Analysis of the transcriptome of human NK lymphocytes

Dissertation

for the award of the degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program Molecular Biology Program
of the Georg-August University School of Science (GAUSS)

submitted by
Olena Zaitseva

born in Kyiv, Ukraine

Göttingen 2016

Thesis Committee

Prof. Lutz Walter
Primate Genetics Laboratory
German Primate Center, Göttingen

Prof Jörg Stülke
Department of General Microbiology
Institute for Microbiology and Genetics, Göttingen

Prof. Matthias Dobbelstein
Department of Molecular Oncology
University Medical Center, Göttingen

Members of the Examination Board

Referee: Prof. Lutz Walter
Primate Genetics Laboratory
German Primate Center, Göttingen

2nd Referee: Prof. Jörg Stülke
Department of General Microbiology
Institute for Microbiology and Genetics, Göttingen

Further members of the Examination Board

Prof. Matthias Dobbelstein
Department of Molecular Oncology
University Medical Center, Göttingen

Prof. Steven Johnsen
Clinic for General, Visceral and Pediatric Surgery
University Medical Center, Göttingen

Prof. Markus Bohnsack
Department of Molecular Biology
University Medical Center Göttingen

Prof. Holger Reichardt
Department of Cellular and Molecular Immunology
University Medical Center Göttingen

TABLE OF CONENTS

LIST OF ABBREVIATIONS

ADCC	Antibody dependent cellular cytotoxicity
ANOVA	Analysis of variance
APC	Antigen presenting cell
BM	Bone marrow
BSA	Bovine serum albumin
CB	Cord blood
ChiP	Chromatin immunoprecipitation
CLP	Common lymphoid progenitor
cNK	Conventional natural killer cells
DC	Dendritic cells
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence activated cell sorting
fpkm	Fragments per kilobase of exon per million fragments mapped
GO	Gene ontology
HC	Hierarchical clustering
HLA	Human leukocyte antigen
HSC	Hematopoietic stem cells
IFC	Integrated Fluidic Circuits
Ig	Immunoglobulin
IL	Interleukin
ILC	Innate lymphoid cell
ITAM	Immunoreceptor tyrosine based activation motif
ITIMs	Immunoreceptor tyrosine based inhibition motifs
KIR	Killer-cell immunoglobuline-like receptor
LMPP	Lymphoid committed or primed progenitor
LN	Lymph nodes
lncRNA	Long non-coding RNA
LoD	Limit of detection
MC	Monocyte
MHC	Major histocompatibility complex
miRNA	Micro RNA
NCR	Natural killer cell receptor
NK	Natural killer cell
NKP	Natural killer cell restricted progenitor
PB	Peripheral blood
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCA	Principal component analysis
SLT	Secondary lymphoid tissues
snoRNA	Small nucleolar RNA

ACKNOWLEDGEMENTS

I am deeply grateful to Prof. Dr. Lutz Walter for giving me the opportunity to work on this project and guiding me through it. Thank you for teaching, advising and encouraging me during all these years.

I want to thank my Thesis Committee members Prof. Dr. Matthias Dobbelstein and Prof. Dr. Jorg Stülke for advises and support throughout my PhD project. I also thank Prof. Dr. Steven Johnsen, Prof. Dr. Markus Bohnsack and Prof. Dr. Holger Reichardt for joining my extended thesis committee.

This project became possible due to our collaboration partners. Samples representing cell populations were obtained from the group of Prof. Dr. Markus Uhrberg (University Clinic of Düsseldorf) and the group of Prof. Salim Kakoo (University of Southampton); RNA-seq was performed in the group of Dr. Gabriela Salinas-Riester.

I am grateful to my colleague Angela Noll as well to my former colleague Markus Brameier for the support in statistical analysis. I also want to thank my colleagues Beatrix Petersen, Christiane Schwarz, Christian Roos, Franziska Aron, Mechthild Pohl, Nico Westaphal not only for the practical help, but also for creating a friendly atmosphere in the group and for their help in improving my German language.

I cannot appreciate enough the support of Dr. Steffen Burkhardt and Kerstin Grüniger, who gave me an opportunity to study in Göttingen took care of ne during these years.

Many-many thanks to my family and friends, who have been supporting me on all this long way from going to Germany for Master's studies till now. They believed in me even when I was completely desperate and were with me in bad and good times.

ABSTRACT

Natural Killer cells are large granular lymphocytes that contribute to both innate and adaptive immunity, as they can not only lyse infected cells or tumors, but also stimulate an adaptive immune response via production of cytokines and chemokines. NK cells function at the frontline of defense against intracellular pathogens and transformed cells due to their ability to respond rapidly and without prior sensitization on multiple signals.

Traditionally the knowledge about NK cell function and development was obtained mainly from mouse models, transplantations and *in vitro* studies, but development of high-throughput methods of gene expression analysis in the last years shed more light on NK cell developmental and functional diversity as well as on ontological relations between their subpopulations. However, our knowledge about NK cells remains incomplete, especially in comparison to T and B lymphocytes, and many questions in this field are yet to be answered.

Together with our collaborators, our group collected an RNA-seq data set that, to our knowledge, is the most complete available for NK cell populations. This data set was comprehensively analyzed to elucidate the specific feature of transcriptomes of developing and mature NK cells.

Typical transcription signature of NK cells that distinguishes them from other peripheral blood mononuclear cells (PBMCs) and changes in expression profiles of NK cells during their development was described. In particular, a set of novel genes was identified in which specific expression or function in NK cells was not described before. Differences in expression profiles of NK cells from tissues and functional groups were also described, in particular the relative immature phenotype of bone marrow-derived stage 3 NK populations in comparison to their counterparts in tonsils.

Further, more mature and activated phenotype of NK cells produced by *in vitro* differentiation assays in comparison to their *ex vivo* counterparts was stated, which can be explained by using a cytokine cocktail for proper growth and development of cells *in vitro*.

In addition, expression levels of non-protein coding genes expressed by NK cells were estimated and the set of non-protein coding genes with regulatory potential in development and function of NK cells (e.g. *LINC00299, RP11-222K16.2, RP11-212I21.4, RP11-456D7.1, RP11-121A8.1*) was identified.

Furthermore, the single cell expression analysis of mature human NK lymphocytes using the microfluidic Fluidigm© technique was established and a pilot experiment was

performed, which allows further estimation of heterogeneity level of mature NK cell, identification of their functional subpopulations, and studying of regulatory gene networks within them.

1.INTRODUCTION

Natural Killer (NK) cells are large granular lymphocytes, which are traditionally described as CD3-negative, CD56-positive by surface phenotype (Kiessling et al., 1975; Robertson et al., 1990). Their function contributes to both innate and adaptive immunity, as they can not only lyse infected cells or tumors, but also stimulate an adaptive immune response via production of cytokines and chemokines (Herberman et al., 1975; Kiessling et al., 1975; Seaman et al., 1987). In the adult human organism NK cells constitute 5 to 15 % of peripheral blood, and also can be found in bone marrow, thymus, secondary lymphoid tissues (e.g. tonsils, lymph nodes, mucosa-associated lymphoid tissue, spleen) as well as in other organs such as liver, lungs, uterus, pancreas and joints (tissue-resident NK cells).

In spite of NK cells being discovered more than 40 years ago (Kiessling et al., 1975; Murphy et al., 1987), they are still studied much less than T and B lymphocytes and most of the knowledge about NK cell function and development is obtained from mouse models, transplantations and *in vitro* studies.

1.1. Innate lymphoid cell lineage

According to the recently established nomenclature, NK cells are placed in a diverse group of innate lymphoid cells (ILCs). ILCs are a family of cells involved in immunity as well as in tissue development and remodeling. Based on the cytokines that they produce and the transcription factors critical for their development and function, ILCs are separated into three groups: ILC1, ILC2 and ILC3 (Spits, 2013).

Group 1 ILCs are characterized by production of IFN-g, but not Th2-cell- and Th17 cell-associated cytokines (e.g. interleukin (IL)-17 and IL-22), and by dependence on T-bet and EOMES transcription factors during development. In mice this cell group expresses the cellular receptors NKp46 and NK1.1, while in humans their characteristic cell surface markers are CD16, CD56 and CD94 (Crellin et al., 2010b; Cupedo et al., 2009). According to these characteristics, NK cells were included into this group. However, a new nomenclature proposal puts NK cells in a separate group named cytotoxic ILCs as opposed to all other so-called helper-like ILCs (i.e. ILC1, ILC2, ILC3 groups) (Diefenbach et al., 2014). In the last years more progress has been made in differentiation of conventional NK (cNK) cells from other ILC1 populations based on differential expression of lineage-specific transcription factors (Ribeiro et al., 2010; Vosshenrich et al., 2006). Thus, differentiation of cNK cells is controlled by the transcription factors ID2, NFIL3 and EOMES, and in addition requires IL-15, while thymic NK cell development depends on GATA-3 transcription factor and requires IL-7 (Ribeiro et al., 2010; Vosshenrich et al.,

2006). Development of another subset of NK cells (TRAIL^{+}VLA-1(CD49a)$^{+}$VLA-2^{-}) identified in the liver is EOMES-independent, but requires T-bet and IL-15 signaling, while NK cells found in the uterus are both T-bet- and EOMES-independent (Sojka et al., 2014). Besides this, NFIL3 is critical for the development of bone marrow-derived NK cells as well as splenic and thymic NK cells, while liver VLA-2^{-}EOMES^{-} NK cells do not require it for development (Male et al., 2012; Seillet et al., 2014).

Unlike ILC1, group 2 ILCs express Th2-cell associated cytokines and depend on the transcription factor GATA3 for their development and function. In mice this group also produces IL-5, IL-13, IL-6 and IL-9; during development they require IL-7 and the transcription factor RORα for development, and their activity is regulated by IL-25 and IL-33 (Hoyler et al., 2012; Koyasu and Moro, 2012; Moro et al., 2010; Wilhelm et al., 2011; Yang et al., 2013).

Group 3 ILCs are characterized by production of Th17-related cytokines (IL-17A and IL-22) and by dependence on RORγt that is encoded by the RORC gene for their development and function and they are also dependent on IL-7Rα. The prototypical ILC3 cell group are LTi (lymphoid tissue-inducer cells), which require RORγt and AHR during development and produce both IL-17 and IL-22 upon activation (Takatori et al., 2009). However, the ILC3 lineage is rather heterogenic and at least two other functionally different subgroups were identified (Luci et al., 2009; Sanos et al., 2011; Satoh-Takavama et al., 2008): (1) ILC3 cells that do not express NK cell receptors (NCR^{-}) depend on RORγt, but not on AHR during development; in mice they are able to produce not only IL-17 and IL-22, but also IFN-γ, while in humans they produce mainly IL-17A (Hoorweg et al., 2012; Luci et al., 2009; Sanos et al., 2011; Satoh-Takavama et al., 2008), and (2) human NCR-positive ILC3 cells, also known as NCR22 or NK22 cells, that express NKp30, NKp44 as well as NKp46 and are able to produce IL-17A and IL-22. In mouse models they were shown to produce only IL-22, but in humans NCR^{+} ILC3s were shown to produce both IL-17A and IL-22. Development of these cells is dependent on RORγt and AHR, but also on T-bet (as well as development of NK cells and other ILC group 1 cells) (Cella et al., 2009; Satoh-Takayama et al., 2010).

1.2. Functions of NK cells

NK lymphocytes belong to the innate immune system and are at the frontline of defense against intracellular pathogens (of viral as well as bacterial nature) and transformed cells due to their ability to respond rapidly and without prior sensitization on multiple signals (Scharton-Kersten et al., 1996). Predominant mechanisms through which

NK cells implement their functions are cell cytotoxicity and production of cytokines (Arase et al., 1996; Robertson et al., 1990).

1.2.1. Cytokine and chemokine production

The most important NK cell-derived cytokine is IFN-γ, which activates a Th1 response and thereby triggers the adaptive immune response. Furthermore, NK cell-derived IFN-γ activates antigen-presenting cells (APCs), stimulating them to upregulate major histocompatibility complex (MHC) class I expression and cytokine expression. It also activates macrophage killing of intracellular pathogens.

Consistent with their function, in secondary lymphoid tissues NK cells are localized close to dendritic cells (DCs), where activated NK cells participate in DC maturation by producing TNF-a and GM-CSF, while DCs stimulate NK cell activation and expansion via cytokine production and secretion (e.g. IL-15, IL-12, IL-1, IL-18).

In addition, IFN-γ production by NK lymphocytes plays a role in human reproduction. When an embryo implants in the uterine wall, fetal trophoblast cells activate maternal NK cells by expression of a unique combination of human leukocyte antigen (HLA) molecules: HLA-C, HLA-E and HLA-G, which are recognized by NK activating receptors. An activation of these NK cells leads to production of IFN-γ, which is important for building of placenta and proper blood supply for a developing embryo (Rajagopalan and Long, 1999; Trundley and Moffet, 2004; Witt et al., 2002).

1.2.2. Cytotoxicity

Cellular cytotoxicity. NK cells mediate cellular cytotoxicity by activation of apoptosis in target cells via either the action of perforins and granzymes, TNF-alpha release, or death receptors (e.g. TRAIL and Fas-related pathways). Cellular cytotoxicity is a predominant function of the $CD56^{dim}$ subset, while $CD56^{bright}$ cells have low cytotoxic activity (Lanier et al., 1986). NK cellular cytotoxicity plays a role in antiviral defense, as virus-infected cells can be lysed prior to virus replication and virion assembly (Ahmad and Menezes, 1996; Biron et al., 1997). Besides this, NK cells are known to be important for surveillance and elimination of tumors, as they can recognize multiple markers of cellular stress and rapidly react by cellular cytotoxicity.

Antibody-dependent cellular cytotoxicity. Antibody-dependent cellular cytotoxicity (ADCC) is the process of recognition and lysis of opsonized target cells. Pathogen's antigens presented on the surface of target cells are recognized by IgG antibodies, which in turn are recognized by Fc gamma receptors on the surface of effector cells (in case of NK cells – by CD16, low affinity Fc gamma RIII). Thus, antibodies

function as a bridge between target and effector cells and provide specificity to the clearance reaction (Biron et al., 1997).

Antibody-independent cellular cytotoxicity (natural cytotoxicity) requires more complicated regulation by a certain balance of signals from inhibitory and activating receptors (Lanier, 2000). Normal cells express different combinations and levels of ligands that engage both stimulatory and inhibitory NK cell receptors (described below in more detail). The affinity of activating receptors to their ligands is generally lower than that of their inhibitory counterparts, so under normal conditions combined signaling from both receptor groups does not lead to NK cell activation. Most of these ligands are MHC class I molecules, which are often downregulated in infected or transformed cells in order to avoid recognition by $CD8^+$ T lymphocytes. This, however, makes affected cells to targets for NK lymphocytes, as resting NK cells become activated when they encounter target cells with low or missing expression of MHC class I molecules. Such recognition and lysis of target cells is called "missing-self recognition" (Karre et al., 1986; Ljunggren and Karre, 1990;).

Mechanism of cell recognition and lysis. Activation of resting NK cells is a complex process that requires combination of several events: cell contact, adhesion (formation of immune synapse), cytotoxic granule polarization and degranulation (exocytosis), and requires simultaneous engagement of multiple NK cell receptors (Leibson, 1997). Contact and adhesion provide a stable contact between NK and target cell and are required for formation of immunological synapse. Interaction of integrins with their ligands at this stage is regulated dynamically, as not only stable contact but also quick release from adherence is necessary for lymphocyte movement (Barber et al., 2003; Bryceson et al., 2005). Polarization of cytotoxic granules (secretory lysosomes) is accompanied by polarization of actin cytoskeleton, microtubules, and Golgi apparatus in the direction of target cell interface. Unlike T cell polarization (which in case of T cells is also required for cytotoxicity), polarization of NK cells was shown to be a relatively complicated stepwise process with a series of checkpoints (Bryceson et al., 2005; Wulfing et al., 2003). Finally, granules are exocytosed at the degranulation stage, releasing cytotoxic effector molecules such as perforin and granzymes to the interface space. When uptaken by a target cell, these molecules lead to its death by inducing apoptosis. Degranulation is a Ca^{2+} dependent process and is performed through PKC and G-protein-dependent pathways (Bryceson et al., 2005; Perez et al., 2004; Wulfing et al., 2003).

1.3. NK cell receptors

Regulation of the activation of resting NK cells is orchestrated by a balance of signals between numerous activating and inhibitory NK cell receptors recognizing MHC class I or class-I-like molecules on the surface of target cells.

Inhibitory receptors ensure self-tolerance to healthy cells and can recognize either different epitopes of self-MHC class I molecules (e.g. killer-cell immunoglobulin-like receptors (KIRs) and NKG2A) or non-MHC ligands (e.g. KLRG1 and NKR-P1) (Lee et al., 1998; Ravetch and Lanier, 2000). All of these receptors use the same signal transduction mechanism via cytoplasmic immunoreceptor tyrosine-based inhibition motifs (ITIMs). Upon engagement of a ligand, the ITIM is phosphorylated and can recruit SHP-1 (phosphatase Src homology 2-domain containing phosphatase 1). NK cells acquire expression of inhibitory receptors stochastically; from 10-20 inhibitory receptors encoded by the genome, each individual NK cell expresses only a subset of three to five receptors in average (Burshtyn et al., 1996; Ravetch and Lanier, 2000).

Activating receptors also contribute to NK cell specificity. Unlike inhibitory receptors, most of the activating receptors are expressed by all NK cells. Also in contrast to inhibitory receptors activating ones do not share a common mechanism of signal transduction: while KIRs and NKG2C interact with the adaptor protein DAP12, which contains an immunoreceptor tyrosine-based activation motif (ITAM), other activating receptors such as NKp30, NKp44, NKp46 and CD16 are associated with the ITAM-containing adaptor proteins CD3Z or FCERG, which transmit signals by recruitment of tyrosine kinase Syk and ZAP-70 (Kaiser et al., 2005; Stewart et al., 2005). One more group of activating NKRs including NKG2D, CD2-family members (CD2, CD244, CD319) and DNAM-1 do not contain ITAMs and do not involve ITAM-carrying adaptors and transduce signals by other various mechanisms (Bottino et al., 2005; Storkus et al., 1991). In particular, NKG2D transduce signals via interaction with DAP10 adaptor molecule, DNAM-1 contains an immunoreceptor tyrosine tail–like motif coupling DNAM-1 to downstream effectors, and CD244 signal transduction requires immunoreceptor tyrosine-based switch motif-containing adaptors that interact with its cytoplasmic domain (Gilfillan, 2002; Veillette et al., 2007; Zhang et al., 2015;)

No single receptor-ligand interaction is sufficient for the complex process of NK cell activation and elimination of target cells; there is a need of synergetic signaling from a bunch of different activating receptors. Only the cooperation between specific receptor combinations, such as NKp46 and CD2, NKp46 and NKG2D, NKp46 and 2B4, NKp46

and DNAM-1, NKG2D and 2B4, 2B4 and DNAM-1 induce degranulation associated with efficient target cell killing (Long et al., 2013). Furthermore, even though stimulation of activating receptors alone can activate cytotoxicity in resting NK cells, it is much more efficient upon combination with IL-stimulation (e.g. IL-2 or IL-12) (Bryceson et al., 2006).

1.4. Development of NK cells

Unlike T and B lymphocytes whose differentiation pathways have been already comprehensively described, the exact scheme and location of NK cell differentiation remains not completely clear.

Bone marrow (BM) is thought to be the main source of early NK cell developmental stages. However, BM-derived early NK cell progenitors are believed to traffic through the blood stream to secondary lymphoid tissues (SLT), where commitment to the NK cell lineage and maturation occur (Luther et al., 2011; Warner et al., 2012). Already in 2005 Freud and colleagues showed that BM-derived $CD34^{+}$ hematopoietic stem cells (HSC) migrate to lymph nodes (LN), where they develop to $CD56^{bright}$ NK cells. Later it was suggested (Eissens et al., 2012) that pre-NK cells undergo commitment in LN and migrate further to the spleen where their receptor repertoire develops until mature NK cells are released into the blood stream. However, detection of at least small amounts of all developmental stages in BM and LN suggests, that *in situ* differentiation also occurs outside of the main trafficking way. NK cell progenitors are also found in SLT (tonsils, thymus and liver), suggesting that at least some NK cells can also develop extra-medullary. It was also reported, that many of the mature NK cell populations that developed in extra-medullary environment differ from most circulating NK cells in peripheral blood (described below) (Eriksson et al., 2004; Koopman et al., 2003). However, it is not yet clear, whether these differences are caused by immunophenotypic changes occurring during NK cell maturation or by tissue-specific development from common precursors.

