better sequencing methodologies being proposed (Smart-seq3, 10x v3, among others), giving higher cell recovery rates, read numbers and overall quality. As *in silico* methodologies are now able to integrate datasets of different origin (donor or even species), it will be exciting to unravel questions not answerable in the past, especially when it comes to rare populations of cells like ILC, in even rarer human tissues.

Applying one -omic technology will soon not be enough to push the field forward. We need to understand how ILC are shaped by their surroundings, by looking at their metabolism, their genetic influence, interactions with other molecules, at a higher scale. New approaches involving a synergetic combination of -omics will most likely take over, forcing inter-field collaborations in order to effectively decipher the multi-layered results. Such studies will provide a more unbiased and systematic approach to ILC biology, and more broadly to immunological research. The three articles presented in this thesis, with emphasis on **Paper III**, will hopefully serve as datasets to be integrated into future studies aiming towards that direction.

# 7 ACKNOWLEDGEMENTS

This thesis is the final product of my educational journey in Sweden, which started in 2013 as a master student. Many people have supported me along the way, and I'd like to thank some of them here.

While thinking about the ramification of my choices during the years, I might not even have ventured into a laboratory if it weren't for **Joseph Rafter**, who pointed out a possible lab position in Novum, Huddinge. Joseph, you ultimately started this whole PhD thing! Shame on you! ☺ Big thanks to **Mauro D'Amato** who took a chance and welcomed me in his lab, when, scientifically speaking, I couldn't tell arse from elbow. Also, grazie for your mentorship and for keeping an eye on me during my PhD studies as co-supervisor, and for all your help with Paper II.

How could I forget my "ciccia" **Aida Assadi**, who invested a lot of time and patience in teaching me a variety of molecular biology techniques that I later used in my PhD, very thankful for that! Thanks also for transferring a bit of your amazing lab organization/planning skills to me, they served me well during my years in the lab. **Ferdinando Bonfiglio** & **Tenghao Zheng**, you've been my role models when it comes to bioinformatics, total badasses, thanks for all the help & tips whenever I was in need. Nando, un grazie particolare per le analisi e l'aiuto nel paper II, much appreciated.

In 2015, as my time as a Master student was coming to an end, I decided to take a basic immunology course in Solna to learn more about the subject (thanks to **Benedikt Chambers** who let me in the PhD course as a Master student, and later was part of my half-time committee). During the course I sat behind a tall German girl with curly hair, who informed me about an open PhD position in her group. **Marianne Forkel**, thanks for completely blocking my visual during the course! If not, we might not have started talking, and I might not have known about the PhD position in Jenny's group that I later applied for and got. Very thankful for all the initial guidance you and **Viktoria Konya** provided, you were the first to introduce me to the joys of working with human samples, FACS techniques & more. You had a huge part in making the work environment in the group as great as it was.

Of course my biggest gratitude goes to my supervisor **Jenny Mjösberg**, from whom I learned how to think like a scientist. Thanks for the opportunity you've given me. The enthusiasm and passion you have for science is contagious. Very grateful for the freedom I had in pursuing my ideas & adventures and for believing in me. I'll always remember the huge work we did together (especially for Paper III), working as a team, spending months analyzing to the tiniest details huge datasets, only to then realize that something was wrong, forcing us to re-analyze whole datasets over and over again. It taught me more than immunology.

Also grateful to my co-supervisors **Niklas Björkström**, who kept an eye on me especially during the first years of doctoral studies, and **Hans-Gustaf Ljunggren**, who I thank for

creating a wonderful institution and work environment, especially at F59, where I performed the majority of the work presented in this thesis.

A big thank you goes to all past and present members of the Mjösbi group with whom I spent many hours in the green lab. **Jovana Maric**, you were on a mission during your time with us, very inspiring to see how hard and efficiently you worked; it was also fun to collaborate on your two amazing papers. **Avinash Ravindran**, I think we spent mostly nights working together, thanks for your friendship, the work in Paper III and for keeping me company in the lab during the many midnight to 4 AM sorts! **Aline Van Acker**, thank you for making the green lab brighter! **Themis Kokkinou**, was great having a PhD-buddy in the group to exchange ideas/notes/gifs with. Thanks for the good memories in the lab and for the better memories in San Diego and Israel. **Anna Rao** and **Anna Carrasco**, "the Annas", thanks for the helpful discussions, and for dragging me to "shady" bars in Tokyo. **Teresa Williams**, it was great having you in the group all the way from Australia/Canada, thank you for all your work in Paper III. **Isabel Meininger**, it was fun being part of your project, you left too soon! To the new additions to the Mjösbi group, **Tea Soini**, **Chris Tibbitt**, **Whitney Weigel**, and **Lorenz Wirth**, the group only got better with your arrival, best of luck on your projects!

Several collaborators helped me reach my goals by supporting me in a variety of ways. For Paper II a big thank you goes to **Jonas Halfvarsson** who provided us with the amazing cohort we analyzed together and key advice thanks to his deep understanding of IBD and immunology. Grateful to all bioinformaticians from NBIS, who had a crucial role in the analysis of datasets in Paper III. **Viktor Jonsson** and **Leif Wigge**, I learned so much from you, from the wonders of bash & slurm to the joys of working on the Bianca cluster. **Markus Rignér**, thanks for your supervision and the many many ideas and different approaches to solving a problem during the many hours spent on Skype/Zoom. **Åsa Björklund** and **Paulo Czarnewski**, Paper III got so much better thanks to you, from the module approach to the VDJ analysis, it was a privilege working with scRNA-seq data gurus, you rock!