The development pathway starts from $Lin^{-}CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}CD49f^{+}$ HSC which gives rise to the lymphoid and myeloid cell lineages, followed by Lin-CD34+CD38-CD90+CD45RA+ lymphoid-commited or primed progenitor (LMPP), which is a common progenitor for lymphoid cell lineages, including T, B, NK cells and ILCs. The LMPP is followed by a common lymphoid progenitor (CLP), characterized as $Lin^{-}CD34^{+}CD38^{-}CD123^{-}CD45RA^{+}CD7^{+}CD10^{+}$, downstream in the developmental hierarchy (Freud et al., 2006).

The first progenitor downstream of LMPP and CLP with developmental potential restricted to the NK cell lineage (NK cell-restricted progenitor, or NKP) was recently

described by Renoux and colleagues (2015). This cell type was shown to be present in fetal tissues as well as in the adult human organism (in bone marrow, umbilical cord blood and tonsils) and is phenotypically characterized as $Lin^{-}CD34^{+}CD38^{+}CD123^{-}$ $CD45RA^{+}CD7^{+}CD10^{+}CD127^{-}$. Unlike the lymphocyte progenitors described before, NKP were shown to give rise to functional NK cells, but not to T- or B-lymphocytes, and not to ILC3s.

According to the classical model of NK cell development, established in 2005-2006, NK cells arise from CLP in five stages, which are distinguished based on the surface expression of CD34, CD117, CD94 and CD56 (Freud et al., 2006). Later studies also revealed CD133, CD33, CD244 and NKG2A as important developmental markers (Eissens et al., 2012).

The first two stages of NK cell development are characterized by commitment to NK cell lineage. Stage 1 (NK progenitor, or pro-NK) is characterized by a $CD34^{+}CD117^{-}$ $CD94^{-}$CD45RA+CD10+CD161- phenotype and is multipotent (Cooper et al, 2009). Besides NK cells, it gives rise to T lymphocytes and plasmacytoid DCs. The part of stage 1 cells with high CD34 expression is CD10-positive and was supposed to be the least mature population of NK cells in adult BM and SLT (Freud et al., 2006).

Stage 2 (preNK) is characterized by $CD34^{+}CD117^{+}CD94^{-}CD45RA^{+}CD10^{-}$ $CD161^{+/-}$ phenotype. All stage 2 cells typically express CD33, CD44, HLA-DR, integrin β7 and do not express CD1a, CD5 or CD123, but are heterogenic regarding expression of certain surface antigens, such as CD2, CD7, CD10, CD56, and CD161 (Freud et al., 2006). It was shown that every stage 2 cell has a capacity to become an NK cell, however cells from this population can still give rise to T cells and DCs. Stages 2 to 4 express high levels of the high-affinity IL-2 receptor, which allows them to successfully compete for picomolar levels of T-cell-derived IL-2, which has an important role for NK cell development in SLT (Freud et al., 2006).

Stage 3 to stage 5 NK cell maturation is characterized by gradual loss of CD34 and CD117 (Kit) expression, while surface expression of CD94, CD16 and KIR molecules increases (Eissens et al., 2012).

Stage 3 (immature NK cells) cells are $CD34^{-}CD117^{+}CD94^{-}$ which are usually described as the first population committed to NK cell development, as they are unable to give rise to any other lymphoid lineages (Freud et al., 2006). Indeed, they all express (on variable levels) NK-associated molecules, such as CD2, CD7, CD56, CD161, 2B4, and NKp44, but lack expression of genes associated with other lineages. In contrast to stage 2

lymphocytes, they lack the surface molecules CD10, integrin beta 7 and HLA-DR, but show high expression of IL-7Rα (Freud et al., 2006). They lack two key marks of mature NK cells: the ability to mediate cellular toxicity and IFN-g production, however they contribute to immunity via production of GM-CSF and type 2 cytokines and/or mediating TRAIL-dependent cell death (Barton et al., 1998; Freud et al., 2006). According to the data of Eissens and colleagues (2012), stage 3 cells can be divided into two substages: first, CD34 expression is lost together with multipotency at stage 3a ($CD34^-CD117^+CD56^-CD94^-$), secondly NK cell commitment is acquired together with CD56 expression at stage 3b ($CD34^-CD117^+CD56^+CD94^-$).

However, lately it has been shown, that the $CD34^-CD117^+CD94^-$ stage 3 population includes phenotypically overlapping, but functionally distinct cells: stage 3 NK cell progenitors themselves that express LFA-1 and are able do differentiate into mature NK cells upon IL-15 stimulation, ILC3 cells expressing RORC, IL-17 and/ or IL-22 and a minor population of ILC2 cells expressing RORα, GATA3 and type 2 cytokines (e.g. IL-5 and IL-13). While ILC2s can be easily identified due to a CD117-CD161+ surface phenotype, differentiation between NK stage 3 cells and ILC3s based on surface markers remains complicated (Ahn et al., 2013; Crellin et al., 2010b; Cupedo et al., 2009; Hughes et al., 2010; Mjösberg et al., 2011; Satoh-Takayama et al., 2011; Spits and di Santo, 2011).

Transition of cells from stage 3 to stage 4 is characterized by acquisition of CD94-NKG2A, NKG2D and NKp46 expression accompanied by downregulation of CD117, CD33 and CD127, so stage 4 NK lymphocytes are usually described as $CD34^-CD117^{+/-}CD94^+CD16^-$. Cells of both stage 4 and 5 express CD56 molecules on the cell surface, however cells of stage 4 are characterized by higher expression of CD56, while stage 5 cells show a low density of CD56. For this reason stage 4 and stage 5 NK cells are also called $CD56^{bright}$ and $CD56^{dim}$ NK cells, respectively. Besides this, stage 4 NK lymphocytes express increasing levels of the transcription factor T-bet (TBX21) and the high-affinity heterotrimeric IL-2 receptor alpha/beta/gamma and can thereby respond to picomolar IL-2 concentrations (Lanier et al., 1986).

According to the data shown by Eissens and colleagues (2012), $CD56^{dim}$ cells can be subdivided into two developmental stages: stage 5a is characterized by downregulation of the CD117 expression ($CD34^-CD117^-CD56^+CD94^+$) which is followed by a loss of CD94 in the course of the development to stage 5b ($CD34^-CD117^-CD56^+CD94^-$). Additionally, stage 5 NK cells acquire expression of KIRs, cytolytic molecules (perforin 1

and granzymes) and chemokines, such as CXCL8 and Mip-1b (Eissens et al., 2012; Hanna et al., 2004; Koopman et al., 2003; Wendt et al, 2006; Yu et al., 2010).

Both $CD56^{bright}$ and $CD56^{dim}$ NK cell subsets play an important role in immunity, however they differ dramatically in their cytotoxic potential, their capacity of cytokine production and their responses to cytokine activation and have different functions. They also have a different tissue distribution that is while the $CD56^{dim}$ subset is predominantly found in peripheral blood, $CD56b^{right}$ NK cells dominate in SLT, neonatal tissues and cord blood (Jakobs, 2001; Lanier, 1986). In the spleen the ratio between $CD56^{dim}$ and $CD56^{bright}$ cells is about 1:1. Consistent with their localization, $CD56^{bright}$ NK cells express CD62L and the chemokine receptors CXCR4, CCR5 and CCR7, which are involved in lymphocyte trafficking, allowing these cells to migrate to SLT. CD56dim NK cells express instead higher levels of CXCR1, CXCR2, CXCR3, CXCR4 and CX3CR1, which enables these cells to be recruited to sites of inflammation (Campbell et al., 2001; Eissens et al., 2012; Fehniger et al., 2003; Ferlazzo et al., 2004; Fu et al., 2014; Hanna et al., 2003).

In comparison to the $CD56^{dim}$ population, $CD56^{bright}$ cells show lower expression levels of KIRs, but higher levels of C-type lectin receptors (CD94/NKG2 heterodimers). Their cytotoxic potential is lower than that of $CD56^{dim}$ cells and is acquired only after prolonged activation (Bjorkstrom et al., 2010); however, they are more potent in cytokine secretion (first of all, IFN-γ, and TNF-a). Furthermore, $CD56^{brigh}$t cells have the ability to express high levels of GM-CSF, TARC, TGF-β3, IL-10, IL-8 and IL-13 following monokine stimulation, while $CD56^{dim}$ cells show higher expression of IGF-1 and IGFBR-3 (Cooper et al., 2001; Hanna et al., 2004). To produce IFN-γ they need to accumulate two activating signals: first, they need IL-12 and secondly either another IL (e.g. IL-1, IL-2, IL-15, IL-18) or the engagement of activating receptors (e.g. CD16 or NKG2D) is necessary. In addition, $CD56^{bright}$ cells express the c-Kit receptor tyrosine kinase (alias CD117), which, together with IL-2 signaling, allows them to proliferate upon activation (Bryceson et al., 2006; Wang et al., 2014).

$CD56^{dim}$ NK cells demonstrate much higher cytotoxicity upon activation. In comparison to the $CD56^{bright}$ subset, they upregulate KIRs, cytotoxic molecules and chemokines. Notably, in resting NK cells, mRNA of granzymes and perforin is present, however translation is blocked; upon activation the mRNA level does not change, but the protein level increases. Recently it was shown that CD56dim cells can undergo further differentiation and functional modifications (Chan et al., 2007; Romagnani et al., 2007;). In contrast to $CD56^{bright}$ cells, resting $CD56^{dim}$ cells are c-Kit negative and are almost

unable to proliferate even upon stimulation with high IL-2 dosages. Instead of this, they respond to IL-2 activation by increased cellular cytotoxicity. (Robertson et al., 1992) Almost all $CD56^{dim}$ NK cells express high levels of CD16 and are capable for ADCC, unlike $CD56^{bright}$ cells (Cooper et al., 2001; Romagnani et al., 2007).

Memory-like NK cells are believed to be a final stage of NK cell development, characterized by enhanced effector response upon secondary treatment with the same antigen. In comparison to naïve NK cells they are able to produce higher amounts of IFN-γ (Miller, 2013; O'Leary et al., 2006; Sun et al., 2009). In a mouse model it was shown that memory-like NK cells could be derived from both $CD56^{bright}$ and $CD56^{dim}$ populations upon cytokine stimulation, exposure to multiple viruses and hapten treatment, however it is not completely clear, whether they can arise only at specific points of development, or may occur from different NK developmental stages depending on environmental stimuli (Cooper et al., 2009; Min-Oo et al., 2013; Paust et al., 2010; Romee et al., 2012). Mouse memory-like NK cells showed higher expression of *Ly49H, Ly6C, Cd43*, and *Klrg1* in comparison to naïve NK cells, while expression of Cd27 was decreased. Besides this, liver-restricted NK memory cells expressed high levels of *Thy1, Cxcr6, Cd49a* and/or *Ly49C/I*. An existence of antigen-specific NK cell memory was also proved in rhesus macaques (Reeves et al., 2015). In humans a corresponding memory-like NK cell population was described even less, but increased expression of CD57, CD62L and NKG2C are supposed to mark virus-induced memory NK cells (Juelke et al., 2010; Lopez-Verges et al., 2010; Sun et al., 2009).

1.5. Licensing of NK cells and self-tolerance

The simple acquisition of specific surface receptors is not enough for NK cells to become completely functional. In addition, they have to undergo a MHC class I-dependent process called "education" or "licensing", which is different from education of T-lymphocytes. This process ensures the tolerance of NK cells meaning that they do not attack healthy cells, but are able to identify efficiently pathogen-infected or transformed cells (Dofman et al., 1997; Kim et al., 2005; Liao et al., 1991; Yokayama and Kim, 2006). As mentioned above, specific ITIM-bearing inhibitory receptors, namely KIRs and CD94/NKG2, recognizing MHC class I molecules are necessary to inhibit NK cell activation upon recognition of cognate ligands. The polymorphic KIR family segregates independently from the MHC locus and KIR proteins are expressed in a clonal manner on individual cells. As KIR proteins show specificity for different MHC class I proteins and even for different MHC class I allotypes, there exist NK cells expressing a KIR receptor,

which does not recognize a MHC class I protein in that individual. To avoid chronic activation of such NK cells due to the missing-self recognition, such NK cells become hyporesponsive and are classified as non-licensed. Inversely, those NK cell expressing KIR proteins capable of interacting with MHC class I are responsive and are classified as licensed (Yokoyama and Kim, 2006). Therefore, in many individuals only a portion of NK cells possess inhibitory KIRs recognizing self-MHC class I molecules (Anfossi et al., 2006; Fauriat et al., 2010; Sim et al., 2015). Although the exact molecular mechanism of NK cell licensing is unclear, it is generally appreciated that it depends on appropriate pairing of inhibitory receptors with their MHC ligands, which is followed by signal transduction. The number of inhibitory receptors engaged by MHC class I molecules and the strength of binding was shown to influence NK cell responsiveness for cytotoxicity and cytokine production (Brodin et al., 2009a; Joncker et al., 2009): the stronger the interaction is between inhibitory receptor and its ligand, the stronger is the effector response of the NK cell. It is argued that licensing occurs during transition from stage 4 to stage 5 in NK cell development (Cruz-Munoz and Veillette, 2010; Kim et al., 2005; Yokoyama and Kim, 2006).

As a result of licensing, two major subpopulations of mature NK cells can be distinguished; both of them are self-tolerant, although in a different manner. Those cells that express one or more inhibitory KIR fitting to a self-MHC phenotype and/or NKG2A receptors become licensed. Due to inhibition through these receptors they remain inactive upon normal conditions. Cells lacking inhibitory receptors to self-MHC molecules remain unlicensed and are tolerant due to functional incompetence. Such cells show reduced cellular cytotoxicity, reduced ability to lyse tumor cells and defects in cytokine production (Dorfman et al., 1997; Fernandez et al., 2005). Even though some studies showed that unlicensed cells can be activated *in vitro* with the help of prior treatment (e.g. with phorbol ester and ionomycin or by cytokine stimulation *in vitro*), *in vivo* they are hyporesponsive and cannot be activated. (Kim et al., 2005; Yokoyama and Kim, 2006).

Four alternative models of the licensing mechanism were proposed: i.e. arming, disarming, cis-interaction and rheostat models. According to the arming model, interaction of NK cell receptors with cognate self-MHC class I molecules promote NK cell licensing in an active way by granting NK cells functional competence. NK cells that lack receptors against self-MHC-I molecules in accordance lack such an influence and remain hyporesponsive by default. According to some studies, signaling via ITIM-containing receptors indeed can lead to phosphorylation of signaling substrates downstream of the

ITIM, proving an active influence of such interactions (Cooley et al., 2007; Peterson and Long, 2008).

According to the disarming model, no active changes happen in NK cells upon interaction with cognate inhibitory targets and responsiveness is a default state of NK cells. While NK cells that possess receptors against self-MHC-I proteins avoid activation and show normal maturation, NK cells that lack corresponding receptors become self-tolerant due to persistent stimulation followed by cell anergy (Fernandez et al., 2005; Raulet et al., 2006).

The cis-interraction model was established in mice and is based on the ability of Ly49 receptors (mouse equivalent of KIRs) to bind MHC class I molecules on the same cell's membrane (Andersson et al., 2007; Doucey et al., 2004;). According to this model, interactions between MHC class I molecules and Ly49 receptors in cis sequesters Ly49 receptors, preventing their relocation to the immunological synapse and thus decreases signaling from activating receptors. However, it was not proven yet, whether all Ly49 receptors as well as human KIRs can engage MHC class I molecules in cis.

The rheostat model partly combines the evidence of the other three mechanisms. According to this model, NK cells have increased (arming-like) or decreased (disarming-like or partially due to cis-interactions) responsiveness depending on the strength of the inhibitory signal that is received, but are not completely "turned on" or "turned off", and that NK cell licensing happens in a quantitative manner (Brodin et al., 2009b; Johansson et al., 2005; Joncker et al., 2009).

A recently published report (Thomas et al., 2013) shed some light on the licensing process. It was shown that licensing increases the ability of NK cells to form an immunological synapse, resulting in the ability to adhere to a target cell via LFA-1. As was described above, LFA-1 is a receptor recognizing ICAM-1; its engagement was shown to be sufficient for NK cell adhesion to a target cell and it can be enhanced by so-called "inside-out" signals from NK cell activating receptors (NKp46 and 2B4) (Barber and Long, 2003; Bryceson et al., 2009). Thomas and colleagues showed that licensing does not influence LFA-1 binding to ICAM-1 itself and thereby an initial binding to the target cell, but signaling from activating receptors, which in turn increases strength and stability of LFA-1-mediated adhesion. These data support rather an active mechanism of licensing and, therefore, an arming model, however details of licensing mechanism remain unclear. In the line with this finding, it was later shown that DNAM-1 forms a functional pair with LFA-1 upon NK cell activation; and both DNAM-1 expression and LFA-1 conformational

changes correlate with licensing status of NK cells. Furthermore, DNAM-1 expression was shown to be the highest on terminally-differentiated memory-like NK cells (Enqvist et al., 2015). It was therefore suggested, that coordinated expression and LFA-1 conformational changes contribute to higher responsiveness of licensed NK cells. These data support rather an active mechanism of licensing and the arming model. However, details of the molecular mechanism underlying licensing still remain unclear.

1.6. Tissue-specific characteristics of NK cells

Tissue- and organ-specific features of NK lymphocytes remain an extensively studied but still complicated topic. First of all, as earlier mentioned, NK cells can develop extra-medullary and many of such mature NK cell populations have specific features. Furthermore, distribution of NK cells is not static because they can recirculate between organs; under normal conditions as well as in the case of inflammation NK cells can be recruited to specific organs where they acquire tissue-specific features (Jiang et al., 2004; Liu et al., 2006; Thapa et al., 2007).

Although it was suggested that variable characteristics of NK cells among different organs are due to their ability to adjust to microenvironment, the developmental location can also influence the NK cells' phenotype and functional specificity (Eriksson et al., 2004; Hamann et al., 2011; Koopman et al., 2003;).

The first and best-described NK cell subsets are those found in peripheral blood. They are usually seen as prototypical NK cells, to which cells from other tissues are compared. In peripheral blood, NK cells usually comprise 5-15% of the total amount of lymphocytes. About 90-95% of the peripheral blood NK cell population are $CD56^{dim}$ cells, and the remaining 5-10% are $CD56^{bright}$ cells. In SLT, in contrast, $CD56^{bright}$ cells constitute an absolute majority of the NK cell population.

It was postulated that NK cell surface phenotype and expression profiles depend on their developmental stage as well as on the local microenvironment (Wang et al., 2015). In particular, it was shown that expression of KIRs (e.g. KIR2DL2, KIR2DL3, KIR2DS2, KIR2DS3 and KIR3DL1) on $CD56^{bright}$ NK cells in lymph nodes and spleen was lower in comparison to corresponding populations in peripheral blood, cord blood and bone marrow, suggesting a more immature phenotype of $CD56^{bright}$ cells in these tissues. Vice versa, $CD56^{bright}$ NK cells derived from the uterus show high expression levels of KIRs and CD9, which was not typically seen in the $CD56^{bright}$ population in peripheral blood (Eriksson et al., 2004; Koopman et al., 2003). In another comparative study, it was shown that the proportion of NKG2A-expressing cells among $CD56^{dim}$ cells is the lowest in bone

marrow, cord blood and peripheral blood, higher (about 50%) in spleen and the highest in lymph nodes. Furthermore, the amount of $CD56^{dim}$ cells expressing activating receptors (besides NKp44) is lower in lymph nodes. Considering this and the fact that loss of NKG2A expression is a marker of NK cell maturation, it was suggested that NK cells developing in lymph nodes have a more immature phenotype in comparison to NK cells in other tissues (Eissens et al., 2012).

The same is also true for earlier stages of NK cell development. In peripheral blood, cord blood and spleen CD33 expression was observed until stage 3b during NK cell development, and even until stage 4 in liver LN, although CD33 was described as a marker specific for stage 2 in bone marrow. Thus it was suggested, that committed NK cells in the spleen and lymph nodes have a more immature phenotype in comparison to cells from other tissues (Eissens et al., 2012).

Even more specific features were described for **NK cell populations in the liver**. The liver is an important organ of the innate immune system, rich in NK cells, natural killer T cells, γδT cells and macrophages. It is an immunotolerant organ with a unique immunological microenvironment, where constant expression of gut-derived antigens does not lead to inflammation (Paust et al., 2010; Racanelli and Rehermann, 2006; Sun et al., 2013). NK cells comprise 30-50% of the lymphocyte population in the human liver and are localized mainly in the hepatic sinusoids adhering to endothelial cells (Sun et al., 2013; Yamagiva et al., 2009). Hepatic $CD56^{bright}$ NK cells were shown to have a more immature phenotype comparing to their counterparts in peripheral blood, probably due to a specific immunological environment characterized by hyporesponsiveness against gut-derived antigens (Eissens et al., 2012; Paust et al., 2010; Racanelli and Rehermann, 2006).

Furthermore, intrahepatic NK cells show an increased surface expression of CXCR6 in comparison to peripheral blood NK cells. $CXCR6^{+}$ NK cells from liver were shown to have an immature phenotype ($CD56^{bright}CD16^{low/-}CD57^{-}$) and correspond to the $CD56^{bright}$ subset in peripheral blood (Cooper et al., 2009). Such cells were shown to develop an antigen-specific memory against haptens and virus-derived antigens (Paust et al., 2010). $CXCR6^{-}$ liver cells have a more mature phenotype and correspond to $CD56^{dim}$ NK cells from the blood (Marquadt et al., 2015; Sun et al., 2012). Furthermore, while peripheral blood NK cells show high expression levels of T-bet and low levels of EOMES. Vice versa, the liver $CXCR6^{+}$ population has low expression of T-bet and high expression of EOMES as well as higher expression of CD49a (ITGA1) in comparison to their counterparts from peripheral blood (Marquadt et al., 2015; Paust et al., 2010).

1.7. The role of high-throughput methods in studying of lymphocyte expression profiles

Development of high-throughput methods of gene expression analysis in the last years allows comprehensive description of lymphocyte transcriptional profiles, better understanding of their development and activation and studying of novel potential roles of molecules in these processes. For these purposes methods such as large-scale mRNA microarrays, miRNA and DNA microarrays, multiparametric mass cytometry, FACS, ChiP-seq and RNA sequencing have been used (Bottcher et al., 2013; Casero et al., 2015; Chen et al., 2014; Choi et al., 2004; Du et al., 2006; Fu et al., 2014; Gard et al., 2012; Gillard-Bocquet et al., 2013; Kim and Lanier, 2013; Park et al., 2010). Furthermore, sophisticated bioinformatical tools, such as WGCNA, GeneMANIA, Inferelator, Cytoscape allow easier consolidation of available information and studying of given molecules in the context of their potential molecular interactions and their role in regulatory networks. Developing single cell technologies, including single cell mass cytometry, RT-qPCR and RNA sequencing, allow in addition analysis of population heterogeneity and individual cellular response on stimuli (Horowitz et al., 2013; Kouno et al., 2013; White et al., 2011;). Single cell gene expression analysis remains, however, challenging for NK cells due to their small size and low amount of RNA in the resting state.

Studies of NK cells performed in recent years using such methods shed more light on different aspects of NK cell biology. In particular, using genome-wide RNA-microarrays Bezman and colleagues demonstrate a close transcriptional relationship between NK cells and T cells (2012), and Dybkaer and colleagues described novel signal transductions pathways involved in NK cell activation by IL-2 (2007). Hanna and colleagues due to oligonucleotide microarrays described differences between gene expression profiles of $CD56^{bright}CD16^{-}$ and $CD56^{dim}CD16^{+}$ NK cells, as well as changes in $CD16^{+}$ NK cells' expression profiles upon activation (2004); whereas Fehniger and colleagues identified micro-RNA transcriptome of NK cells using next-generation sequencing (2010).

However, our knowledge about NK cells remains incomplete, especially in comparison to T and B lymphocytes (Casero et al., 2015; Chen et al., 2014) and many questions in this field remain unanswered.

Therefore, this research pursued the following scientific aims:

1. To identify the typical transcription signature of NK cells that distinguishes them from other PBMCs.
2. To define, how expression profiles of NK cells change during their development
3. To define, how expression profiles of NK cells differ between tissues and functional groups.
4. To estimate the level of heterogeneity of mature NK cell peripheral blood population at the single cell level.
5. To estimate the level of expression of non-protein coding genes expressed by NK cells.
6. To define, whether expression profiles of NK cells produced by *in vit*ro differentiation assays correspond to their natural counterparts.
7. To identify genes that are co-expressed in mature and developing NK cells.

2. MATERIALS AND METHODS

2.1. Chemicals, kits and consumables

ArrayControl RNA Spikes	Life Technologies
Biocol Seprating Solution	Biochrom AG
BSA	Sigma
C1 IFC for mRNA seq (5-10 µm)	Fluidigm
C1 Single cell Auto Prep Reagent Kit for mRNA Seq	Fluidigm
Conical tubes (50ml)	Greiner bio-one
EDTA	Roth
High Sensitivity DNA Assay kit	Agilent Technologies
$KHCO_3$	Merck
MS Columns	Myltenyi Biotec
NH_4Cl	Merck
NK Cell Isolation Kit	Myltenyi Biotec
PBS	Gibco
Pipette double-filter tips (2.5ul, 100ul, 1000ul)	Eppendorf
Reaction tubes (1.5ml, 2ml)	Eppendorf
Separation tubes Luecosep (50ml)	Greiner bio-one
SMARTer Ultra Low RNA Kit for the Fluidigm C1 System	Clontech
Sterile piettes (5ml, 10ml, 25ml)	Sarstedt

2.2. Buffers

MACS buffer: 0.5 % BSA and 2 mM EDTA in PBS.

Erythrocyte lysis buffer: 155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA, pH 8.0.

2.3. Equipment

BioAnalyzer 21000	Agilent Technologies
C1TM Single cell autoprep System	Fluidigm
Centrifuge Heraeus Multifuge 1S	Thermo Scientific
Centrifuge Heraeus Fresco 21	Thermo Scientific
MACS multy stand	Myltenyi Biotec
Microscope Observer A1	Zeiss
Pipette controller	Hirshmann pipetus

Pipettes "Research" series (2.5ul, 100ul, 1000ul)	Eppendorf

2.4. Software

SINGuLAR	Fluidigm
PANTHER Classification System 10.0	Paul Thomas ©
R	Bell Laboratories

2.5. Blood and tissue samples

Samples representing cell populations were obtained from our collaborators: Natural Immunity Group (University Clinic of Düsseldorf, Germany) and Faculty of Medicine (University of Southampton, UK).

Taking blood and tissue samples with subsequent cell isolation was conducted with the approval of ethical authorities. Sample collection in the University Clinic of Düsseldorf was approved by the ethical committee of the Heinrich Heine University Düsseldorf (study numbers 3756, 3240, 3880, and 4723); blood and bone marrow samples were obtained from healthy donors, while tonsils were resected due to medical recommendations. Liver sample collection was approved by the National Research Ethics Service (reference number 13/WA/0329).

In all cases cells were isolated by density gradient centrifugation using Ficoll medium and specific lymphocyte subpopulations were selected by fluorescence-activated cell sorting (FACS) based on known surface markers. Details about origin and sorting strategy of cell populations are summarized in Table 1.

The analyzed sample set included human NK cells (n=57), T lymphocytes (n=17), B lymphocytes (n=2), monocytes (n=2), and $CD34^+$ lymphocyte precursors isolated from peripheral blood, umbilical cord blood, bone marrow, tonsils, and liver. Besides this, NK cells from two independent *in vitro* differentiation assays were analyzed (n=9). These NK cells were generated from bone marrow-derived $CD34^+$ progenitors by induction with the combination of cytokines including IL-2, IL-3, IL-6, IL-7, IL-12, IL-15, IL-18, and trombopoetin. The differentiation assays were performed by the Natural Immunity Group (University Clinic of Düsseldorf).

Table 1 Characteristic of studied lymphocyte populations

Name	Source	Cell Type	Phenotype	Specific characteristic
pbNKst4_1	Peripheral blood	NK	$CD56^{bright}CD16^{-}$	Stage 4
pbNKst4_2	Peripheral blood	NK	$CD56^{bright}CD16^{+}$	Stage 4
pbNKst5_1	Peripheral blood	NK	$CD56^{dim}KIR2DL3^{+}$	Stage 5
cbNKst4	Cord blood	NK	$CD56^{bright}$	Stage 4
pbNKst5_2	Peripheral blood	NK	$KIR^{-}NKG2A^{-}CD56^{+}CD3^{-}$	Stage 5
pbNKst5_3	Peripheral blood	NK	$KIR^{-}NKG2A^{-}CD56^{+}CD3^{-}$	Stage 5
pbNKst5lic_1	Peripheral blood	NK	$KIR3DL1^{-}NKG2A^{-}KIR2DL1^{-}$ $KIR2DL3^{+}CD56^{+}CD3^{-}$	Stage 5 licensed
pbNKst5lic_2	Peripheral blood	NK	$KIR3DL1^{-}NKG2A^{-}$ $KIR2DL1^{+}KIR2DL3^{-}CD56^{+}CD3^{-}$	Stage 5 licensed
cbNKst5	Cord blood	NK	$CD56^{dim}KIR2DL3^{+}$	Stage 5
cbCD34pos_1	Cord blood	CD34	$CD34^{+}CD38^{+}$	Lymphocyte precursors
cbCD34pos_2	Cord blood	CD34	$CD34^{+}CD38^{+}$	Lymphocyte precursors
cbTCD4pos_1	Cord blood	T	$CD4^{+}$	T lymphocytes
cbTCD4pos_2	Cord blood	T	$CD4^{+}$	T lymphocytes
cbTCD8pos_1	Cord blood	T	$CD8^{+}$	T lymphocytes
cbTCD8pos_2	Cord blood	T	$CD8^{+}$	T lymphocytes
pbNKst5lic_3	Peripheral blood	NK	$KIR2DL3^{+}KIR2DL1^{+}$ $KIR3DL1^{+}$	Stage 5 licensed
cbCD34_1	Cord blood	CD34	$CD34^{+}$	Cultured without feeder layer
cbCD34_2	Cord blood	CD34	$CD34^{+}$	Cultured with feeder layer
pbNKst5lic_4	Peripheral blood	NK	$KIR2DL1^{+}$	Stage 5 licensed
pbNKst5lic_5	Peripheral blood	NK	$KIR2DL3^{+}$	Stage 5 licensed
pbNKst5lic_6	Peripheral blood	NK	$KIR2DL3^{+}$	Stage 5 licensed
pbNKst5lic_7	Peripheral blood	NK	$KIR2DL1^{+}$	Stage 5
daNKst2_1	Differentiation assay	NK	$CD34^{+}CD117^{+}CD56^{-}CD94^{-}$	Stage 2
daNKst3_1	Differentiation assay	NK	$CD34^{+}CD117^{+}CD56^{+}CD94^{-}$	Stage 3
daNKst4_1	Differentiation assay	NK	$CD34^{-}CD117^{-}$ $CD56^{+}CD94^{+}CD62L^{+}$/ $CD117^{bright}CD56^{+}CD94^{-}KIR^{-}$	Stage 4
daNKst5_1	Differentiation assay	NK	$CD117^{dim}CD56^{+}CD94^{+}$ KIR^{+}	Stage 5
pbTKIRpos_1	Peripheral blood	T	$CD4^{+}KIR2DL2^{+}$	KIR-positive
pbTKIRpos_2	Peripheral blood	T	$CD4^{+}KIR3DL2^{+}$	KIR-positive
pbTKIRneg_1	Peripheral blood	T	$CD4^{+}KIR^{-}$	KIR-negative
daNKst4_2	Differentiation assay	NK	$CD117^{bright}CD56^{+}CD94^{+}$ $CD62L^{+}$/$CD117^{bright}CD56^{+}$ $CD94^{-}KIR^{-}$	Stage 4
daNKst2_2	Differentiation assay	NK	$CD34^{+}CD3^{-}CD94^{-}CD117^{+}CD56^{-}$	Stage 2 day 7
daNKst3_2	Differentiation assay	NK	$CD34^{-}CD117^{+}CD56^{+}CD94^{-}$ $NKG2A^{dim}$	Stage 3 day 11
daNKst3_3	Differentiation assay	NK	$CD94^{-}CD117^{+}CD56^{+}NKG2A^{-}$ KIR^{-}	Stage 3 day 25

daNKst4_3	Differentiation assay	NK	$CD94^{+}CD117^{-}$ $CD56^{+}NKG2A^{+}KIR^{-}$	Stage 4 day 25
daNKst5_2	Differentiation assay	NK	$CD94^{+}CD117^{-}$ $CD56^{+}NKG2A^{+}KIR^{+}$	Stage 5 day 25
pbTCKIRneg_2	Peripheral blood	T	$CD4^{+}KIR^{-}$	KIR-negative
pbTKIRpos_4	Peripheral blood	T	$CD4^{-}KIR3DL2^{+}$	KIR-positive
pbTKIRpos_5	Peripheral blood	T	$CD4^{+}KIR3DL2^{+}$	KIR-positive stimulated
pbTKIRpos_6	Peripheral blood	T	$CD4^{+}KIR3DL2^{+}$ CD3/CD28	KIR-positive stimulated
pbTKIRneg_3	Peripheral blood	T	$CD8^{+}KIR^{-}$	KIR-negative
pbTKIRpos_7	Peripheral blood	T	$CD8^{+}KIR3DL2^{+}$	KIR-positiv
pbNKst5mem_1	Peripheral blood	NK	$CD56^{dim}NKG2A^{-}CD57^{-}$ $KIR2DL1^{+}NKG2C^{+}CD3^{-}$	Stage 5 memory-like
pbNKst5mem_2	Peripheral blood	NK	$CD56^{dim}NKG2A^{+/-}$ $CD57^{+}KIR2DL1^{+}NKG2C^{-}CD3^{-}$	Stage 5 memory-like
pbNKst5mem_3	Peripheral blood	NK	$CD56^{dim}NKG2A^{-}$ $CD57^{+}KIR2DL1^{+}NKG2C^{+}CD3^{-}$	Stage 5 memory-like
pbNKst5lic_8	Peripheral blood	NK	$CD56^{dim}NKG2A^{-}CD57^{-}$ $KIR2DL1^{+}NKG2C^{-}CD3^{-}$	Stage 5 licensed
pbNKst5lic_9	Peripheral blood	NK	$CD56^{dim}NKG2A^{+}KIR2DL1^{-}$ $KIR2DL3^{-}KIR3DL1^{-}CD3^{-}$	Stage 5 licensed
pbNKst5nonl_2	Peripheral blood	NK	$CD56^{dim}NKG2A^{-}$ $KIR2DL1^{+}KIR2DL3^{-}KIR3DL1^{-}$ $CD3^{-}$	Stage 5 nonlicensed
pbNKst5lic_10	Peripheral blood	NK	$CD56^{dim}NKG2A^{-}$ $KIR2DL1^{-}$ $KIR2DL3^{+}KIR3DL1^{-}CD3^{-}$	Stage 5 licensed
pbNKst4_3	Peripheral blood	NK	$CD56^{bright}$	Stage 4
pbNKst5nonl_3	Peripheral blood	NK	$CD56^{dim}NKG2A^{-}KIR2DL1^{-}$ $KIR2DL3^{-}KIR3DL1^{-}CD3^{-}$	Stage 5 nonlicensed
cbB	Cord blood	B	$CD19^{+}$	B lymphocytes
pbTKIRneg_4	Peripheral blood	T	$CD4^{+}KIR^{+}$	Stimulated
pbTKIRneg_5	Peripheral blood	T	$CD4^{+}KIR^{+}$ CD3/CD28	KIR-positive
pbNKst5mem_4	Peripheral blood	NK	$CD56^{dim}CD3^{-}$ $CD57^{+}KIR2DL3^{+}NKG2C^{-}$ $NKG2A^{-}KIR2DL1^{-}$	Stage 5 memory-like
pbNKst5mem_5	Peripheral blood	NK	$CD56^{dim}CD3^{-}CD57^{-}$ $KIR2DL3^{+}NKG2C^{+}NKG2A^{-}$ $KIR2DL1^{-}$	Stage 5 memory-like
pbNKst5lic_12	Peripheral blood	NK	$CD56^{dim}CD3^{-}CD57^{-}$ $KIR2DL3^{+}NKG2C^{-}NKG2A^{-}$ $KIR2DL1^{-}$	Stage 5 licensed
pbNKst5mem_6	Peripheral blood	NK	$CD56^{dim}CD3^{-}$ $CD57^{+}KIR2DL3^{+}NKG2C^{+}NKG2$ $A^{-}KIR2DL1^{-}$	Stage 5 memory-like
pbNKst5nonl_4	Peripheral blood	NK	$CD56^{dim}CD3^{-}CD57^{-}KIR2DL3^{-}$ $NKG2C^{-}NKG2A^{-}KIR2DL1^{+}$	Stage 5 nonlicensed
bmNKst1_1	Bone marrow	NK	$CD34^{+}CD117^{-}CD56^{-}CD94^{-}CD3^{-}$ $CD14^{-}CD19^{-}$	Stage 1
bmNKst2_1	Bone marrow	NK	$CD34^{+}CD117^{+}CD56^{-}CD94^{-}CD3^{-}$ $CD14^{-}CD19^{-}$	Stage 2
bmNKst3_1	Bone marrow	NK	$CD34^{+}CD117^{+}CD56^{-}CD94^{-}CD3^{-}$ $CD14^{-}CD19^{-}$	Stage 3
bmNKst4_1	Bone marrow	NK	$CD117^{-}CD34^{-}$ $CD56^{bright}CD94^{+}CD3^{-}CD14^{-}$ $CD19^{-}$	Stage 4

bmNKst5_1	Bone marrow	NK	$CD117^{-}CD34^{-}$ $CD56^{dim}CD94^{+}CD3^{-}CD14^{-}CD19^{-}$	Stage 5
cbNKst2	Cord blood	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{+}$	Stage 2
cbNKst3	Cord blood	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{-}CD94^{-}$	Stage 3
cbNKst4_2	Cord blood	NK	$CD3^{-}CD14^{-}CD19^{-}$ $CD56^{bright}CD94^{+}$	Stage 4
cbNKst5_2	Cord blood	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{-}CD34^{-}$ $CD56^{dim}CD94^{+}$	Stage 5
toNKst3_1	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{+/-}CD94^{-}$	Stage 3
toNKst4_1	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{bright}CD94^{+}$	Stage 4
liNKCXCR6n_1	Liver	NK	$CD56^{dim}CXCR6^{-}$	CXCR-negative
liNKCXCR6n_2	Liver	NK	$CD56^{dim}CXCR6^{-}$	CXCR-negative
liNKCXCR6p	Liver	NK	$CD56^{dim}CXCR6^{+}$	CXCR-positive
bmNKst1_2	Bone marrow	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{-}CD34^{+}$	Stage 1
bmNKst2_3	Bone marrow	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{+}$	Stage 2
bmNKst3_2	Bone marrow	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{+}$	Stage 3
bmNKst4_2	Bone marrow	NK	$CD3^{-}CD14^{-}CD19^{-}$ $CD56^{bright}CD94^{+}$	Stage 4
bmNKst5_2	Bone marrow	NK	$CD3^{-}CD14^{-}CD19^{-}CD56^{dim}CD94^{+}$	Stage 5
toNKst2	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{+}$	Stage 2
toNKst3_2	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{+/-}CD94^{-}$	Stage 3
toNKst4_2	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{bright}CD94^{+}$	Stage 4
toNKst5	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{-}CD34^{-}$ $CD56^{dim}CD94^{+}$	Stage 5
toNKst3a	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{-}CD94^{-}$	Stage 3a
toNKst3b	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{+}CD94^{-}$	Stage 3b
toNKst3_4	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{-}CD94^{-}NKp44^{+}$	Stage 3 NKp44-positive
toNKst3_5	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{+}CD94^{-}NKp44^{+}$	Stage 3 NKp44-positive
toNKst3_6	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{+}CD94^{-}NKp44^{-}$	Stage 3
toNKst3_7	Tonsils	NK	$CD3^{-}CD14^{-}CD19^{-}CD117^{+}CD34^{-}$ $CD56^{-}CD94^{-}NKp44^{-}$	Stage 4
pbB	Peripheral blood	B	$CD19^{+}$	B lymphocytes
cbMono	Cord blood	MC	$CD14^{+}$	Monocytes
pbMono	Peripheral blood	MC	$CD14^{+}$	Monocytes
pbNKst5lic_11	Peripheral blood	NKs	$NKG2A^{+}CD56^{dim}$	Stage 5 licensed

2.5. Preparation of single cell samples

For NK cell transcriptome analysis at single cell level, mature NK cells from peripheral blood were selected as described below.

2.5.1. Isolation of peripheral blood mononuclear cells

PBMCs for single cell analysis were isolated by density gradient centrifugation using the Leucosep^{TM} system. First, 50 ml separation tubes containing 15 ml of Biocoll separation medium were centrifuged for 20 s at 1000 x g, so that the medium is localized below the porous barrier. After addition of 30 ml blood, diluted 1:1 with PBS, the tubes were centrifuged for 10 min at 1000 x g. After centrifugation, the PBMC-containing fraction was harvested using a pasteur pipet and washed three times with PBS for 10 min at 250 x g. After the second washing step all PBMCs were combined in one tube.

To remove residual erythrocytes, additional treatment with erythrocyte lysis buffer was performed. For this purpose, PBS was aspirated after the last washing step and PBMCs were resuspended in 30 ml Erythrocyte lysis buffer. After incubation for 5 min at room temperature, the cells were pelleted for 10 min at 250 x g and resuspended in MACS® buffer.