Thanks to all the CIM colleagues for all the fikas, lunches, and consequent discussions that often lead to new ideas (too many to name, you know who you are!). Especially grateful for the ones I've shared an office with, from the first months of my PhD down in the "cave" (HERM 4$^{th}$ floor, 10 minutes from the lab, cubicle, no windows) with the only company of **Michal Sobkowiak**, through F59, with **Magda Lourda**, **Kim Blom**, **Monika Enqvist**, **Nicole Marquardt**, **Sofia Björnfot**, **Christine Zimmer**, **Helena Bergsten** and finally ANA8 with **Alvaro Haroun-Izquierdo**, **Quirin Hammer and Tiphaine Parrot**. A special mention goes to **Caroline Boulouis**, who put up with me for nearly the entirety of my studies, on my (left) side. Kudos to you per avermi sopportato tutti questi anni! ☺ E un grazie per non avermi fatto dimenticare l'italiano!

Stort tack till alla kollegor jag träffade på Janssen! Det var en otrolig upplevelse att kunna lära mig av er hur ett stort läkemedelsföretag fungerar, och hur många steg som ska tas innan forskning blir till klinisk praxis. Särskilt tacksam för möjligheten jag fick för två år sedan från

**Fredrik Hjelm**, **Erika Siljebrand Kruse**, **Ole Forsberg**, **Anna Tjärnlund** och nyligen från **Martin Johansson** och hela MAF teamet. More to come ;)

To the people who filled my days outside of the lab, **El Oso**, **Nico**, **Bartolo**, **Zinga**, **Samurai**, **Don**, **Tigre**. It was fun (& needed) balancing out the lab work with big rides, runs, swims, or a mix of the above. A big thanks to **Stockholm CK** & **Team Envol,** for all the adventures and battles around Sweden, Finland, Germany, Norway and Denmark. Many more to come. Agli amici di sempre in Italia, **Cutti**, **Esch**, **Capuzzo, Rosi, Ile, Michi**. È 14 anni che ho lasciato Bolzano ma voi ci siete sempre quando ne ho bisogno, grazie.

Un grazie speciale **alla mia famiglia**, che mi sostiene da sempre. Delle tante avventure passate assieme, questo traguardo è uno dei piu belli (e sudati). Chi l'avrebbe mai detto che il palpanuvole diventasse dottore? ☺

Last but not least, ett stort tack till min (p)älskling **Daniella**, mitt ankare i livet. Du ger mig stöd, stabilitet och är en stor inspirationskälla. Tack för att du alltid finns där för mig, speciellt när det stormar. Jag är glad att jag hittade dig på ön och följde med till Sverige, vilket aventyr det har varit hittills! Ser fram emot en framtid med dig ☺

# 8 FUN STATS

Tonsil processed: 124

Buffy coats processed: 15

Gut biopsies processed: 64

Lung biopsies processed: 5

Antibody tubes + MACS beads + cytokines vials used: 444

Number of FACS experiments: 363

Number of sorted samples (BD AriaIIu, Aria III, Aria FUSION, Melody and Sony): 90

Single cell sorted 384-well plates: 52

Nights spent in F59’s vilorum: 28

# 9 REFERENCES


1. Buchmann, K. Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. *Front. Immunol.* **5**, 459 (2014).

2. Vivier, E., van de Pavert, S. A., Cooper, M. D. & Belz, G. T. The evolution of innate lymphoid cells. *Nat. Immunol.* **17**, 790–4 (2016).

3. Spits, H. *et al.* Innate lymphoid cells--a proposal for uniform nomenclature. *Nat. Rev. Immunol.* **13**, 145–9 (2013).

4. Artis, D. & Spits, H. The biology of innate lymphoid cells. *Nature* **517**, 293–301 (2015).

5. Boulenouar, S. *et al.* Adipose Type One Innate Lymphoid Cells Regulate Macrophage Homeostasis through Targeted Cytotoxicity. *Immunity* **46**, 273–286 (2017).

6. Björklund, Å. K. *et al.* The heterogeneity of human CD127+ innate lymphoid cells revealed by single-cell RNA sequencing. *Nat. Immunol.* **17**, 451–460 (2016).

7. Konya, V. *et al.* Vitamin D downregulates the IL-23 receptor pathway in human mucosal group 3 innate lymphoid cells. *J. Allergy Clin. Immunol.* **141**, 279–292 (2018).

8. Dudakov, J. A. *et al.* Loss of thymic innate lymphoid cells leads to impaired thymopoiesis in experimental graft-versus-host disease. *Blood* **130**, 933–942 (2017).

9. Kim, B. S. Innate lymphoid cells in the skin. *J. Invest. Dermatol.* **135**, 673–678 (2015).

10. Forkel, M. *et al.* Composition and functionality of the intrahepatic innate lymphoid cell-compartment in human nonfibrotic and fibrotic livers. *Eur. J. Immunol.* **47**, 1280–1294 (2017).

11. Lai, D.-M., Shu, Q. & Fan, J. The origin and role of innate lymphoid cells in the lung. *Mil. Med. Res.* **3**, 25 (2016).

12. Doisne, J.-M. *et al.* Composition, Development, and Function of Uterine Innate Lymphoid Cells. *J. Immunol.* **195**, 3937–45 (2015).

13. Moretta, F. *et al.* The generation of human innate lymphoid cells is influenced by the source of hematopoietic stem cells and by the use of G-CSF. *Eur. J. Immunol.* **46**, 1271–1278 (2016).

14. Mjösberg, J. M. *et al.* Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are defined by expression of CRTH2 and CD161. *Nat. Immunol.* **12**, 1055–1062 (2011).

15. Roan, F. *et al.* CD4 $^{+}$ Group 1 Innate Lymphoid Cells (ILC) Form a Functionally Distinct ILC Subset That Is Increased in Systemic Sclerosis. *J. Immunol.* **196**, 2051–2062 (2016).

16. Turner, J.-E. & Gasteiger, G. Innate lymphoid cells: key players in tissue-specific immunity. *Semin. Immunopathol.* **40**, 315–317 (2018).