2.5.2. NK cell enrichment

Enrichment of NK cells from PBMCs was performed using the MACS® magnetic NK Cell Isolation Kit. Due to labeling of non-target cells (i.e. T cells, B cells, stem cells, dendritic cells, monocytes, granulocytes, and erythroid cells) with lineage-specific biotinylated antibodies, this kit allows enrichment of intact, unlabeled NK cells.

After treatment with erythrocyte lysis buffer, every 10^7 PBMCs were resuspended in 40µl MACS® buffer and incubated with 10 µl NK Cell Biotin-Antibody Cocktail for 5 min at 8°C. Then 30 µl MACS® buffer were added and cells were incubated with 20 µl NK Cell MicroBead Cocktail for 10 min at 8°C.

The mixture was adjusted to 500 µl with MACS® buffer and applied to a MS separation column, preliminary equilibrated with 500 µl MACS® buffer. The flow-through containing unlabeled NK cells was collected. Then the column was washed with 500 µl MACS® buffer and the flow-through was combined with the unlabeled NK cell obtained from the previous step.

The NK cell suspension was washed with 5 ml PBS for 10 min at 250 x g and resuspended in fresh PBS before being applied to the Integrated Fluidic Circuits (IFCs).

2.5.3. Generating of cDNA libraries from single cells

NK cells obtained from MACS® enrichment were loaded on the IFC for mRNA sequencing at a final concentration of 200 cells/µl.

Capturing of single cells, their lysis, first-strand reverse transcription on polyA$^+$ RNA templates, long-distance PCR, and addition of SMARTer® universal tag sequences on 3’ and 5’ cDNA ends was performed in IFC according to the manufacturer’s protocol

using the C1™ Single cell Auto Prep System to generate single cell tagged cDNA libraries for mRNA sequencing (Fluidigm Corporation ©, 2014a). Efficiency of cell capturing in the IFC was controlled in light microscope (Observer A1, Zeiss).

After harvesting from the IFC, cDNA libraries (total volume of 3 µl for each) were diluted with 10 µl C1 DNA Dilution Reagent. DNA concentrations in the libraries were measured using the BioAnalyzer and the High Sensitivity DNA Assay kit; 18 samples with total DNA concentrations of >0.2 ng/µl were selected for preparation of sequencing libraries.

2.6. Sequencing

The preparation of sequencing libraries from complete sample transcriptomes of single cell samples and lymphocyte populations using Nextera DNA Library Preparation Kit (Illumina) as well as deep sequencing using IlluminaHighSeq™ technology was performed in the Transcriptome Analysis Laboratory (University Medical Center Göttingen).

2.7. Statistical analysis

Normalized expression levels used for all analysis were calculated as fragments per kilobase of exon per million fragments mapped (fpkm).

Analysis of normalized data was performed using R software. Most of the analyses, including principal component analysis (PCA), ANOVA, unsupervised hierarchical clustering (HC) on gene expression, and visualizing results by violin plots was done using SINGuLAR Analysis Toolset, a shared-source, proprietary data analysis resource for gene expression data (Fluidigm Corporation ©, 2014b).

2.7.1. Log_2EX and Limit of Detection (LoD)

The SINGuLAR Analysis Toolset performs all statistical analyses based on the expression values in the Log_2 domain.

Normalized levels of gene expression levels are converted to Log_2 values (Log_2Ex), as the primary interest in expression analysis is the fold change. Because small numbers (less than 1) can become very large negative values in the Log_2 domain, a LoD is required to eliminate background noise. Normally in RNA-seq experiments LoD is set to 1 so that background expression values are zeros in the Log_2 domain.

Linear expression values are converted to Log_2Ex by the following formula:

$Log_2Ex = Log_2$(expression level) if LoD ≤ expression level, and Log_2(LoD) = 0 if LoD > expression level.

2.7.2. Principal component analysis

The PCA is a statistical data reduction technique that uses successive orthogonal transformations to narrow down variation within a data set in as few variables as possible, while retaining most of the original variation. PCA reduces the dimensionality of a data set by transforming it into a new set of uncorrelated variables, called principal components, with decreasing degrees of variability. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn explains the next highest variance in the data set under the constraint that its relationship with the previous PC is zero (Jolliffe, 1986).

2.7.3. The Outlier Identification Method

The outlier analysis is based on the assumption that samples of the same type also have a set of commonly-expressed genes. The algorithm iteratively trims the low-expressing genes in an expression file, until 95% of the remaining genes are expressed above the defined LoD value in half of the samples included in the analysis, according to the assumption that the sample set contains less than 50% outliers. This means that subsequent calculations will only include the half of the samples that have the highest expression for the trimmed gene list. The trimmed gene list represents the most evenly expressed genes.

For the 50% of the samples that remain, a distribution is calculated that represents their combined expression values for the gene list defined above. For this distribution, the median represents the 50th percentile expression value for the set of data.

The Outlier Threshold is defined as the expression value at which 85% of the gene expression values are above that line. Outliers are then identified as samples whose median expression values are less than the Outlier Threshold. Then, the outlier identification function automatically identifies outliers and displays a box plot of expression values for each sample, with the outlier candidates labeled. At this point it is also possible to display a PCA score plot of normal and outlier data points with sample names labeled to visualize each sample's location in relation to each other. One can either accept an automatically identified list of outliers or, based on this data, indicate the list of outliers manually, if needed.

In further analyses only samples identified as normal are used and only genes that are expressed over the LoD in at least 50% of the analyzed samples.

When samples are divided into groups, an outlier analysis is performed for each group separately.

2.7.4. Identification of top genes

When an analysis is performed with SINGuLAR Analysis Toolset, the algorithm automatically identifies the top N PCA genes, thus, those genes that are the mostly differentially expressed ones among analyzed samples. Then the algorithm trims the expression data with those top genes. All subsequent analyses, including ANOVA, PCA, and HC, are performed using the edited gene list. The number of genes used for the analysis in a standard case was 100, however it could be raised up to 400.

2.7.5. ANOVA

Analysis of Variance (ANOVA) is a statistical method used to test for differences between two or more means. It is based on the idea that variability in the quantity being measured (gene expression, for example) can be partitioned into a number of identifiable sources. The one-way ANOVA, which is used in the SINGuLAR Analysis Toolset, compares the means between the groups of interest and determines whether any of those means are significantly different from each other.

When ANOVA is used for gene expression data analysis, it allows identification of genes that are differentially expressed among analyzed groups. In this study ANOVA was normally used to test the set of the top 100 genes upon LoD=1.

2.7.6. Hierarchical Clustering Analysis

A cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). The SINGuLAR Analysis Toolset uses an algorithm of unsupervised HC. The goal of HC is to build a binary tree or a dendrogram of data (in this case – of gene expression levels) in the form of a heatmap that successively merges similar groups of data points. HC algorithms connect data points to form clusters based on distance, with a cluster being described largely by the maximum distance needed to connect parts of the cluster.

Co-profiled genes are clustered together using the Pearson method and samples are clustered together using the Euclidean method. In each case, the complete linkage method is then used to find similar clusters. The data is then visualized as a heatmap with a dendrogram that shows both clusters of genes and samples, based on the expression levels of these genes. In this study two visualizing strategies were used depending on the purpose: “gene_z_score” and “global_z_score”.

The so-called "gene_z_score" display method was used in most cases to visualize the gene expression data. In this case expression values per gene were normalized with the mean and standard deviation for each gene, which better graphical demonstration of similarities in expression patterns between genes. The “global_z_score” display method was used to visualize single cell gene expression data. In this case expression values per gene were normalized with the global mean and global standard deviation, which allows better graphical demonstration of similarities between samples.

2.7.7. Violin plots

Violin plots depict the probability density of the data at different values and are used to visualize gene expression levels within the analyzed sample group. The algorithm included in the SINGuLAR Analysis Toolset builds histograms for multiple genes; each of them visualizes which portion of samples from the analyzed group expresses the gene at a given level.

2.7.8. Pairwise comparison between samples and sample groups

The algorithm allowing for pairwise comparison between sample groups was also included in the SINGuLAR Analysis Toolset. Pairwise scattered plots between sample groups allow to visualize the average gene expression values for each sample group and calculate the correlation coefficient between two sample groups.

Correlations between expression levels of single samples were calculated separately in R using Spearman’s correlation method, which was part of the Hmisc R package. Scattered plots of gene expression levels used for visualization of correlation between individual samples were obtained from Dr. Angela Noll.

2.7.9. Spearman’s correlation analysis

Spearman’s rank correlation is a non-parametric test, which is used to identify whether there is an association between two ranked variables (in this case between expression levels of two genes). Spearman's coefficient ρ (rho), or r shows the strength of association between two variables, where the value r = 1 means a perfect positive correlation and the value r = -1 means a perfect negative correlation.

The correlation analysis was performed in R. The algorithm is a part of the Hmisc R package, which allows calculating matrixes of Spearman's rank correlation coefficients r and corresponding significance tests (p-values). Only coefficients with $p<0.05$ were accepted as significant. Spearman’s correlation analysis requires at least five values to calculate the coefficient, meaning that groups of at least five samples have to be used to analyze correlation of gene expression within the samples.

2.7.10. Gene ontology analysis

Gene ontology (GO) analysis, specially a statistical the overrepresentation test, was performed using the PANTHER (Protein ANalysisTHrough Evolutionary Relationships) classification system (Mi et al., 2016). This online analysis tool was designed to classify proteins and genes encoding them, and is a part of the Gene Ontology Reference Genome Project. It allows classifying proteins and corresponding genes according to different characteristics, in particular according to the biological process in which they are involved.

In this study the statistical overrepresentation test was used to determine representatives of which biological processes are significantly enriched in a given list in comparison to the reference list (in particular case - in complete human genome).

When a list with genes of interest is given to the tool, genes are mapped to particular GO-slim Biological Process. Then, an expected value is calculated, which is the number of genes of this category expected to be present in the list, based on their presence in a reference list. Then, using binominal statistics, an expected value is compared to a real number of genes from a given category present in the list, resulting in the fold enrichment. If the fold enrichment is greater than 1, it indicates that the category is overrepresented in the analyzed list. Conversely, the category is underrepresented if it is less than 1. Fold enrichment values with $p<0.05$ were considered as significant.

3. RESULTS

3.1 Clustering of different lymphocyte group: proof of data quality and the method

To prove the data quality, PCA and HC was first performed using data from all samples available (n=91), including NK, T, and B lymphocytes, and monocytes. When the limit of detection (LoD) 1 was used, 20858 genes among the complete transcriptome were included in the analysis. PCA and hierarchical cluster analysis were performed using top 100 of the mostly differentially expressed genes (Appendix 1). Based on this expression data, both on PCA score plot and clustered heatmap (Figure 1, 2) the following sample groups were observed: lymphocyte precursors (stage 1 and 2), NK cell precursors (stage 3), NK cells stage 4 and 5, T lymphocytes; B lymphocytes, and monocytes. Thus, the sequencing data is complete enough and its quality is sufficient for the analysis.

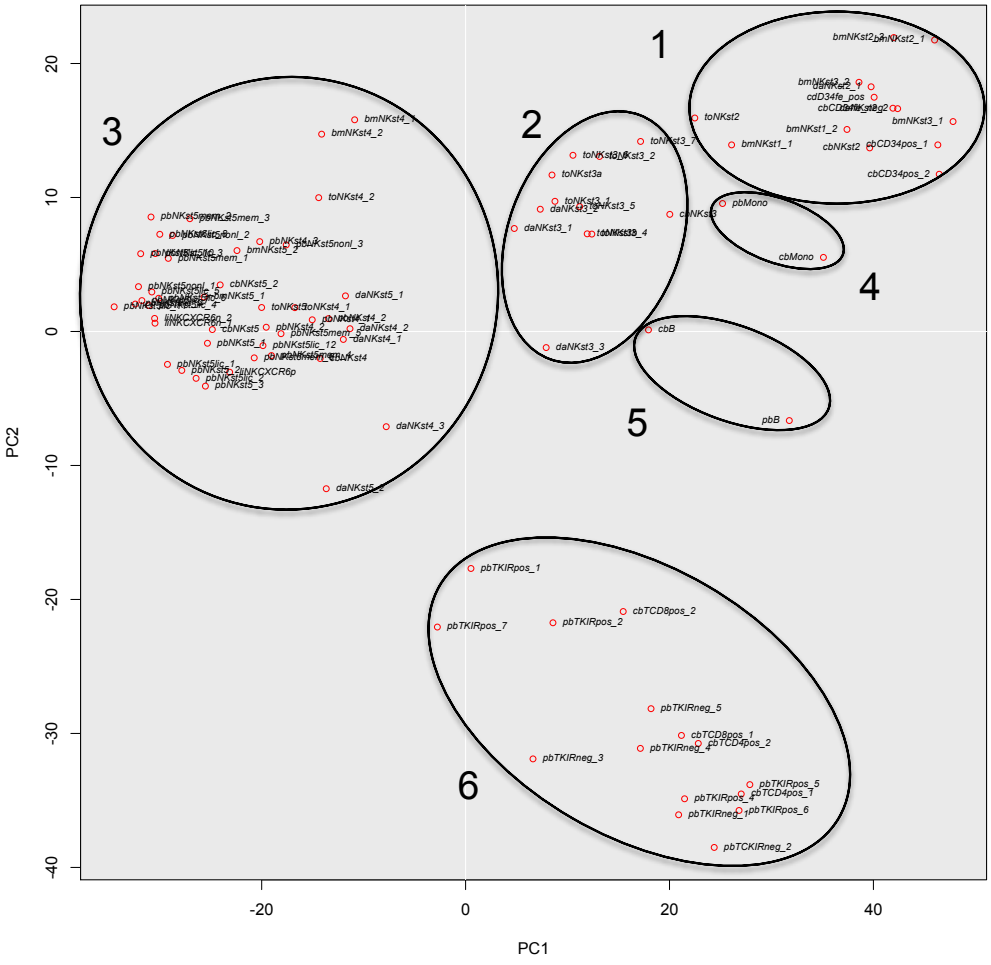

Figure 1 PCA score plot of analyzed samples based on expression levels of the top 100 differentially expressed genes.
Genes were selected from the total number of 20858, LoD=1. 1 - lymphocyte precursors; 2 – immature NK cells (stage 3); 3 – NK stage 4 and 5; 4 – monocytes; 5 - B lymphocytes; 6 – T lymphocytes.

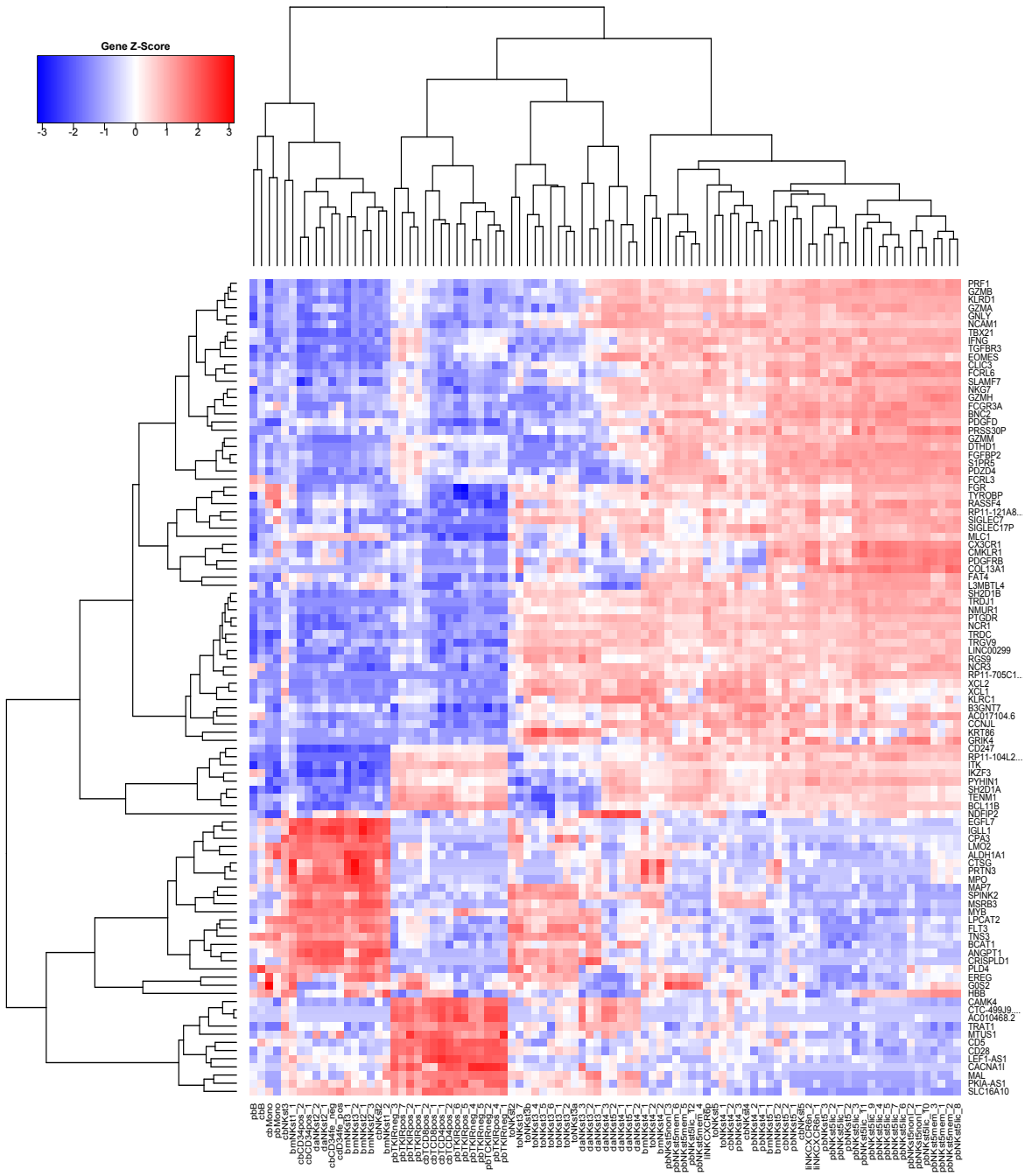

Figure 2 HC heatmap of analyzed samples based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 20858, LoD=1.

As can be seen on the clustered heatmap (Figure 2), for most of these groups specific expression patterns were observed (Table 2). Furthermore, known functions of most of these specifically expressed genes were consistent to observed expression patterns. In particular, among genes expressed in the group of all committed NK cells (stage 3 to stage 5) were classical NK cell markers such as *NCR1, NCR3, NCAM1*, and *KLRC1*. Furthermore, genes encoding NK effector molecules, such as PRF1, IFNG, CST7, GNLY and granzymes (*GZMA, GZMB, GZMH, GZMM*), and other molecules typical for mature

NK cells (*FCRL3, FCRL6, CX3CR1, FCGR3A, FGFBP2, S1PR5*) were expressed by NK cells stage 4 and stage 5, but not by stage 3 NK cells.

Table 2 Expression patterns of 100 top differentially expressed genes among all samples, LoD1

Lymphocyte precursors	T lymphocytes	T lymphocytes and NK stage 5	NK stages 3-5	NK stages 4 and 5
ALDH1A1	AC010468.2	BCL11B	AC017104.6	ATP8B4
ANGPT1	CACNA1I	CD247	ATP8B4	BNC2
BCAT1	CAMK4	KZF3	B3GNT7	CLIC3
CPA3	CCR4	PYHIN1	CCNJL	CMKLR1
CTSG	CD28	SH2D1A	CD300A	COL13A1
EGFL7	CD5	TENM1	FAT4	CX3CR1
FLT3	CTC-499J9.1		FGR	DTHD1
IGLL1	LEF1-AS1		GNLY	FCGR3A
LMO2	MAL		GRIK4	FCRL3
LPCAT2	NELL2		IL18RAP	FCRL6
MAP7	PKIA-AS1		ITGAM	FGFBP2
MPO	SLC16A10		KLRC1	GZMA
MSRB3	TRAT1		KRT86	GZMB
MYB			LINC00299	GZMH
PLD4			MLC1	GZMM
PRTN3			NCAM1	IFNG
SPINK2			NCR1	ITK
TNS3			NCR3	KLRD1
			NMUR1	L3MBTL4
			PDGFD	LINGO2
			PTGDR	NKG7
			RASSF4	PDGFRB
			RGS9	PDZD4
			RNF165	PRF1
			RP11-104L21.3	PRSS30P
			RP11-121A8.1	S1PR5
			RP11-705C15.5	SLAMF7
			SH2D1B	TBX21
			SIGLEC17P	
			SIGLEC7	
			TRDC	
			TRDJ1	
			TRGV9	

			TYROBP	
			XCL1	
			XCL2	

Notably, $CD56^{dim}CXCR6^{-}$ hepatic NK cells were clustered together with stage 5 NKs from other tissues and shared a mature expression profile, while the $CD56^{dim}CXCR6^{+}$ sample clustered together with stage 4 NK cells. In comparison to stage 5 NK cells, $CXCR6^{-}$ samples had lower expression levels of markers of mature NK cells such as *GNLY, GZMB, GZMH, TBX21, CX3CR1, CMKLR1, COL13A1 and FGFBP2*, as well as of *KRT86, LINGO2, PDGFRB PRSS30P*. Furthermore, the expression of some genes (e.g. *RGS9, FGFBP2*, *GZMB* and *GZMH*) was lower than in both stage 4 and stage 5 NK cells.

In the group of lymphocyte precursors hematopoiesis-related genes were expressed, such as *MYB, CPA3, FLT3, LMO2, MPO, IGLL1* and genes involved in angiogenesis (*EGFL7* and *ANGPT1*) (Su et al., 2004; Surmiak et al., 2012; Yang et al., 2002). Notably, stage 3 samples from bone marrow and cord blood clustered together with stages 1 and 2 and shared with them a specific cluster of 18 highly expressed genes (Figure 2).

Among the genes that are expressed mainly in mature T lymphocytes were classical T-cell markers like *CD28, CD5, TRAT1* and other gens known being expressed in T lymphocytes such as *CAMK4, CCR4*, and *MAL* (Illario et al., 2008).

To prove whether the obtained data are reproducible, a correlation analysis of the complete transcriptomic data was performed on sample pairs that were most similar considering the sorting strategy and the tissue origin, and were derived from the same donor. Two such pairs were present in the sample set: pbTKIRpos_1 and pbTKIRpos_2, pbNKst5lic_6 and pbNKst5lic_7. In both cases a high similarity of gene expression data was observed on linear regression plots (Figure 3) and corresponding correlation coefficients were 0.91 and 0.92 respectively, proving that the obtained data is reproducible.

To prove that functional enrichment of immune system-related genes was significant, Panther GO overrepresentation test was performed. Out of the top 100 most differentially expressed genes, 88 were mapped to particular biological processes (Table 3); among the significantly overrepresented ones were genes involved into immune system processes (GO:0002376), in particular NK activation (GO:0030101) and B cell mediated immunity (GO:0019724).

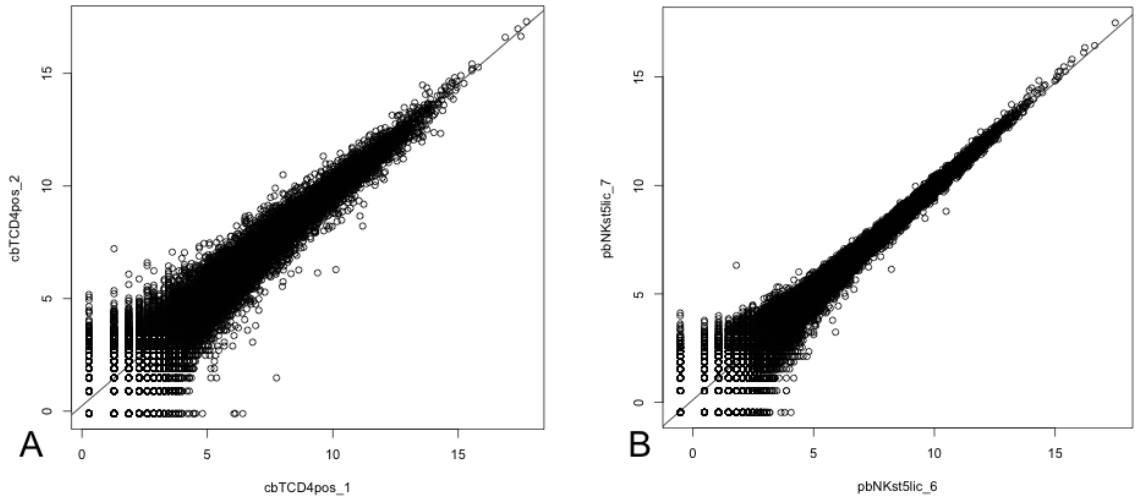

Figure 3 Scattered plots of gene expression CD4⁺ T lymphocytes (A) and licensed NK cells (B) Correlation analysis was based on complete transcriptome, compared samples represented the same tissues of origin and were derived from the same donors.

The analogical analysis, including PCA, HC and Panther GO overrepresentation was also performed using top 400 differentially expressed genes (data not shown). As no differences in distinguishing between sample groups and clustering was observed in comparison to analysis based on top 100 genes, it was concluded, that 100 of genes are sufficient for analysis.

Table 3 Gene ontology overrepresentation among top 100 differentially expressed genes

PANTHER GO-Slim Biological Process	Gene number	Fold enrichment	P-value
Cellular process (GO:0009987)	45	1.59	4.02E-02
Response to stimulus (GO:0050896)	33	3.6	4.24E-09
Immune system process (GO:0002376)	27	4.59	2.43E-09
Immune response (GO:0006955)	18	8.2	1.35E-09
B cell mediated immunity (GO:0019724)	6	10	7.20E-03
Natural killer cell activation (GO:0030101)	6	14.28	9.93E-04

In summary, a successful differentiation between blood cell populations is possible based on the available transcriptomic data and top 100 genes are sufficient for this. The groups observed upon PCA and HC correspond to FACS sorting strategies and most of the

genes that were differentially expressed between the groups and defined the specific clustering are well-known and typical for the corresponding cell populations.

3.2 Identification of specific expression patterns between mature PBMC populations

To further investigate cell type specific gene expression patterns, expression profiles of the following mature PBMC sample groups were compared: stage 5 NK lymphocytes (n=27), T lymphocytes (n=10), B lymphocytes (n=2), and monocytes (n=2). As it is was suggested, that KIR^+ T lymphocytes have specific features in comparison typical T lymphocyte phenotype (Uhrberg et al., 2001; Young et al., 2001), and also our data showed that they differ in their expression profiles (not shown), six KIR^+ T cell samples were not included in this analysis. Based on LoD=1, 20180 genes were included into the analysis. ANOVA analysis and PCA were performed to identify top 100 genes that are significantly ($p<0.05$) differentially expressed between abovementioned groups.

On the PCA score plot (Fig 4A) specific clusters of NK, B and T cells as well as monocyte samples was observed.

Notably, among genes specifically expressed in NK samples (Figure 4B) were known NK cell markers (e.g. *FCGR3A, FCRL6, FGR, KLRD1, KLRF1, NCAM1, NCR1, NKG7, PRF1, GZMA, GZMB, GZMH, TYROBP*), but also genes without well-described function, in particular non-protein coding genes *LINC00299, AC017104.6, CTC-499J9.1*, and *AC010468.2*. Analogically, the group of genes specifically expressed among T cell samples included known T lymphocyte markers, such as *CD5, CD28, CAMK4, NELL2*, and *TRAT1*, but also non-protein coding genes *CTC-499J9.1, AC010468.2, RP11-203J28.8, LEF1-AS1, PKIA-AS1*.

Eleven genes were highly expressed in B lymphocytes: six of them were also expressed in NK cells (*ITGAM, FGR, SYK, B3GNT7, AC017104.6, HBB*), and the other five (*SPI1, TNS3, JUP, ST14, CFD*) were also expressed in monocytes, some NK or T cells.

Due to the analysis algorithm, the relative size of a given sample group in comparison to other groups influences the ability to define group-specific genes. For this reason most of the top 100 differentially expressed genes identified for this sample set represented NK- and T-specific genes. Even though B lymphocyte and monocyte samples could be successfully differentiated from other samples, no genes expressed exclusively in these cell types could be identified within the data set of 100 differentially expressed genes.

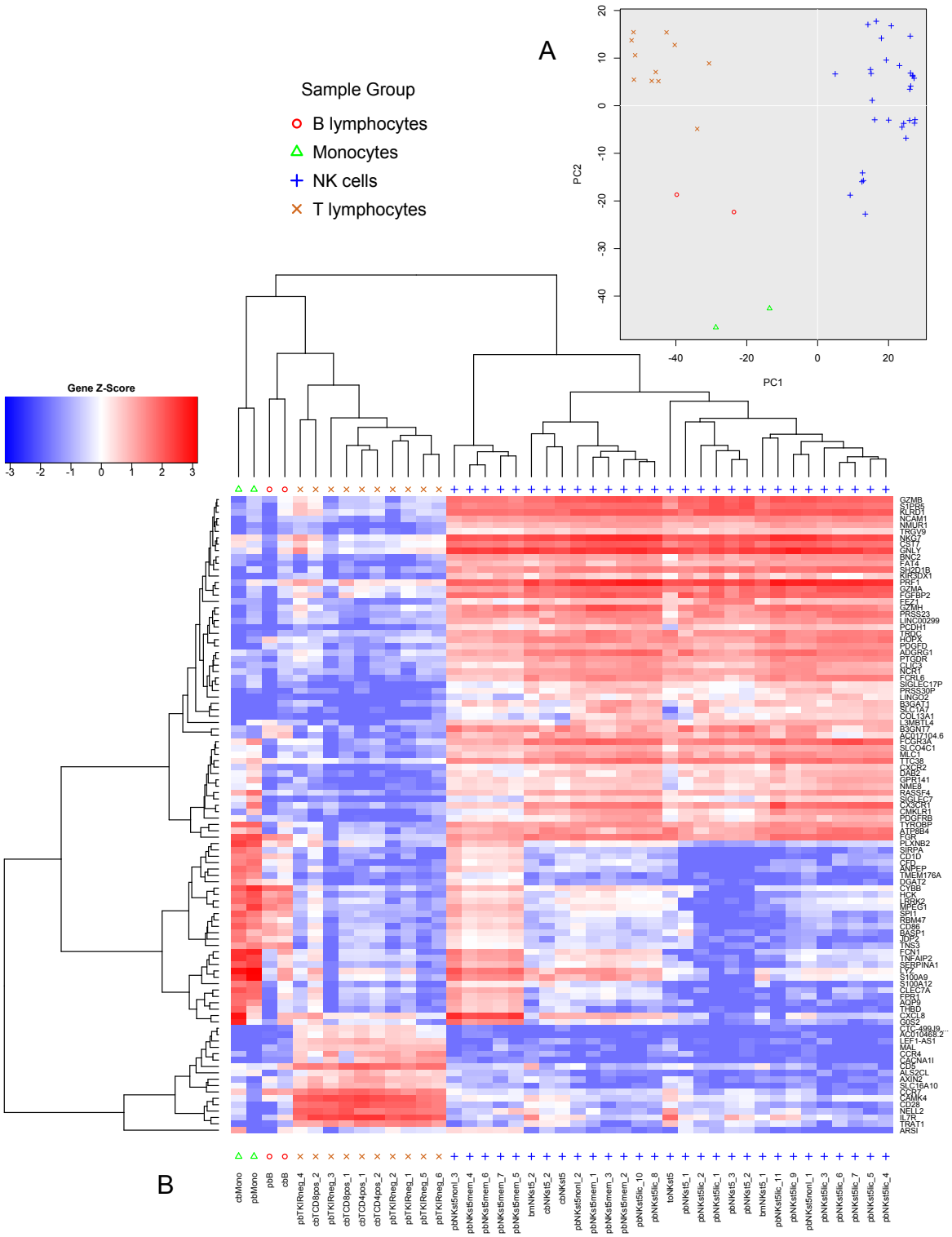

Figure 4 PCA score plot (A) and HC heatmap (B) of mature PBMC samples based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 20180, LoD=1.

Despite this, when ANOVA analysis was performed, clear significant difference between all sample groups was observed. Among the top 100 genes, 80 genes were differentially expressed between NK and T lymphocytes, 69 between NK and B

lymphocytes, 66 between NK lymphocytes and monocytes, 56 between T lymphocytes and monocytes, 32 between B and T lymphocytes, and 23 between monocytes and B lymphocytes (Appendix 2).

So, despite the fact that genes specifically expressed in small sample groups are underrepresented, significant differences between gene expression levels within populations of mature PBMCs could be identified. These populations possess specific gene expression patterns, which include both genes with known functional relevance, as well as not well-described genes.

3.3 Expression profiling of *ex vivo* NK populations

The specific clustering using T and B lymphocytes as outgroups has confirmed the high RNA-seq data quality, as we could successfully distinguish between the different lymphocyte groups. However, as the main focus of my project was on NK lymphocyte expression profiling, in further studies I have concentrated on analysis of *ex vivo* NK subpopulations.

3.3.1 Changes in expression landscape during NK cell development

To check how expression profiles change during NK cell development, the analysis was done on *ex vivo* NK cell samples (n=53) including two samples from NK stage 1, four samples from stage 2, eleven samples from stage 3, nine samples from stage 4, and 29 samples from stage 5. Among all analyzed genes, 20459 exceeded the threshold of LoD=1; PCA and HC were performed based on the top 100 differentially expressed genes.

On PCA score plot (Figure 5) samples were grouped according to developmental stages and partly according to the origin of the tissue. Groups including stage 4 and stage 5 samples were clearly separated; only one out of 30 stage 5 samples (toNKst5) was localized together with samples from stage 4. Also eight samples from tonsil-originated stage 3 were grouped together, while stage 3 samples from bone marrow were localized closer to earlier developmental stages. The stage 3 sample from cord blood and stage 2 sample from tonsils had an intermediate position between the group consisting of other stage 3 samples and the group including stages 1 and 2 samples. For such an outlying position of these two samples two explanations are possible: either their expression phenotypes are intermediate between early progenitors (stages 1 and 2) and stage 3 NK precursor, or cells included in them represent a mixture of several cell types.

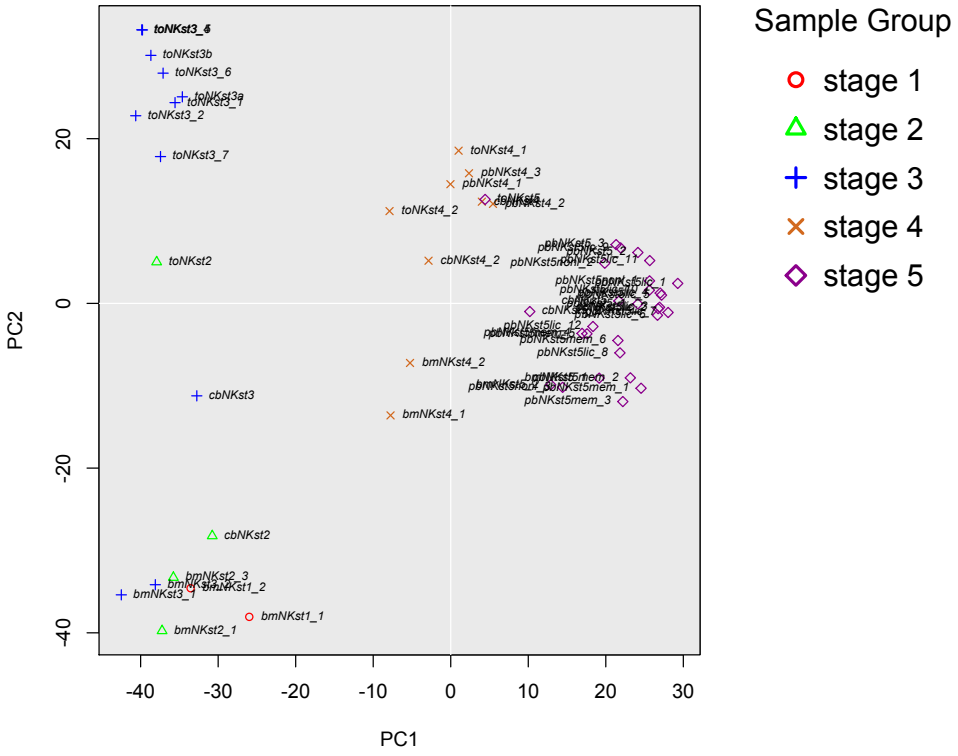

Figure 5 PCA score plot of *ex vivo* NK samples based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 20459, LoD=1.

According to the HC heatmap (Figure 6), a gradual change in the gene expression profiles was observed from stage 1 to stage 5. The following gene expression patterns were observed:

• Genes with the highest expression among stage 1 and 2 samples included myeloid-lineage related genes (*MMP8, LTF, DEFA3, MPO, AZU1, CTSG, CLC, PRTN3, TAL1, AHSP, RHAG, SPTA1, HBD*), as well as other genes known being expressed by $CD34^{+}$ NK progenitor cells (*MS4A3, CPA3, ELANE*) (Lakschevitz et al., 2015; Sánches et al., 1992; Thomas et al., 2014).

• Among the genes, expression of which continues till stage 3, were typical markers of early NK cell development (*MAP7, KIT, MYB, SPINK2, GNAI1, GRB10, CDH1*), other NK-lineage specific genes (e.g. *ADGRG6, CXCL8*), T-lineage associated gene *TGM2*, and myeloid-specific genes *CA2, CDC42BPA, PRSS57* (Eissens et al., 2012; Surmiak et al., 2012). Besides this, *VWDE, CASC15, MRC2, STAC, EREG* were also expressed at the highest levels within stages 1 to 3.

• Genes with the highest expression in stage 3 and lower expression at stage 4 samples included ILC3-related genes IL23R, IL1R1 and RORC, genes specific for the

myeloid lineage (*LIF* and *SLC4A10*) and a set of genes known to be expressed in lymphocytes, for which, however, no specific function for NK stage 3 is yet known (*IL4I1, B3GALT5, CSPG4, ENPP1, GPR55, KRT81)* (Gurnett et al., 2008; Eissens et al., 2012; Takatori et al., 2009). In addition, a gene group was identified for which no specific function in lymphocytes was yet described (*KIAA1211L, KIAA1324, LINGO4, LPAR1, PKDCC, RRAD, THEM5, SMIM10L2A*), as well as non-protein-coding genes *LINC00892, RP11-330A16.1, RP5-1028K7.2, and AL450992.2.*

• A group of genes with equally high expression among stage 3 and 4 samples included mainly NK-lineage specific ones (*DLL1, XCL1, TNFSF11, TOX2*), but also IL7R (ILC3-specific), *IGFBP4* and *KRT86* (Cupedo et al., 2009; Eissens et al., 2012; Spits et al., 2013).

• Genes expressed among stages 3 to 5 were all typical for NK cells: *CD2, GNLY, GPR68, TNFRSF18, XCL2* (Seillet et al., 2014).

• Genes with the highest expression among stage 4 and 5 samples included mainly markers of mature NK cells: *TBX21, GZMB, PDZD4, IFNG, GZMM, GZMH, GZMA, EOMES, FCRL6, KLRD1, TGFBR3, KLRF1, KIR3DX1, CD8A, FGFBP2, S1PR5, FCGR3A, FCRL3, PDZD4*, but also several genes for which expression in NK lymphocytes was not yet described (*DTHD1, PYHIN1, RP11-222K16.2, TENM1*) (Eissens et al., 2012; Seillet et al., 2014).

• Finally, genes expressed only among stage 5 samples included *FEZ1, B3GAT1, AKR1C3* known to be expressed in $CD56^{+}$ NK cells, as well as *SLC1A7, BNC2, LGR6* (Jakobs et al., 2001).

Consistent to the picture observed on the PCA score plot (Figure 5), on the HC heatmap, the stage 3 samples of bone marrow origin were clustered together with developmental stages 1 and 2, and shared gene expression pattern with them, suggesting that they might have a more immature phenotype in comparison to stage 3 samples of tonsil origin.

To check, how expression profiles change during development, pairwise comparisons between developmental stages were performed (Figure 7A). If gene expression changes sequentially from one stage to the next one, total expression profiles should correlate between neighboring stages, but not with those that are far from each other along the NK cell developmental flow. The correlation was calculated using the top 100 differentially expressed genes selected upon LoD=1; correlation coefficients of >0.5 were accepted as a true positive correlation. Indeed, the correlation coefficient between

stages 1 and 2, stages 2 and 3, as well as between stage 4 and 5 were highly positive, namely 0.74, 0.679, and 0.664 respectively, while there was no significant correlation between stages 3 and 4. Correlation coefficients between non-neighboring stages were in the range from -0.507 to 0.331.