17. Vivier, E. *et al.* Leading Edge Review Innate Lymphoid Cells: 10 Years On. (2018) doi:10.1016/j.cell.2018.07.017.

18. Conference and Workshop on Cellular Immune Reactions to Human Tumor-Associated Antigens. in *Conference and Workshop on Cellular Immune Reactions to Human Tumor-Associated Antigens* (ed. Herberman RB, G. C.) monograph 37 (Nat Cancer Inst., 1973).

19. Rosenberg, E. B. *et al.* Lymphocyte cytotoxicity reactions to leukemia-associated antigens in identical twins. *Int. J. cancer* **9**, 648–58 (1972).

20. Herberman, R. B., Nunn, M. E., Holden, H. T. & Lavrin, D. H. Natural cytotoxic reactivity of mouse lymphoid cells against syngeneic and allogeneic tumors. II. Characterization of effector cells. *Int. J. cancer* **16**, 230–9 (1975).

21. 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. *Eur. J. Immunol.* **5**, 117–121 (1975).

22. Mebius, R. E., Rennert, P. & Weissman, I. L. Developing lymph nodes collect CD4+CD3- LTbeta+ cells that can differentiate to APC, NK cells, and follicular cells but not T or B cells. *Immunity* **7**, 493–504 (1997).

23. Sanos, S. L. *et al.* RORγt and commensal microflora are required for the differentiation of mucosal interleukin 22–producing NKp46+ cells. *Nat. Immunol.* **10**, 83–91 (2009).

24. Price, A. E. *et al.* Systemically dispersed innate IL-13-expressing cells in type 2 immunity. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11489–94 (2010).

25. Moro, K. *et al.* Innate production of TH2 cytokines by adipose tissue-associated c-Kit+Sca-1+ lymphoid cells. *Nature* **463**, 540–544 (2010).

26. Neill, D. R. *et al.* Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* **464**, 1367–1370 (2010).

27. Klose, C. S. N. *et al.* Differentiation of Type 1 ILCs from a Common Progenitor to All Helper-like Innate Lymphoid Cell Lineages. *Cell* **157**, 340–356 (2014).

28. Constantinides, M. G., McDonald, B. D., Verhoef, P. A. & Bendelac, A. A committed precursor to innate lymphoid cells. *Nature* **508**, 397–401 (2014).

29. Spits, H. *et al.* Innate lymphoid cells--a proposal for uniform nomenclature. *Nat. Rev. Immunol.* **13**, 145–9 (2013).

30. Bernink, J. H. *et al.* Human type 1 innate lymphoid cells accumulate in inflamed mucosal tissues. *Nat. Immunol.* **14**, 221–229 (2013).

31. Vonarbourg, C. *et al.* Regulated expression of nuclear receptor RORγt confers distinct functional fates to NK cell receptor-expressing RORγt(+) innate lymphocytes. *Immunity* **33**, 736–51 (2010).

32. Kang, M.-J. *et al.* Cigarette smoke selectively enhances viral PAMP- and virus-induced pulmonary innate immune and remodeling responses in mice. *J. Clin. Invest.* **118**, 2771–84 (2008).

33. Schoenborn, J. R. & Wilson, C. B. Regulation of Interferon‐γ During Innate and Adaptive Immune Responses. in *Advances in immunology* vol. 96 41–101 (2007).

34. Klose, C. S. N. *et al.* A T-bet gradient controls the fate and function of

CCR6−RORγt+ innate lymphoid cells. *Nature* **494**, 261–265 (2013).

35. Fuchs, A. *et al.* Intraepithelial type 1 innate lymphoid cells are a unique subset of IL-12- and IL-15-responsive IFN-γ-producing cells. *Immunity* **38**, 769–81 (2013).

36. Male, V., Nisoli, I., Gascoyne, D. M. & Brady, H. J. M. E4BP4: an unexpected player in the immune response. *Trends Immunol.* **33**, 98–102 (2012).

37. Simoni, Y. & Newell, E. W. Toward Meaningful Definitions of Innate-Lymphoid-Cell Subsets. *Immunity* **46**, 760–761 (2017).

38. Bernink, J. H., Mjösberg, J. & Spits, H. Human ILC1: To Be or Not to Be. *Immunity* **46**, 756–757 (2017).

39. Roan, F. & Ziegler, S. F. Human Group 1 Innate Lymphocytes Are Negative for Surface CD3ε but Express CD5. *Immunity* **46**, 758–759 (2017).

40. Simoni, Y. *et al.* Human Innate Lymphoid Cell Subsets Possess Tissue-Type Based Heterogeneity in Phenotype and Frequency. *Immunity* **46**, 148–161 (2017).

41. Poli, A. *et al.* CD56 $^{bright}$ natural killer (NK) cells: an important NK cell subset. *Immunology* **126**, 458–465 (2009).

42. Moretta, L. Dissecting CD56dim human NK cells. *Blood* **116**, 3689–91 (2010).

43. Björkström, N. K., Ljunggren, H.-G. & Michaëlsson, J. Emerging insights into natural killer cells in human peripheral tissues. *Nat. Rev. Immunol.* **16**, 310–20 (2016).

44. Mandelboim, O. *et al.* Recognition of haemagglutinins on virus-infected cells by NKp46 activates lysis by human NK cells. *Nature* **409**, 1055–1060 (2001).

45. Moretta, L. & Moretta, A. Unravelling natural killer cell function: triggering and inhibitory human NK receptors. *EMBO J.* **23**, 255–9 (2004).

46. Jiao, L. *et al.* NK cells promote type 1 T cell immunity through modulating the function of dendritic cells during intracellular bacterial infection. *J. Immunol.* **187**, 401–11 (2011).

47. Lanier, L. L. NK CELL RECOGNITION. *Annu. Rev. Immunol.* **23**, 225–274 (2005).

48. Topham, N. J. & Hewitt, E. W. Natural killer cell cytotoxicity: how do they pull the trigger? *Immunology* **128**, 7–15 (2009).

49. Mistry, A. R. & O'Callaghan, C. A. Regulation of ligands for the activating receptor NKG2D. *Immunology* **121**, 439–447 (2007).

50. Voskoboinik, I., Whisstock, J. C. & Trapani, J. A. Perforin and granzymes: function, dysfunction and human pathology. *Nat. Rev. Immunol.* **15**, 388–400 (2015).

51. Fort, M. M. *et al.* IL-25 induces IL-4, IL-5, and IL-13 and Th2-associated pathologies in vivo. *Immunity* **15**, 985–95 (2001).

52. Voehringer, D., Reese, T. A., Huang, X., Shinkai, K. & Locksley, R. M. Type 2 immunity is controlled by IL-4/IL-13 expression in hematopoietic non-eosinophil cells of the innate immune system. *J. Exp. Med.* **203**, 1435–46 (2006).

53. Fallon, P. G. *et al.* Identification of an interleukin (IL)-25–dependent cell population that provides IL-4, IL-5, and IL-13 at the onset of helminth expulsion. *J. Exp. Med.*

**203**, 1105–1116 (2006).

54. Mjösberg, J. *et al.* The Transcription Factor GATA3 Is Essential for the Function of Human Type 2 Innate Lymphoid Cells. *Immunity* **37**, 649–659 (2012).

55. Wong, S. H. *et al.* Transcription factor RORα is critical for nuocyte development. *Nat. Immunol.* **13**, 229–236 (2012).

56. Salimi, M. *et al.* A role for IL-25 and IL-33–driven type-2 innate lymphoid cells in atopic dermatitis. *J. Exp. Med.* **210**, 2939 (2013).

57. Monticelli, L. A. *et al.* Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nat. Immunol.* **12**, 1045–54 (2011).

58. Maric, J. *et al.* Prostaglandin E 2 suppresses human group 2 innate lymphoid cell function. *J. Allergy Clin. Immunol.* (2017) doi:10.1016/j.jaci.2017.09.050.

59. Konya, V. & Mjösberg, J. Lipid mediators as regulators of human ILC2 function in allergic diseases. *Immunol. Lett.* **179**, 36–42 (2016).

60. Kim, B. S. *et al.* Basophils promote innate lymphoid cell responses in inflamed skin. *J. Immunol.* **193**, 3717–25 (2014).

61. Morimoto, M. *et al.* Functional importance of regional differences in localized gene expression of receptors for IL-13 in murine gut. *J. Immunol.* **176**, 491–5 (2006).

62. Wynn, T. A. IL-13 E FFECTOR F UNCTIONS. *Annu. Rev. Immunol.* **21**, 425–456 (2003).

63. Halim, T. Y. F. *et al.* Group 2 innate lymphoid cells are critical for the initiation of adaptive T helper 2 cell-mediated allergic lung inflammation. *Immunity* **40**, 425–35 (2014).

64. Hamann, K. J. *et al.* In vitro killing of microfilariae of Brugia pahangi and Brugia malayi by eosinophil granule proteins. *J. Immunol.* **144**, 3166–73 (1990).

65. Shin1, E.-H., Osada1, Y., Sagara2, H., Takatsu3, K. & Kojima1, S. Involvement of complement and fibronectin in eosinophil-mediated damage to Nippostrongylus brasiliensis larvae. *Parasite Immunol.* **23**, 27–37 (2001).

66. Turner, J.-E. *et al.* IL-9–mediated survival of type 2 innate lymphoid cells promotes damage control in helminth-induced lung inflammation. *J. Exp. Med.* **210**, 2951 (2013).

67. Zaiss, D. M. W., Gause, W. C., Osborne, L. C. & Artis, D. Emerging functions of amphiregulin in orchestrating immunity, inflammation and tissue repair. *Immunity* **42**, 216 (2015).

68. Chen, F. *et al.* An essential role for TH2-type responses in limiting acute tissue damage during experimental helminth infection. *Nat. Med.* **18**, 260–266 (2012).

69. Huang, Y. *et al.* IL-25-responsive, lineage-negative KLRG1hi cells are multipotential 'inflammatory' type 2 innate lymphoid cells. *Nat. Immunol.* **16**, 161–169 (2014).

70. Golebski, K. *et al.* IL-1β, IL-23, and TGF-β drive plasticity of human ILC2s towards IL-17-producing ILCs in nasal inflammation. *Nat. Commun.* **10**, 2162 (2019).

71. Crellin, N. K. *et al.* Regulation of Cytokine Secretion in Human CD127+ LTi-like Innate Lymphoid Cells by Toll-like Receptor 2. *Immunity* **33**, 752–764 (2010).

72. Satoh-Takayama, N. *et al.* Microbial flora drives interleukin 22 production in intestinal NKp46+ cells that provide innate mucosal immune defense. *Immunity* **29**, 958–70 (2008).

73. Cella, M. *et al.* A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. *Nature* **457**, 722–725 (2009).

74. Hughes, T. *et al.* Interleukin-1beta selectively expands and sustains interleukin-22+ immature human natural killer cells in secondary lymphoid tissue. *Immunity* **32**, 803–14 (2010).

75. Liang, S. C. *et al.* Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. *J. Exp. Med.* **203**, 2271–2279 (2006).

76. Sugimoto, K. *et al.* IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *J. Clin. Invest.* **118**, 534–44 (2008).

77. Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nat. Med.* **14**, 282–289 (2008).

78. Gladiator, A., Wangler, N., Trautwein-Weidner, K. & LeibundGut-Landmann, S. Cutting edge: IL-17-secreting innate lymphoid cells are essential for host defense against fungal infection. *J. Immunol.* **190**, 521–5 (2013).

79. Hoorweg, K. *et al.* Functional Differences between Human NKp44(-) and NKp44(+) RORC(+) Innate Lymphoid Cells. *Front. Immunol.* **3**, 72 (2012).