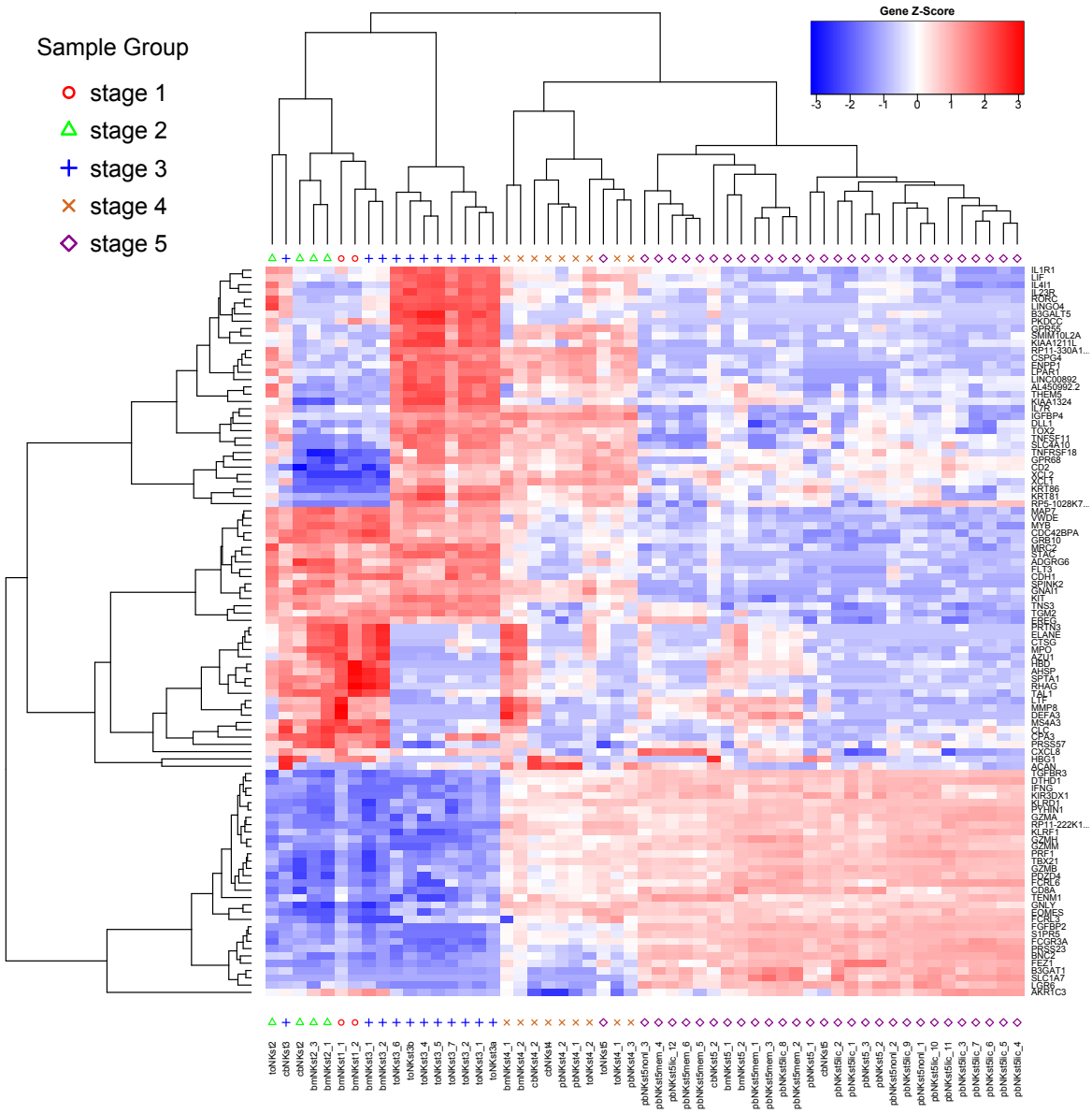

Figure 6 HC heatmap of *ex vivo* NK samples based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 20459, LoD=1.

To prove this tendency, an analogical analysis was performed based on the top 400 mostly differentially expressed genes (Figure 7B). In that case the highest correlations were observed between stages 1 and 2 (correlation coefficient 0.787) and between stages 4 and 5 (0.735). The correlation coefficient between stages 2 and 3 was lower (0.568), but

still higher than between any of non-neighboring stages, while correlation between stages 3 and 4 remained under the threshold (coefficient 0.265).

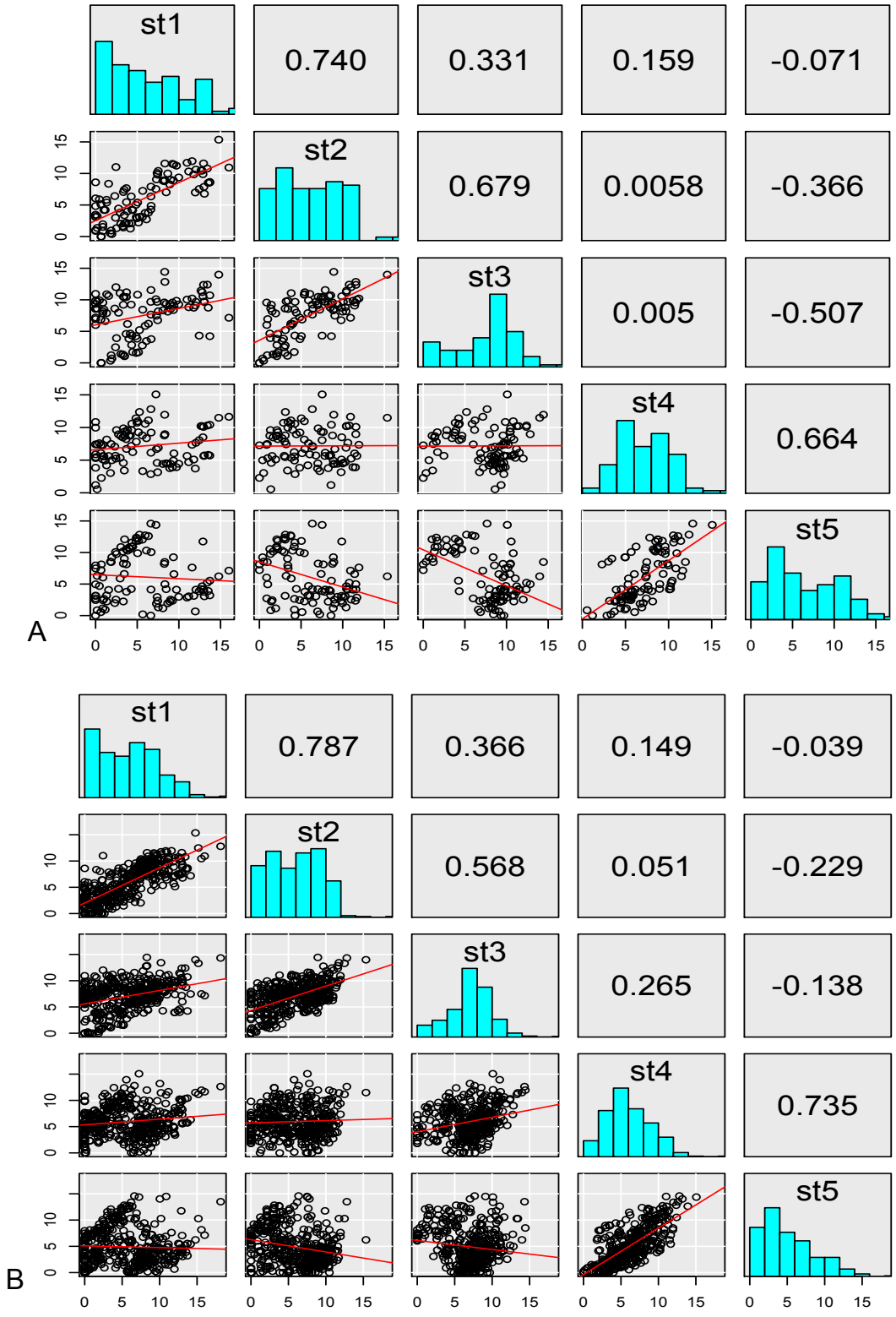

Figure 7 Pairwise comparison of gene expression levels within *ex vivo* NK developmental stages
A – based on top 100 differentially expressed genes; B – based on top 400 genes.

A plausible reason for the exceptional position of stage 3 samples is the known heterogeneity of stage 3. First, it is known that stage 3 cells sorted in FACS according to their $CD34^{-}CD117^{+}CD94^{-}$ surface phenotype are not a homogeneous population of cells, but also can include precursors of T lymphocytes, DCs and ILC3 cells. Another theoretical possibility would be heterogeneity between the analyzed stage 3 samples. To test this hypothesis, a correlation analysis between gene expression levels of individual stage 3 samples was performed (Table 4). According to Spearman's correlation analysis, correlation between all samples was between 0.84 and 0.93. Considering that this ratio was not lower than correlation rates between individual samples within other stages (not shown), it is unlikely that heterogeneity between stage 3 samples is the cause of the unique position of stage 3.

Table 4 Correlation between gene expression profiles of individual stage 3 *ex vivo* NK samples

	bmNKst3_1	cbNKst3	bmNKst3_2	toNKst3_1	toNKst3_2	toNKst3a	toNKst3b	toNKst3_4	toNKst3_5	toNKst3_6	toNKst3_7
bmNKst3_1	1										
cbNKst3	0.85	1									
bmNKst3_2	0.92	0.85	1								
toNKst3_1	0.86	0.86	0.88	1							
toNKst3_2	0.86	0.86	0.88	0.92	1						
toNKst3a	0.86	0.86	0.88	0.91	0.93	1					
toNKst3b	0.82	0.83	0.83	0.87	0.89	0.88	1				
toNKst3_4	0.83	0.83	0.85	0.88	0.90	0.89	0.87	1			
toNKst3_5	0.83	0.83	0.84	0.87	0.90	0.89	0.87	0.91	1		
toNKst3_6	0.78	0.79	0.80	0.82	0.84	0.83	0.82	0.85	0.86	1	
toNKst3_7	0.81	0.82	0.83	0.85	0.87	0.87	0.83	0.87	0.87	0.87	1

ANOVA analysis was then performed to specify the change in gene expression during transition from one developmental stage to the next one. Among a total of 100 of the most differentially expressed genes, 56 genes were differentially expressed between stages 3 and 4, 54 genes between stages 4 and 5, 45 genes between stages 2 and 3, and only 4 were differentially expressed between stages 1 and 2 (Appendix 3). In order to account for genes specifically expressed in small size groups, the whole analysis and ANOVA comparison was performed for only stage 1 and stage 2 groups separately. However, even in that case no additional genes were identified (data not shown).

3.3.2 Analysis of *ex vivo* NK stage 3 expression profiles

3.3.2.1 Tissue-specific differences in gene expression profiles of stage 3 *ex vivo* NK cells

To study tissue-specific differences in gene expression among stage 3 NK samples, 11 *ex vivo* NK samples were used: 8 of tonsil origin, 2 from bone marrow, and one from cord blood. Using LoD=1, 20160 genes from the complete transcriptome were included in the analysis; among them the top 100 of the most differentially expressed genes were selected to perform PCA and HC. As cord blood was represented by only a single sample, it was impossible to form sample groups and to compare gene expression between tissues with ANOVA analysis.

The stage 3 samples from tonsil origin are further distinguished based on the expression of the NKp44 receptor expression and were, therefore, sorted and analyzed separately. Both PCA plot and HC heatmap show this heterogeneity (Figure 8A,B). Samples toNKst3_4 and toNKst3_5 were sorted as NKp44+ while for sample toNKst3b this marker was not used, but it might contain many NKp44-positive cells. NKp44+ and NKp44- ILC3 cells have been described to differ in function (Crellin et al., 2010b; Hughes et al., 2009).

According to a clustered heatmap (Figure 8B), differences in sample expression profiles could be observed as well: bone marrow samples clustered closer to the sample derived from cord blood, while tonsil-derived samples clustered separately and could be also differentiated into two sub-clusters.

Among the top 100 genes, 9 were expressed in all tonsil-derived, but much lower in other samples; they express ILC3 markers (*KRT86, KRT81, SMIM10L2A, IL23R, GPR55*), NK-marker *XCL1*, as well as *GIMAP6, LPAR1,* and *KIAA1211L* (Gurnett et al., 2008; Eissens et al., 2012).

A bigger set of 31 genes was expressed in the cord blood stage 3 sample and in five out of eight tonsil stage 3 samples (on a lower level), but not in bone marrow and other tonsil samples (the same separate group as in PCA including samples toNKst3b, toNKst3_4 and toNKst3_5). This group included genes expressed in NK cells (*DHRS9, PTGDR2, FCGR2B*), other lymphoid lineage-related genes (*FAM49A, AKAP12, SLC7A11, IL1RL1, MS4A2, LILRB2, CD1C, HDC*), genes related to myeloid lineage (*THBD, PADI2, CLEC10A, IL1B, ADGRE3, TMEM176B, PTX3, APOBEC3A, AGPAT9, FCER1A*) as well as other genes (*HCAR3, C10orf105, HCAR2, CYP2S1, SLC45A3, CACNA2D3, HLA-*

DQB2, MS4A4E, RPL10P3, ANKRD55, ADGRE4P) (Chetyrkin et al., 2001; Sánches et al., 1992; Su et al., 2004; Thomas et al., 2014).

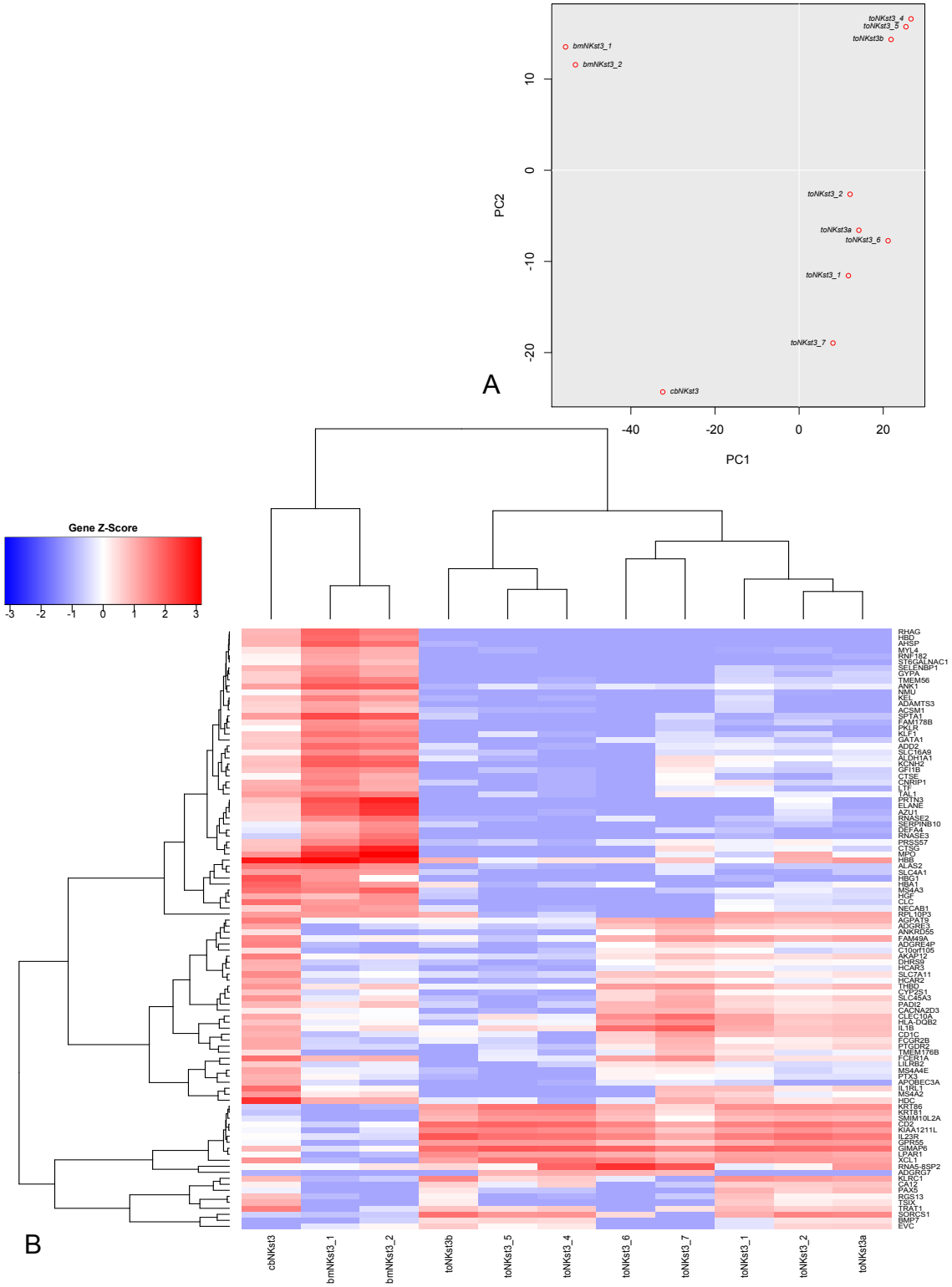

Figure 8 PCA score plot (A) and HC heatmap (B) of stage 3 *ex vivo* NK samples based on expression levels of the top100 differentially expressed genes
Genes were selected from the total number of 20160, LoD=1.

Only 3 out of the top 100 genes were expressed by bone marrow samples exclusively and they were characteristic for erythroid lineage (*SERPINB10, DEFA4* and *RNASE3*), while a cluster of 32 genes was expressed also by cord blood cell, but at lower level. This gene set mainly included genes that are characteristic for early myeloid and lymphoid lineages (*HBD, AHSP, MYL4, RNF182, ST6GALNAC1, SELENBP1, ANK1, NMU, KEL, ADAMTS3, ACSM1, SPTA1, FAM178B, PKLR, KLF1, GATA1, ADD2, ALDH1A1, KCNH2, GFI1B, CTSE, CNRIP1, LTF, TAL1, PRTN3, ELANE, AZU1, RNASE2*) and also *RHAG, GYPA, TMEM56, SLC16A9* (Lakschevitz et al., 2015; Surmiak et al., 2012; Yang et al., 2002;). In addition, 11 more genes, which also were mainly involved in hematopoiesis, (*PRSS57, CTSG, MPO, HBB, ALAS2, SLC4A1, HBG1, HBA1, MS4A3, HGF,* and *CLC*) were expressed on high levels by both cord blood and bone marrow-derived samples, but not by any tonsil-derived ones (Kinkade et al., 1983; Perera et al., 2013;).

3.3.2.2 Expression differences between stage 3 NK cells and ILC3

As it was mentioned before, samples with stage 3 phenotype sorted by FACS are a heterogeneous cell mixture containing ILC3 and NK precursors. It is also known from the literature that ILC3 cells are present among $CD34^{-}CD117^{+}CD94^{-}$ stage 3 phenotype in secondary lymphoid tissues and cord blood, but not in bone marrow (Crellin et al., 2010b; Cupedo et al., 2009; Geremia et al., 2011; Tang et al., 2011); and one of the aims of this project was to find new potential gene markers that would allow differentiating successfully real NK stage 3 cells from ILC3. For this purpose, the difference in gene expression profiles between bone marrow-derived and other *ex vivo* samples was examined.

Samples were divided into two groups: the first one contained two samples from bone marrow, the second one contained the other nine *ex vivo* stage 3 samples (cord blood and tonsil-derived). ANOVA analysis identified the top 100 differentially expressed genes and differences in gene expression between groups were visualized by violin plots.

Consistent with expression profiles observed before in a clustered heatmap, differences in gene expression levels were also observed in violin plots (Figure 9). The following genes were clearly higher expressed in bone marrow samples than in others: myeloid-related *RNASE3, DEF4A, SERPINB10, NMUH, AZU1, ELANE, RNASE2, GATA1, MPO, PRTN3, MYL4, AHSP, KLF1, CTSG, ANK1, CTSE, SPTA1, KEL, PRSS57, HBD*; genes known to be expressed by $CD34^{+}$ cells (e.g. *FAM178B, PKLR, KCNH2, ADAMTS3*); *SMIM10L2A* (ILC3-marker), *ST6GALNAC1* (ubiquitously expressed in

lymphocytes), *ADD2* (expressed in hematopoietic tissues), and other genes, whose expression patterns in blood cells is not yet described (*TMEM56, RNF182, RHAG, SLC16A9, COL6A5, ALDH1A1, MFSD2B, LIN7A*) (Su, 2004; Suhre et al., 2011; Zhang et al., 2000). Expression levels of NK-cell markers, such as *XCL1, GIMAP6, CD2, KLRC1*, as well as ILC3-cell markers *IL23R, KRT81, KRT86, GPR55*, and some other genes (*LPAR1, KIAA1211L, SORCS,* and *CA12*) were instead higher in non-bone marrow group. However, for most of the samples the expression of the respective genes was observed on different levels. In almost half of the cases, expression levels of both groups overlapped completely or mainly (e.g. for *TRAT1, HBA1, FCGR2B, AGPART9, RPL10P3, CLEC10A, IL1B, FAM49A, RNA5-8SP2, HLA-DQB2, HDC, DHRS9, SLC7A11, RP11-225N10.3, FCER1A, HCAR2, HCAR3, PADI2, CYPS1, AKAP12, CACNA2D3, LILRB2, SLC45A3, APOBEC3A, EVC, MSA4E, IL1RL1*).