80. Lim, A. I. *et al.* Systemic Human ILC Precursors Provide a Substrate for Tissue ILC Differentiation. *Cell* **168**, 1086-1100.e10 (2017).

81. Cupedo, T. *et al.* Human fetal lymphoid tissue-inducer cells are interleukin 17-producing precursors to RORC+ CD127+ natural killer-like cells. *Nat. Immunol.* **10**, 66–74 (2009).

82. Shikhagaie, M. M. *et al.* Neuropilin-1 Is Expressed on Lymphoid Tissue Residing LTi-like Group 3 Innate Lymphoid Cells and Associated with Ectopic Lymphoid Aggregates. *Cell Rep.* **18**, 1761–1773 (2017).

83. Robinette, M. L. *et al.* Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. *Nat. Immunol.* **16**, 306–17 (2015).

84. Yu, X. *et al.* The basic leucine zipper transcription factor NFIL3 directs the development of a common innate lymphoid cell precursor. *Elife* **3**, (2014).

85. Seehus, C. R. *et al.* The development of innate lymphoid cells requires TOX-dependent generation of a common innate lymphoid cell progenitor. *Nat. Immunol.* **16**, 599–608 (2015).

86. Yokota, Y. *et al.* Development of peripheral lymphoid organs and natural killer cells depends on the helix–loop–helix inhibitor Id2. *Nature* **397**, 702–706 (1999).

87. Yang, Q. *et al.* TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nat. Immunol.* **16**, 1044–50 (2015).

88. Zook, E. C. *et al.* The ETS1 transcription factor is required for the development and cytokine-induced expansion of ILC2. *J. Exp. Med.* **213**, 687–96 (2016).

89. Gordon, S. M. *et al.* The Transcription Factors T-bet and Eomes Control Key Checkpoints of Natural Killer Cell Maturation. *Immunity* **36**, 55–67 (2012).

90. Yagi, R. *et al.* The transcription factor GATA3 is critical for the development of all IL-7Rα-expressing innate lymphoid cells. *Immunity* **40**, 378–88 (2014).

91. Constantinides, M. G. *et al.* PLZF expression maps the early stages of ILC1 lineage development. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5123–8 (2015).

92. Montaldo, E. *et al.* Human RORγt+CD34+ Cells Are Lineage-Specified Progenitors of Group 3 RORγt+ Innate Lymphoid Cells. *Immunity* **41**, 988–1000 (2014).

93. Scoville, S. D. *et al.* A Progenitor Cell Expressing Transcription Factor RORγt Generates All Human Innate Lymphoid Cell Subsets. *Immunity* **44**, 1140–50 (2016).

94. Gasteiger, G., Fan, X., Dikiy, S., Lee, S. Y. & Rudensky, A. Y. Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. *Science* **350**, 981–5 (2015).

95. Bernink, J. H. *et al.* Interleukin-12 and -23 Control Plasticity of CD127+ Group 1 and Group 3 Innate Lymphoid Cells in the Intestinal Lamina Propria. *Immunity* **43**, 146–60 (2015).

96. King, P. T. Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer. *Clin. Transl. Med.* **4**, 68 (2015).

97. Bal, S. M. *et al.* IL-1β, IL-4 and IL-12 control the fate of group 2 innate lymphoid cells in human airway inflammation in the lungs. *Nat. Immunol.* **17**, 636–645 (2016).

98. Ohne, Y. *et al.* IL-1 is a critical regulator of group 2 innate lymphoid cell function and plasticity. *Nat. Immunol.* **17**, 646–655 (2016).

99. Silver, J. S. *et al.* Inflammatory triggers associated with exacerbations of COPD orchestrate plasticity of group 2 innate lymphoid cells in the lungs. *Nat. Immunol.* **17**, 626–635 (2016).

100. Ohne, Y. *et al.* IL-1 is a critical regulator of group 2 innate lymphoid cell function and plasticity. *Nat. Immunol.* **17**, 646–655 (2016).

101. Gao, Y. *et al.* Tumor immunoevasion by the conversion of effector NK cells into type 1 innate lymphoid cells. *Nat. Immunol.* **18**, 1004–1015 (2017).

102. Cortez, V. S. *et al.* SMAD4 impedes the conversion of NK cells into ILC1-like cells by curtailing non-canonical TGF-β signaling. *Nat. Immunol.* **18**, 995–1003 (2017).

103. Raykova, A. *et al.* Interleukins 12 and 15 induce cytotoxicity and early NK-cell differentiation in type 3 innate lymphoid cells. *Blood Adv.* **1**, 2679–2691 (2017).

104. Mazzurana, L., Rao, A., Van Acker, A. & Mjösberg, J. The roles for innate lymphoid cells in the human immune system. *Semin. Immunopathol.* **40**, 407–419 (2018).

105. Seyedian, S. S., Nokhostin, F. & Malamir, M. D. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. *J. Med. Life* **12**, 113–122 (2019).

106. Li, J. & Glover, S. C. Innate Lymphoid Cells in Inflammatory Bowel Disease. *Arch. Immunol. Ther. Exp. (Warsz).* **66**, 415–421 (2018).

107. Stokkers, P. C. F. & Hommes, D. W. New cytokine therapeutics for inflammatory bowel disease. in *Cytokine* vol. 28 167–173 (2004).

108. Yu, Y. R. & Rodriguez, J. R. Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes. *Semin. Pediatr. Surg.* **26**, 349–355 (2017).

109. Habibi, F. *et al.* Quality of life in inflammatory bowel disease patients: A cross-sectional study. *J. Res. Med. Sci.* **22**, 104 (2017).

110. F, C., MJ, A., L, P., JR, M. & JR, M. Assessment of the influence of disease activity on the quality of life of patients with inflammatory bowel disease using a short questionnaire. *Am. J. Gastroenterol.* **99**, (2004).

111. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).

112. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–24 (2012).

113. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–25 (2010).

114. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246 (2011).

115. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).

116. Assadi, G. *et al.* Functional Analyses of the Crohn's Disease Risk Gene LACC1. *PLoS One* **11**, e0168276 (2016).

117. Cader, M. Z. *et al.* C13orf31 (FAMIN) is a central regulator of immunometabolic function. *Nat. Immunol.* **17**, 1046–56 (2016).

118. Ananthakrishnan, A. N. Environmental risk factors for inflammatory bowel disease. *Gastroenterol. Hepatol. (N. Y).* **9**, 367–74 (2013).

119. Cosnes, J., Carbonnel, F., Beaugerie, L., Le Quintrec, Y. & Gendre, J. P. Effects of cigarette smoking on the long-term course of Crohn's disease. *Gastroenterology* **110**, 424–31 (1996).

120. Cosnes, J. *et al.* Effects of current and former cigarette smoking on the clinical course of Crohn's disease. *Aliment. Pharmacol. Ther.* **13**, 1403–11 (1999).

121. Andersson, R. E., Olaison, G., Tysk, C. & Ekbom, A. Appendectomy is followed by increased risk of Crohn's disease. *Gastroenterology* **124**, 40–6 (2003).

122. Jantchou, P., Morois, S., Clavel-Chapelon, F., Boutron-Ruault, M.-C. & Carbonnel, F. Animal protein intake and risk of inflammatory bowel disease: The E3N prospective study. *Am. J. Gastroenterol.* **105**, 2195–201 (2010).

123. Chan, S. S. M. *et al.* Aspirin in the aetiology of Crohn's disease and ulcerative colitis: a European prospective cohort study. *Aliment. Pharmacol. Ther.* **34**, 649–55 (2011).

124. Virta, L., Auvinen, A., Helenius, H., Huovinen, P. & Kolho, K.-L. Association of repeated exposure to antibiotics with the development of pediatric Crohn's disease--a nationwide, register-based finnish case-control study. *Am. J. Epidemiol.* **175**, 775–84 (2012).

125. Sun, Y. *et al.* Stress Triggers Flare of Inflammatory Bowel Disease in Children and Adults. *Front. Pediatr.* **7**, (2019).

126. Graff, L. A., Walker, J. R. & Bernstein, C. N. Depression and anxiety in inflammatory bowel disease: a review of comorbidity and management. *Inflamm. Bowel Dis.* **15**, 1105–18 (2009).

127. H, K. *et al.* Oral contraceptives, reproductive factors and risk of inflammatory bowel disease. *Gut* **62**, (2013).

128. Ananthakrishnan, A. N. *et al.* A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. *Gastroenterology* **145**, 970–7 (2013).

129. AN, A. *et al.* Higher predicted vitamin D status is associated with reduced risk of Crohn's disease. *Gastroenterology* **142**, (2012).

130. PS, de S. *et al.* An association between dietary arachidonic acid, measured in adipose tissue, and ulcerative colitis. *Gastroenterology* **139**, (2010).

131. IBD in EPIC Study Investigators, T. *et al.* Linoleic acid, a dietary n-6 polyunsaturated fatty acid, and the aetiology of ulcerative colitis: a nested case-control study within a European prospective cohort study. *Gut* **58**, 1606–11 (2009).

132. S, L. *et al.* Psychological stress and disease activity in ulcerative colitis: a multidimensional cross-sectional study. *Am. J. Gastroenterol.* **89**, (1994).

133. Khalili, H. *et al.* Hormone therapy increases risk of ulcerative colitis but not Crohn's disease. *Gastroenterology* **143**, 1199–1206 (2012).

134. Higuchi, L. M. *et al.* A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. *Am. J. Gastroenterol.* **107**, 1399–406 (2012).

135. RE, A., G, O., C, T. & A, E. Appendectomy and protection against ulcerative colitis. *N. Engl. J. Med.* **344**, (2001).

136. Steel, A. W., Mela, C. M., Lindsay, J. O., Gazzard, B. G. & Goodier, M. R. Increased proportion of CD16+ NK cells in the colonic lamina propria of inflammatory bowel disease patients, but not after azathioprine treatment. *Aliment. Pharmacol. Ther.* **33**, 115–126 (2011).

137. Ng, S. C. *et al.* A novel population of human $CD56^+$ human leucocyte antigen D-related ($HLA\text{-}DR^+$) colonic lamina propria cells is associated with inflammation in ulcerative colitis. *Clin. Exp. Immunol.* **158**, 205–218 (2009).

138. Schroder, K., Hertzog, P. J., Ravasi, T. & Hume, D. A. Interferon-γ: an overview of signals, mechanisms and functions. *J. Leukoc. Biol.* **75**, 163–189 (2004).

139. Martín-Fontecha, A. *et al.* Induced recruitment of NK cells to lymph nodes provides

IFN-γ for TH1 priming. *Nat. Immunol.* **5**, 1260–1265 (2004).

140. Morandi, B., Bougras, G., Muller, W. A., Ferlazzo, G. & Münz, C. NK cells of human secondary lymphoid tissues enhance T cell polarizationvia IFN-γ secretion. *Eur. J. Immunol.* **36**, 2394–2400 (2006).

141. Gerosa, F. *et al.* Reciprocal Activating Interaction between Natural Killer Cells and Dendritic Cells. *J. Exp. Med.* **195**, 327–333 (2002).

142. Forkel, M. *et al.* Distinct Alterations in the Composition of Mucosal Innate Lymphoid Cells in Newly Diagnosed and Established Crohn's Disease and Ulcerative Colitis. *J. Crohns. Colitis* **13**, 67–78 (2019).

143. Li, J., Doty, A. L., Iqbal, A. & Glover, S. C. The differential frequency of Lineage(-)CRTH2(-)CD45(+)NKp44(-)CD117(-)CD127(+)ILC subset in the inflamed terminal ileum of patients with Crohn's disease. *Cell. Immunol.* (2016) doi:10.1016/j.cellimm.2016.05.001.

144. Adams, R. B., Planchon, S. M. & Roche, J. K. IFN-gamma modulation of epithelial barrier function. Time course, reversibility, and site of cytokine binding. *J. Immunol.* **150**, 2356–2363 (1993).

145. Boehm, U., Klamp, T., Groot, M. & Howard, J. C. CELLULAR RESPONSES TO INTERFERON-γ. *Annu. Rev. Immunol.* **15**, 749–795 (1997).

146. Lim, A. I. *et al.* IL-12 drives functional plasticity of human group 2 innate lymphoid cells. *J. Exp. Med.* **213**, 569–83 (2016).

147. Monticelli, L. A. *et al.* IL-33 promotes an innate immune pathway of intestinal tissue protection dependent on amphiregulin-EGFR interactions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10762–7 (2015).

148. von Moltke, J., Ji, M., Liang, H.-E. & Locksley, R. M. Tuft-cell-derived IL-25 regulates an intestinal ILC2-epithelial response circuit. *Nature* **529**, 221–5 (2016).

149. Hengst, J. *et al.* High-resolution determination of human immune cell signatures from fine-needle liver aspirates. *Eur. J. Immunol.* **45**, 2154–2157 (2015).

150. Okoniewski, M. J. & Miller, C. J. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* **7**, 276 (2006).

151. Rao, M. S. *et al.* Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Front. Genet.* **9**, 636 (2018).

152. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).

153. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–5 (1994).

154. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–90 (1998).

155. Warren, L., Bryder, D., Weissman, I. L. & Quake, S. R. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl. Acad. Sci. U.*

*S. A.* **103**, 17807–12 (2006).

156. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).

157. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–7 (2011).

158. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).

159. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–9 (2014).

160. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

161. Zhang, M. *et al.* Highly parallel and efficient single cell mRNA sequencing with paired picoliter chambers. *Nat. Commun.* **11**, 2118 (2020).

162. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).

163. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).

164. Soumillon, M., Cacchiarelli, D., Semrau, S., Oudenaarden, A. van & Mikkelsen, T. S. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv* 003236 (2014) doi:10.1101/003236.

165. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).

166. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–8 (2013).

167. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

168. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631-643.e4 (2017).

169. John, L. B. & Ward, A. C. The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. *Mol. Immunol.* **48**, 1272–8 (2011).

170. Georgopoulos, K., Moore, D. D. & Derfler, B. Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. *Science* **258**, 808–12 (1992).

171. Georgopoulos, K. *et al.* The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**, 143–56 (1994).

172. Morgan, B. *et al.* Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J.* **16**, 2004–13 (1997).

173. Quintana, F. J. *et al.* Aiolos promotes TH17 differentiation by directly silencing Il2 expression. *Nat. Immunol.* **13**, 770–777 (2012).

174. Holmes, M. L. *et al.* Peripheral natural killer cell maturation depends on the transcription factor Aiolos. *EMBO J.* **33**, 2721–34 (2014).

175. Kelley, C. M. *et al.* Helios, a novel dimerization partner of Ikaros expressed in the earliest hematopoietic progenitors. *Curr. Biol.* **8**, 508–15 (1998).

176. Singh, K., Hjort, M., Thorvaldson, L. & Sandler, S. Concomitant analysis of Helios and Neuropilin-1 as a marker to detect thymic derived regulatory T cells in naïve mice. *Sci. Rep.* **5**, 7767 (2015).

177. Sebastian, M. *et al.* Helios Controls a Limited Subset of Regulatory T Cell Functions. *J. Immunol.* **196**, 144–55 (2016).

178. Yeung, F., Chung, E., Guess, M. G., Bell, M. L. & Leinwand, L. A. Myh7b/miR-499 gene expression is transcriptionally regulated by MRFs and Eos. *Nucleic Acids Res.* **40**, 7303–7318 (2012).

179. Hu, R. *et al.* Eos, MITF, and PU.1 Recruit Corepressors to Osteoclast-Specific Genes in Committed Myeloid Progenitors. *Mol. Cell. Biol.* **27**, 4018–4027 (2007).

180. Perdomo, J., Holmes, M., Chong, B. & Crossley, M. Eos and pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities. *J. Biol. Chem.* **275**, 38347–54 (2000).

181. Marquardt, N. *et al.* Cutting edge: identification and characterization of human intrahepatic CD49a+ NK cells. *J. Immunol.* **194**, 2467–71 (2015).

182. Ealey, K. N. & Koyasu, S. How Many Subsets of Innate Lymphoid Cells Do We Need? *Immunity* **46**, 10–13 (2017).

183. Hoorweg, K. *et al.* Functional Differences between Human NKp44(-) and NKp44(+) RORC(+) Innate Lymphoid Cells. *Front. Immunol.* **3**, 72 (2012).

184. S, L., JW, B. & L, Z. Regulation of Innate Lymphoid Cells by Aryl Hydrocarbon Receptor. *Front. Immunol.* **8**, (2018).

185. Mjösberg, J., Bernink, J., Peters, C. & Spits, H. Transcriptional control of innate lymphoid cells. *Eur. J. Immunol.* **42**, 1916–23 (2012).

186. Kiss, E. A. & Diefenbach, A. Role of the Aryl Hydrocarbon Receptor in Controlling Maintenance and Functional Programs of RORγt(+) Innate Lymphoid Cells and Intraepithelial Lymphocytes. *Front. Immunol.* **3**, 124 (2012).

187. Li, S. *et al.* Aryl Hydrocarbon Receptor Signaling Cell Intrinsically Inhibits Intestinal Group 2 Innate Lymphoid Cell Function. *Immunity* **49**, 915-928.e5 (2018).