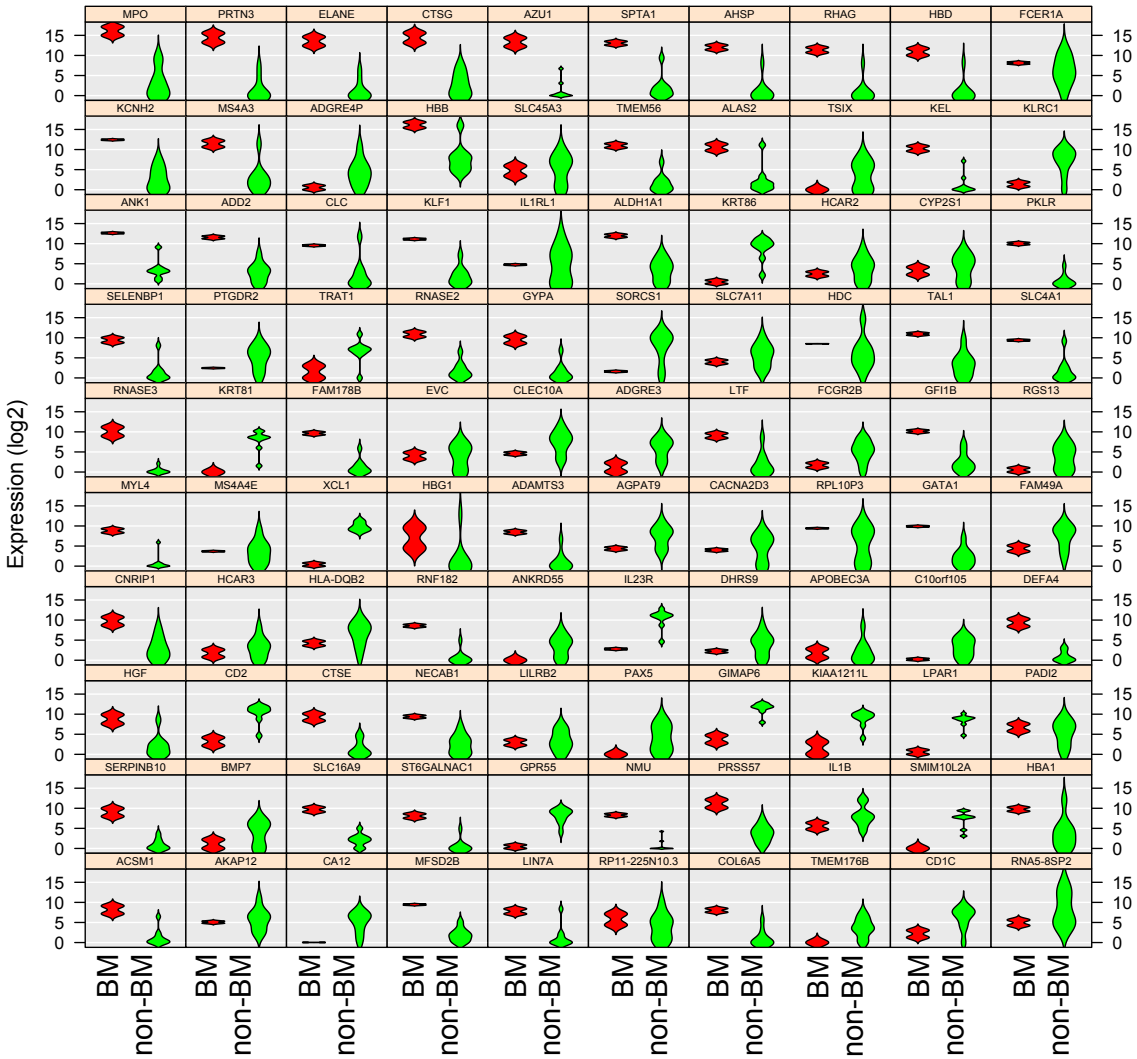

Figure 9 Violin plots of top 100 gene differentially expressed within bone marrow-derived (BM) and non-bone marrow-derived (non-BM) stage 3 *ex vivo* NK samples

Even though differences in gene expression could be observed both in heatmap and violin plots, when ANOVA analysis was performed, no statistically significant difference could be identified between two sample groups. A possible reason for this might be a high level of heterogeneity within the "non-bone marrow" group or a low number of samples derived from bone marrow. Even when the analysis was extended up to 400 top genes (data not shown), no significant difference was found.

However, already among the top 200 genes (Figure 10), some important differences between the groups were observed. In particular, among the genes with a higher expression in non-bone marrow group were *RORC* and *IL22*, which are typically expressed in ILC3, but not in NK cells. *IL23R* (present already among top 100 genes) and *IL7R*, which were also highly expressed by that sample group, are important for ILC3, as *IL23* and *IL7* are required for their development, but not for the development of NK cells (Hoorweg et al., 2012; Sanos et al., 2011).

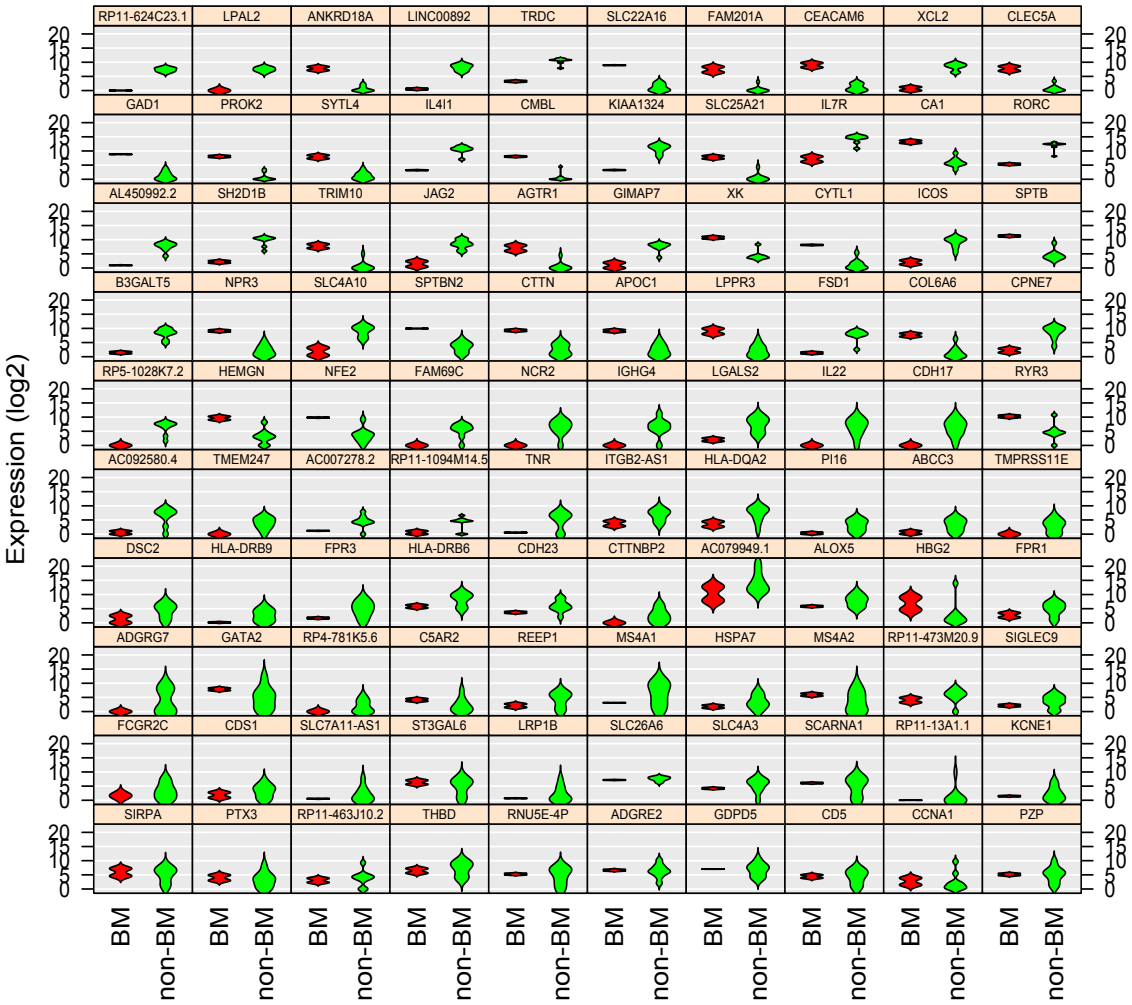

Figure 10 Violin plots of top 101-200 gene differentially expressed within bone marrow-derived (BM) and non-bone marrow-derived (non-BM) stage 3 *ex vivo* NK samples

Furthermore, *IL1R1*, which is critical for ILC3 development, but is not expressed by NK cells (Killig et al., 2014), was expressed much higher by non-bone marrow samples, than by bone marrow (data not shown). These observations suggested that differences in expression profiles of the bone marrow-derived stage 3 samples from the other stage 3 samples could indeed be caused by absence of ILC3 cells from bone marrow.

Besides genes already mentioned above, the non-bone marrow group had clearly a higher expression of NK-related genes *XCL2, GIMAP7, NCR2* and *IL4I1*, T- and DC-related genes *TRDC, LGALS2* and *ICOS,* as well as *KIAA1324, AL450992.2, SH2D1B, B3GALT5, FSD1, RP5-1028K7.2, FAM69CIGHG4, CDH17, AC092580.4, RP11-624C23.1*, and *TNR*. In the bone marrow sample group a higher expression of genes expressed on lymphocyte precursors (*CEACAM6, GAD1, CYTL1, HEMGN*), myeloid-related (*CLEC5A, CA1, TRIM10, XK, SPTBN2,* and *NFE2*) and other such as *ANKRD18A, FAM201A, PROK2, SYTL4, CMBL, SLC25A21, AGTR1, SPTB, NPR3, CTNN, LPPR3, COL6A6* was observed (Ai et al., 2016; Surmiak et al., 2012; Li et al., 2004).

In summary, these data suggest, that the difference between bone marrow and non-bone marrow-derived stage 3 samples is more due to a less mature phenotype of the stage 3 cells from bone marrow samples. However, the presence of ILC3 cells could make an influence.

3.3.2.3 Heterogeneity of non-bone marrow stage 3 samples

The heterogeneity of non-bone marrow stage 3 samples was examined additionally. Using LoD=1, 20201 genes were included in the analysis. PCA and HC were performed using the top 100 differentially expressed genes. Indeed, both in PCA plot and clustered heatmap a high level of heterogeneity between the samples could be observed (Figure 11).

In particular, a set of 14 genes among the top 100 was highly expressed in a cord blood sample in contrast to tonsil samples and consisted mainly of non-NK specific genes e.g. genes expressed in B cells, erythrocytes, eosinophils and basophils: *BCL11B, HDC, GATA2, CR1, CLC, TRPM6, MS4A3, GCSAML, HBG2, HBG1, RGS13, RPL10P3, RP11-225N10.3, IGF2BP3, HBA1*.

Also the tonsil stage 3 samples were quite heterogeneous. A cluster of 16 genes with higher expression in 6 tonsil samples, but much lower in two NKp44$^-$ samples, as well as in the cord blood sample included the following genes: *AC092580.3, MAS1, BFSP2-AS1, SORCS1, CCL20, IL22, NCR2* (*=NKp44*), *AC092580.4, REEP1, BMP7, EVC, SEC14L6, OPCML, TNR, SEMA3A*. A set of 18 genes showed higher expression in both the cord blood sample and 5 out of 8 tonsil samples, but much lower expression in 3 other

samples (including both NKp44^{+} samples and the CD56^{-} sample from stage 3a): *CLEC10A, HLA-DQB2, SIRPA, CYP2S1, CACNA2D3, PADI2, FCN1, SLC7A11, MS4A4E, ADGRE4P, SLC45A3, PTGDR2, FCER1A, HCAR2, LILRB2, CFD*.

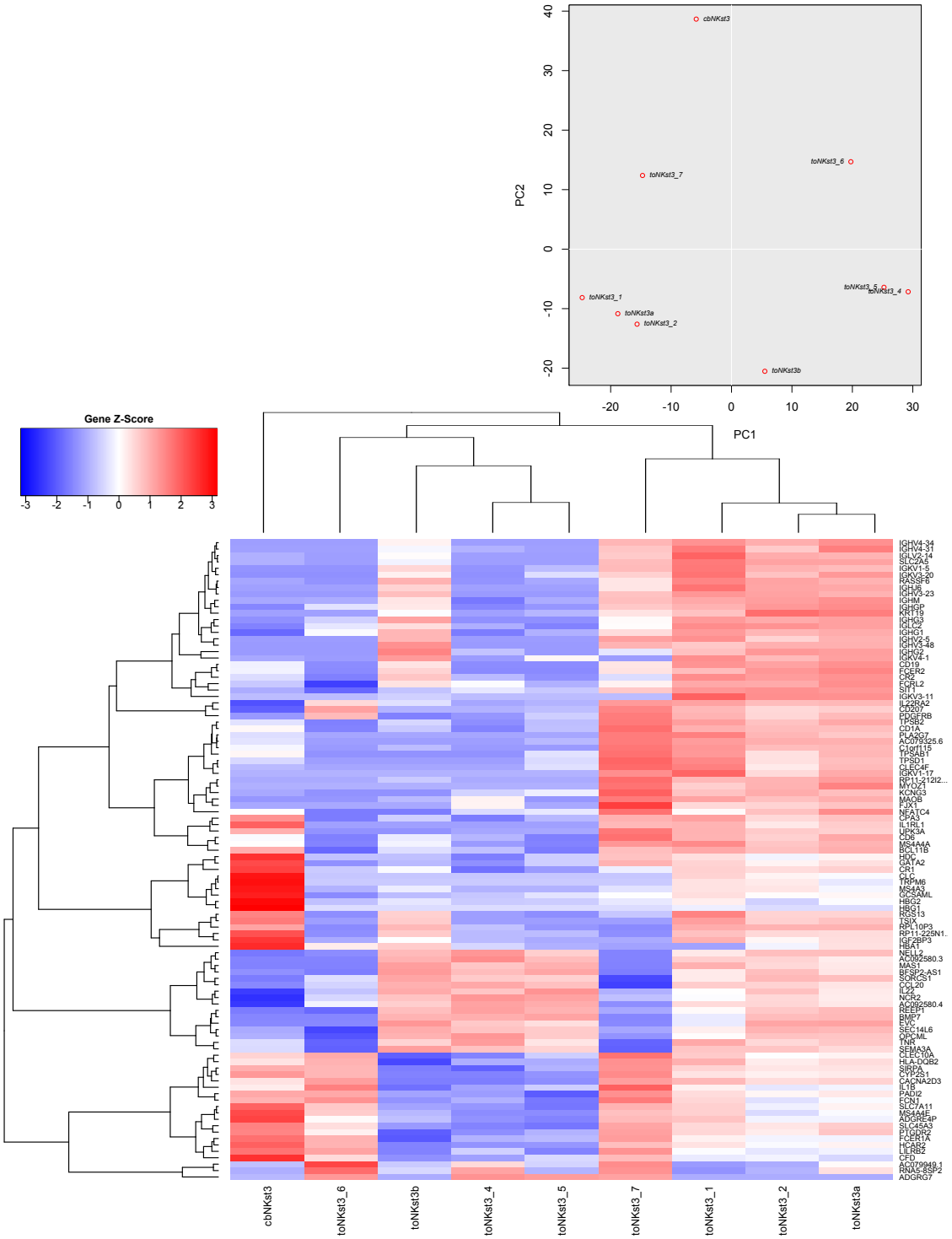

Figure 11 PCA score plot (A) and HC heatmap (B) of stage 3 *ex vivo* NK samples of tonsil- and cord blood origin, based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 20201, LoD=1.

Besides this, 4 out of 8 tonsil samples (toNKst3_1, toNKst3_2, toNKst3a and toNKst3_7) expressed a cluster of 28 genes with higher expression than in other samples. Most of these genes were related to the B cell lineage: 17 of them were gene segments of immunoglobulin chains: *IGHV4-34, IGHV4-31, IGLV2-14, IGKV1-5, IGKV3-20, IGHJ6, IGHV3-23, IGHM, IGHGP, IGKV3-11, IGHG3, IGLC2, IGHG1, IGHV2-5, IGHV3-48, IGHG2*, and *IGKV4-1*. Other genes included the specific B cell marker *CD19, FCER2, CR2, FCRL2, KRT19, SIT1, RASSF6, IL22RA2, CD207, PDGFRB*, and *SLC2A5*. Furthermore, this cluster also expressed CD22 – an inhibitory receptor and another B cell marker (Otipoby et al., 2001).

Expression of these genes in abovementioned samples was between 100 and 600 fpkm, which is lower than in B cells (3000-6000 fpkm), but still higher than average for stage 3 NK cells and NK cells in general. Considering that FACS sorting should have excluded $CD19^+$ cells, it can be that although these genes are transcribed, their products are not present at the protein level. Alternatively, the phenotype of given tonsil-derived NK cells could differ due to immune response, as tonsils were resected due to medical recommendations and there is high probability of inflammation in the tonsil samples. The theoretical possibility of B lymphocyte contamination also cannot be completely excluded.

3.3.2 Analysis of *ex vivo* NK stage 4 expression profiles

To examine whether the gene expression landscape differs between stage 4 NK cells from different tissues, 9 samples from *ex vivo* stage 4 samples were analyzed: 2 originated from bone marrow, 2 from cord blood, 2 from tonsils, and 3 from peripheral blood. Using LoD=1, 19924 genes were included in the analysis and based on the top 100 differentially expressed genes (Appendix 4), ANOVA and PCA were performed. A HC heatmap was created to visualize differences in expression of selected genes.

Similar to stage 3, also stage 4 samples turned out to be heterogeneous in their gene expression profiles. According to the PCA plot (Figure 12A), both bone marrow samples were localized far away from all other samples, samples from tonsils were also separated, while samples from cord blood and peripheral blood were localized relatively close to each other.

According to a clustered heatmap (Figure 12B), a specific gene set expressed in bone marrow samples consisted mainly of genes specific for the myeloid cell lineage (*CEACAM8, BPI, DEFA3, MMP8, CRISP3, LTF, PGLYRP1, RETN, CHI3L1, LCN2, PRTN3, ELANE, MPO, RNASE2, RNASE3, CTSG, DEFA4, AZU1, S100A9, S100A12, S100A8, FCN1, LYZ, PLBD1*) as well as *SLPI* and *CAMP* (expressed by different

leukocyte groups, including NKs), *MS4A3*, which is normally expressed by CD34^{+} cells, and *ABCA13* (Arai et al., 2013).

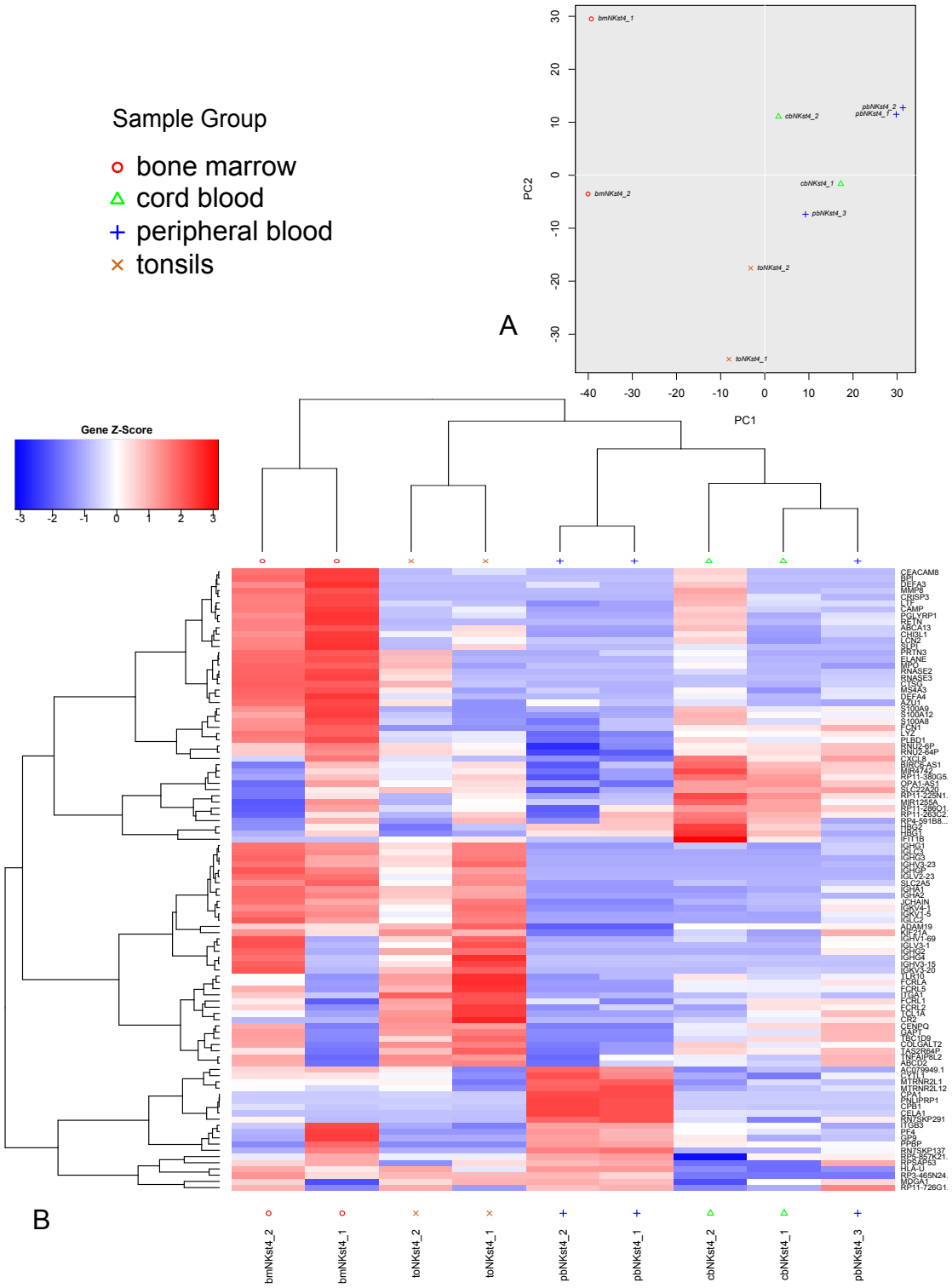

Figure 12 PCA score plot (A) and HC heatmap (B) of stage 4 ex vivo NK samples based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 20160, LoD=1.