188. Li, J., Doty, A. & Glover, S. C. Aryl hydrocarbon receptor signaling involves in the human intestinal ILC3/ILC1 conversion in the inflamed terminal ileum of Crohn's disease patients. *Inflamm. cell Signal.* **3**, (2016).

189. Veldhoen, M. *et al.* The aryl hydrocarbon receptor links TH17-cell-mediated autoimmunity to environmental toxins. *Nature* **453**, 106–9 (2008).

190. Quintana, F. J. *et al.* Control of T(reg) and T(H)17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453**, 65–71 (2008).

191. Hughes, T. *et al.* The transcription Factor AHR prevents the differentiation of a stage 3

innate lymphoid cell subset to natural killer cells. *Cell Rep.* **8**, 150–62 (2014).

192. Tageja, N. Lenalidomide – Current Understanding of Mechanistic Properties. *Anticancer. Agents Med. Chem.* **11**, 315–326 (2011).

193. Kronke, J. *et al.* Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma Cells. *Science (80-. ).* **343**, 301–305 (2014).

194. Lu, G. *et al.* The Myeloma Drug Lenalidomide Promotes the Cereblon-Dependent Destruction of Ikaros Proteins. *Science (80-. ).* **343**, 305–309 (2014).

195. Kritharis, A., Coyle, M., Sharma, J. & Evens, A. M. Lenalidomide in non-Hodgkin lymphoma: biological perspectives and therapeutic opportunities. *Blood* **125**, 2471–2476 (2015).

196. Bar-Ephraim, Y. E. *et al.* CD62L Is a Functional and Phenotypic Marker for Circulating Innate Lymphoid Cell Precursors. *J. Immunol.* **202**, 171–182 (2019).

197. Creyns, B. *et al.* Biological Therapy in Inflammatory Bowel Disease Patients Partly Restores Intestinal Innate Lymphoid Cell Subtype Equilibrium. *Front. Immunol.* **11**, 1847 (2020).

198. Momozawa, Y. *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 2427 (2018).

199. GTEx Portal. https://gtexportal.org/home/.

200. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20120362 (2013).

201. Gunesch, J. T. *et al.* CD56 regulates human NK cell cytotoxicity through Pyk2. *bioRxiv* 2020.03.19.998427 (2020) doi:10.1101/2020.03.19.998427.

202. Rao, A. *et al.* Cytokines regulate the antigen-presenting characteristics of human circulating and tissue-resident intestinal ILCs. *Nat. Commun.* **11**, 2049 (2020).

203. Munneke, J. M. *et al.* Activated innate lymphoid cells are associated with a reduced susceptibility to graft-versus-host disease. *Blood* **124**, 812–21 (2014).

204. Van Der Maaten, L. & Hinton, G. *Visualizing Data using t-SNE*. *Journal of Machine Learning Research* vol. 9 http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf (2008).

205. Crinier, A. *et al.* High-Dimensional Single-Cell Analysis Identifies Organ-Specific Signatures and Conserved NK Cell Subsets in Humans and Mice. *Immunity* **49**, 971-986.e5 (2018).

206. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e20 (2019).

207. Cella, M. *et al.* Subsets of ILC3−ILC1-like cells generate a diversity spectrum of innate lymphoid cells in human mucosal tissues. *Nat. Immunol.* **20**, 980–991 (2019).

208. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).

209. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).

210. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. *Fast unfolding of communities in large networks*. https://arxiv.org/pdf/0803.0476.pdf (2008).

211. Wojno, E. D. T. *et al.* The prostaglandin $D_2$ receptor CRTH2 regulates accumulation of group 2 innate lymphoid cells in the inflamed lung. *Mucosal Immunol.* **8**, 1313–23 (2015).

212. Winkler, C. *et al.* Activation of group 2 innate lymphoid cells after allergen challenge in asthmatic patients. *J. Allergy Clin. Immunol.* **144**, 61-69.e7 (2019).

213. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–1 (2015).

214. Phillips, J. H. *et al.* Ontogeny of human natural killer (NK) cells: fetal NK cells mediate cytolytic function and express cytoplasmic CD3 epsilon,delta proteins. *J. Exp. Med.* **175**, 1055–66 (1992).

215. Shin, S. B. *et al.* Abortive γδTCR rearrangements suggest ILC2s are derived from T-cell precursors. *Blood Adv.* **4**, 5362 (2020).

216. Nagasawa, M., Germar, K., Blom, B. & Spits, H. Human CD5+ Innate Lymphoid Cells Are Functionally Immature and Their Development from CD34+ Progenitor Cells Is Regulated by Id2. *Front. Immunol.* **8**, 1047 (2017).

217. Wang, B. *et al.* Over-expression of CD3ε transgenes blocks T lymphocyte development. *Int. Immunol.* **7**, 435–448 (1995).

218. Yannoutsos, N. *et al.* The Role of Recombination Activating Gene (RAG) Reinduction in Thymocyte Development in Vivo. *J. Exp. Med.* **194**, 471 (2001).

219. Gool, F. Van *et al.* Interleukin-5–producing group 2 innate lymphoid cells control eosinophilia induced by interleukin-2 therapy. *Blood* **124**, 3572 (2014).

220. Korn, L. L. *et al.* Conventional CD4+ T cells regulate IL-22-producing intestinal innate lymphoid cells. *Mucosal Immunol.* **7**, 1045–57 (2014).

221. Roediger, B. *et al.* Cutaneous immuno-surveillance and regulation of inflammation by group 2 innate lymphoid cells. *Nat. Immunol.* **14**, 564 (2013).

222. Karo, J. M., Schatz, D. G. & Sun, J. C. The RAG recombinase dictates functional heterogeneity and cellular fitness in natural killer cells. *Cell* **159**, 94 (2014).

223. Bando, J. K. & Colonna, M. Innate lymphoid cell function in the context of adaptive immunity. *Nat. Immunol.* **17**, 783–9 (2016).

224. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

225. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).