Besides these genes, these samples highly expressed a cluster that included *ADAM19* (shown to be expressed by DCs), *KIF21A* (known to be expressed by B and NK lymphocytes), *SLC2A5* and 12 genes encoding immunoglobulin chains (*GHG1, IGLC3, IGHG3, IGHV3-23, IGHGP, IGLV2-23, IGHA1, IGHA2, JCHAIN, IGKV4-1, IGKV1-5, IGLC2*) (Lakschevitz et al., 2015; Miyake et al., 2004; Su et al., 2004; Surmiak et al., 2012; Yang et al., 2002). The same cluster was also expressed in both tonsil-derived samples.

Six more immunoglobulin-coding genes (*IGHV1-69, IGLV3-1, IGHG2, IGHG4, IGHV3-15, IGKV3-20*) together with other B lymphocyte specific genes (*FCRLA, FCRL5, FCRL1, FCRL2, CR2, GAPT, TBC1D9*) and other immune system-related genes (*COLGALT2, TAS2R64P, TNFAIP8L2, TCL1A, ABCD2, CENPQ, TLR10,* and *ITGA1*) formed a cluster, that was also expressed by tonsil-derived samples, but by only one of bone marrow-derived. Notably, only one gene in this cluster, *ITGA1*, is known to be important for NK cells, in particular for their trafficking (Grégoir et al., 2007).

A group of ten genes was expressed exclusively by two out of three peripheral blood samples. It included non-protein coding genes *AC079949.1, MTRNR2L1, MTRNR2L12, RN7SKP291* and protein-coding genes *PNLIPRP1, CYTL1, CPB1, CPA1, CELA1*. Interestingly, among these genes only *CYTL1* is known being expressed by lymphocytes (in particular by PBMCs and $CD34^+$ cells), while the other protein-coding genes known to play a role in other tissues (e.g. pancreas and skin). The third sample from peripheral blood shared an expression of a specific gene cluster with cord blood- and one of bone marrow-derived samples. This group included *CXCL8* and non-protein coding genes, namely *BIRC6-AS1, MIR4742, RP11-380G5.2, OPA1-AS1, SLC22A20, RP11-225N10.3, MIR1255A, RP11-286O18.1,* and *RP11-263C24.1*.

When ANOVA pairwise comparison between groups was performed based on the top 100 genes, it revealed that the number of genes whose expression significantly differed ($p<0.05$) between groups was the highest, when the bone marrow sample group was compared to any other: 37 genes were differently expressed between bone marrow and peripheral blood, 23 genes between bone marrow and cord blood, 16 genes between bone marrow and tonsils. When any other two groups were compared, the number of differentially expressed genes was lower: eleven genes were differentially expressed between peripheral blood and tonsils, nine genes between cord blood and tonsils, and only one, *RPSAP53* (ribosomal protein SA pseudogene 53), was differentially expressed between peripheral and cord blood (Appendix 5). Genes differentially expressed between

tonsils and cord blood, as well as between tonsil and peripheral blood, included mainly immunoglobulin coding genes (six and seven respectively), B cell specific gene *CR2* and immature NK marker *ITGA1* (*CD11a*).

To specify, which genes significantly differentiate bone marrow stage 4 samples from other samples of the same developmental stage, an additional ANOVA analysis was performed. In that case, a group of two bone marrow samples was compared to a group containing all other stage 4 samples (n=7) and genes differentially expressed between groups were identified. The expression of 34 out of the top100 genes significantly differed ($p<0.05$) between groups (Table 4). All these genes were higher expressed in the bone marrow sample group than in another one. Among them, 21 genes were characteristic for the myeloid lineage (expressed normally in erythrocytes, granulocytes or monocytes), seven genes encoded immunoglobulin chains and one was characteristic for early stages of hematopoiesis. The same gene groups were also observed, when bone marrow samples were compared to samples from tonsil, peripheral blood and cord blood separately. Furthermore, when ANOVA was performed among the top 400 genes, 88 genes differed between bone marrow and the group of other samples (Appendix 7). Among them 75 genes with higher expression in bone marrow-derived samples also mainly represented myeloid lineage and $CD34^+$ hematopoietic cell; 13 genes with higher expression in non-bone marrow group included seven non-protein coding genes and six protein-coding genes with broad expression within lymphocytes.

Table 4 Genes differentially expressed between BM-derived and non-BM derived stage 4 NK samples identified among top 100 genes using ANOVA analysis. Expression levels are indicated in Log_2 domain.

		Expression levels				Expression levels	
GeneID	**p Value**	**BM**	**non-BM**	**GeneID**	**p Value**	**BM**	**non-BM**
ABCA13	0.006997875	10.00	2.18	IGLC3	0.031006099	8.67	2.09
AZU1	0.000980616	12.13	3.60	IGLV2-23	0.00964024	9.31	1.49
BPI	0.000133552	10.46	0.63	LCN2	0.001579508	11.04	2.49
CAMP	0.001083936	11.12	1.98	LTF	0.003082246	13.25	3.42
CEACAM8	0.000440334	9.77	0.90	LYZ	0.006791651	13.16	6.73
CHI3L1	0.003755841	8.18	1.78	MMP8	0.002840424	9.98	1.09
CRISP3	0.002570828	9.37	1.15	MPO	0.003409805	13.61	3.91
CTSG	0.001571086	12.44	1.66	MS4A3	0.000812184	9.74	2.19
DEFA3	0.001465622	11.30	1.19	PGLYRP1	0.002404518	8.73	1.45
DEFA4	1.90E-05	11.55	0.97	PRTN3	0.002778924	12.13	1.98
ELANE	0.0013738	12.19	1.47	RNASE2	0.000295559	9.01	0.91
FCN1	0.02213523	10.12	2.96	RNASE3	9.08E-05	9.66	0.67
IGHA1	0.031039151	10.59	3.02	S100A12	0.013417495	9.92	2.67

IGHG1	0.042965989	13.01	3.81	S100A8	0.011079726	13.23	4.45
IGHG3	0.048660702	10.87	2.56	S100A9	0.007013196	13.93	5.95
IGHGP	0.00931305	9.39	1.58	SLC2A5	0.020687636	7.91	2.10
IGKV1-5	0.029075362	9.17	2.21	SLPI	0.000638316	7.87	0.76

Notably, when the analogical analysis was performed comparing sequentially peripheral blood, cord blood and tonsil samples against all others, no significant difference could be identified neither among the top 100, nor among the top 400 genes (not shown). These results suggest, that the expression landscape of stage 4 NK cells from bone marrow is significantly different from the other stage 4 samples that were analyzed, probably due to a less mature phenotype of bone marrow-derived samples. However, heterogeneity was also observed between stage 4 NK cells from tonsils, peripheral blood and cord blood.

3.3.3 Analysis of *ex vivo* NK stage 5 expression profiles

To study tissue- and functional-specific differences in gene expression among stage 5 NK samples, 29 samples in total were used: two of them originated from bone marrow, two – from cord blood, one – from tonsils, and 24 – from peripheral blood. Eighteen samples were derived from licensed and five samples from nonlicensed NK cells, and for five samples the licensing status was not defined. Furthermore, seven among the licensed samples were identified as memory-like NK cells.

Using LoD=1, 21903 genes from the complete transcriptome were included in the analysis; among them 100 of the most differentially expressed genes were selected for further analysis (Appendix 7). First, PCA and HC analysis were performed to estimate which factors influence the heterogeneity of samples and their clustering.

3.3.3.1 Tissue-specific differences

Unlike stages 3 and 4, sample groups observed in PCA (Figure 13A) plot did not correspond to sample origin. Even though bone marrow samples were localized together, they were included in a bigger group consisting of samples from cord blood and part of peripheral blood samples. The only sample group localized separately consisted of five peripheral blood samples: four of memory-like NK cells and one of nonlicensed. This sample group was also clearly visible on a HC heatmap (Figure 13B), as its members demonstrated dramatically higher expression of 79 out of top 100 genes (marked green in Appendix 7).

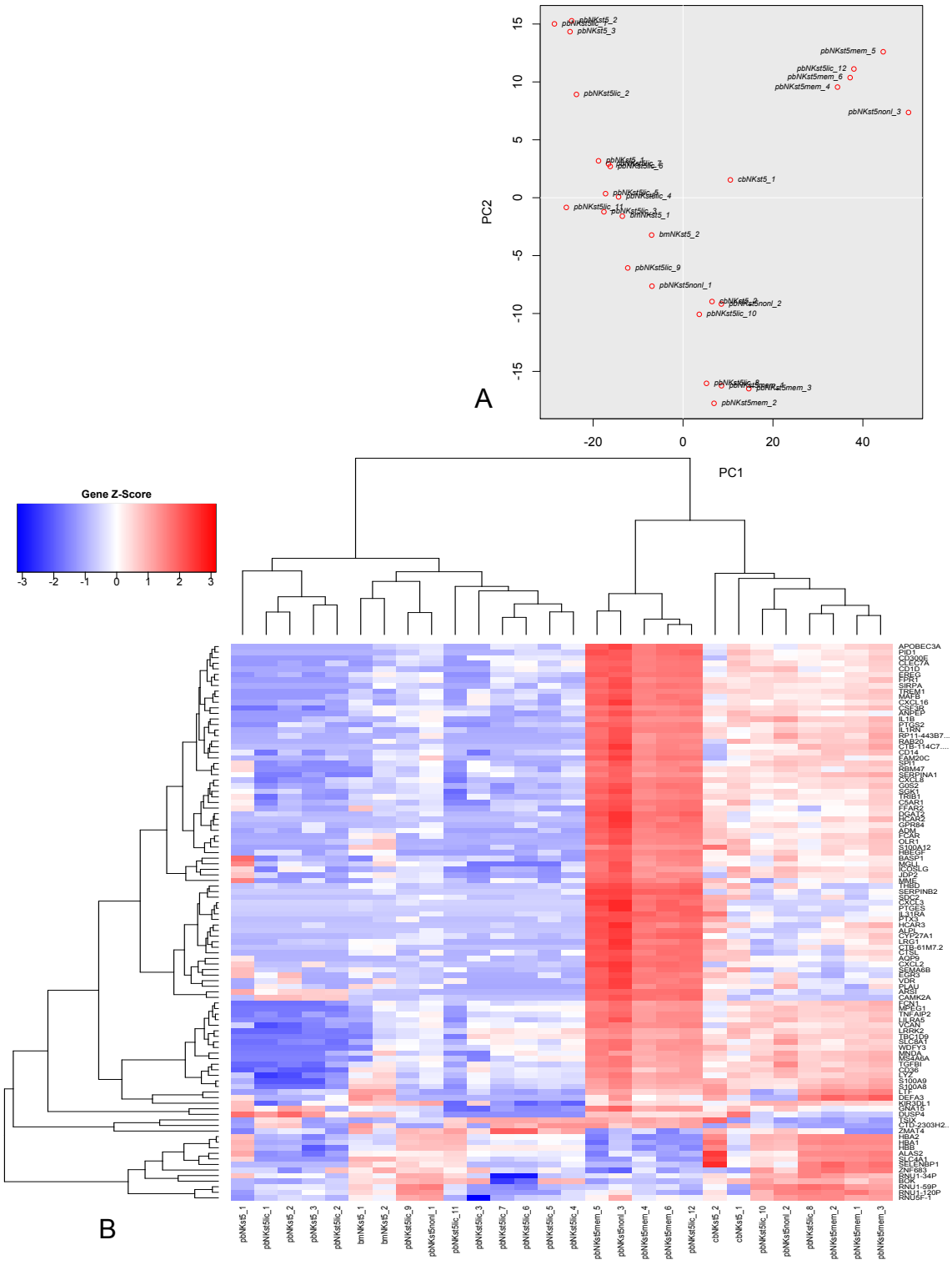

Figure 13 PCA score plot (A) and HC heatmap (B) of stage 5 *ex vivo* NK samples based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 21903, LoD=1.

This group included genes of different specificity: typically expressed by NK cells (e.g. *CXCL8* and *DUSP4*), markers of myeloid lineage (e.g. *CD14*, *CD300E, CXCL16, FCAR*), genes with ubiquitous expression within leukocyte populations (e.g. *G0S2,*

GPR84, IL1RN, MAFB) or within different human tissues (e.g. *ADM, EREG, PTGS2*), as well as genes, for which a specific function in lymphocytes was not yet described (protein coding, such as *ARSI, CAMK2A, PROK2* and three non-protein coding genes). Within all other stage 5 samples expression of these genes was highly variable.

Besides this, another group of five peripheral blood samples (including nonlicensed, licensed and memory-like samples) expressed a smaller cluster of another 16 genes, including those associated with erythroid lineage (*ALAS2, HBA2, HBA1, HBB, SLC4A1, SELENBP1*), T-specific (*ZNF683*), NK-specific (*KIR3DL1*), ubiquitously expressed among lymphocytes (GSTM1), six genes of small nuclear RNAs (*RNU1-34P, RNVU1-18, RNVU1-6, RNU1-59P, RNU1-120P, RNU5F-1*), and proapoptotic gene *BOK* (Figure 13B).

To check, whether there is any significant difference between samples with a different tissue origin, ANOVA analysis based on 100 of the most differentially expressed genes was performed. As tonsils were represented by only one sample, it was excluded from the analysis, while bone marrow, cord blood and peripheral blood samples were compared to each other. However, no significant difference between gene expression groups could be detected (not shown). Furthermore, even when ANOVA was performed using top 400 instead of top 100 genes, still no difference could have been detected (not shown).

3.3.3.2 Functional-specific differences

At the next step of an analysis it was investigated, how the licensing status may influence the NK cell expression profile. For this purpose only those stage 5 samples were selected, where the licensing status was known. Besides this, two KIR^-NKG2A^- samples were also excluded, as they might have represented less mature cell population in comparison to KIR^+ and/or $NKG2A^+$ NK cells. In this case, the analysis was based on 22 samples from peripheral blood: 3 of nonlicensed, 11 licensed, and 7 memory-like NK cells. Of note, when PCA and HC for these groups were performed, no group-dependent distribution was observed at the plot (Figure 14A, B).

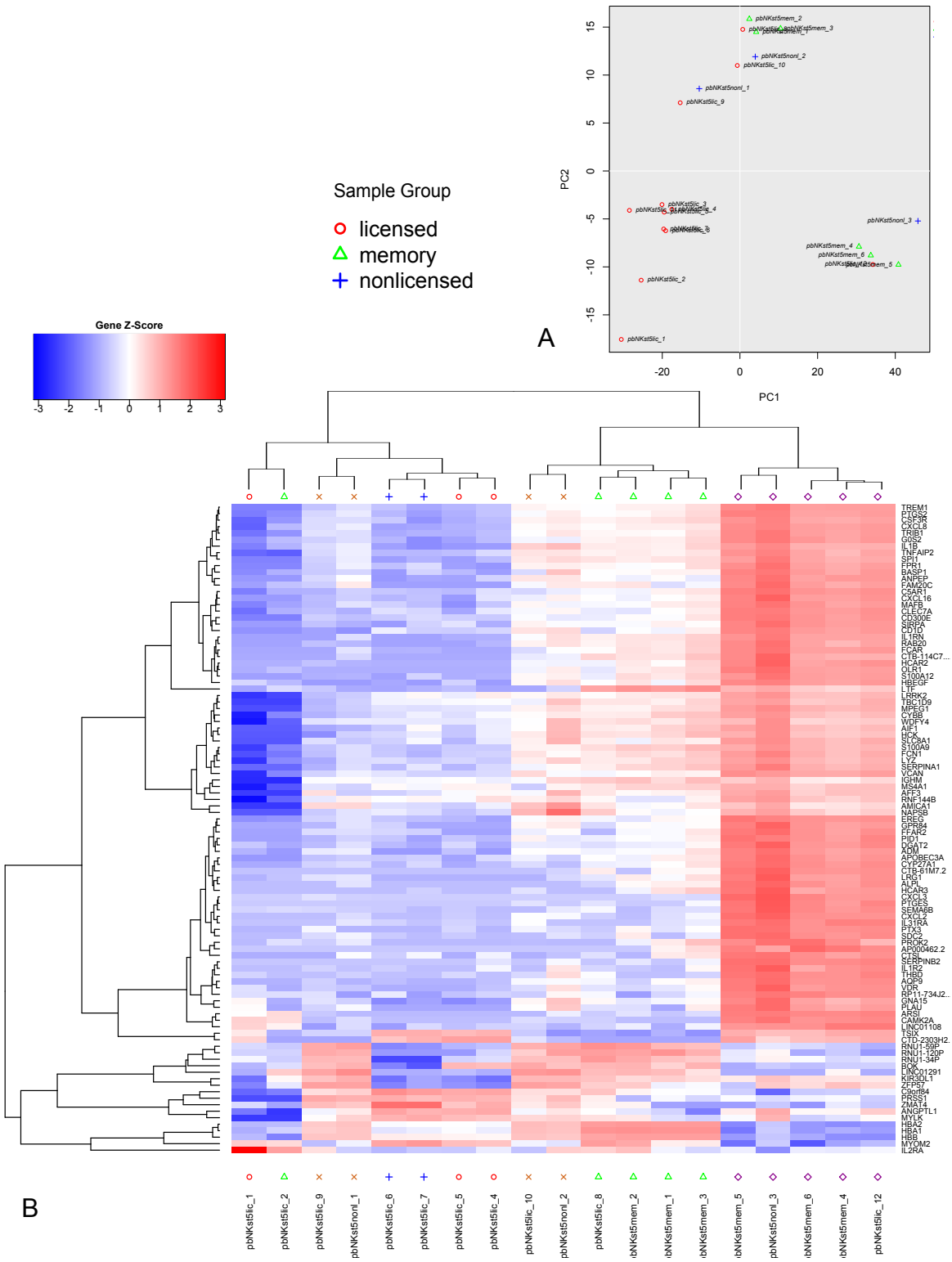

Figure 14 PCA score plot (A) and HCA (B) of licensed, nonlicensed and memory stage 5 *ex vivo* NK samples based on expression levels of the top 100 differentially expressed genes

When these three groups were compared pairwise by ANOVA, the analysis revealed that there is indeed significant difference in gene expression between them (listed in Appendix 8). Among the top 100 genes, significantly ($p<0.05$) differently expressed between licensed and nonlicensed were 22 of them (all with higher expression in

nonlicensed group, than in licensed), and 66 genes between licensed and memory (only *ADGRA3* was expressed higher in licensed group, all other genes had higher expression in memory group), and 2 genes between nonlicensed and memory group (*LTF* and *ADGRA3*). This, however, contradicted current knowledge about memory-like cells. As memory NK cells further differentiated licensed NK cells, they were supposed to share at least the same differences in expression landscape, as between licensed and nonlicensed sample groups, so the observed result was unlikely due to functional differences.

To search for another explanation, the data were revised again. As it was shown before, among stage 5 was a group of five sample, the expression profile of which differed obviously from all other stage 5 samples, and that group included both nonlicensed (n=1), licensed (n=1) and memory-like (n=3) NK samples, and such prominent difference could have affected the analysis. Looking back at the sorting data, it was found out, that all these five samples came from the same donor (donor 6, Table 5). Taking this into account, it was suggested, that specific characteristics of the donor determined an outstanding expression profile of these samples; this dominating expression profile, in turn, influenced an analysis, so that no other differences could be identified.

Table 5 Donor origin of selected stage 5 NK samples

Sample	Donor	Sample	Donor
pbNKst5lic_1	1	pbNKst5lic_9	5
pbNKst5lic_2	2	pbNKst5nonl_1	5
pbNKst5lic_3	3	pbNKst5lic_10	5
pbNKst5lic_4	1	pbNKst5nonl_2	5
pbNKst5lic_5	1	pbNKst5mem_4	6
pbNKst5lic_6	4	pbNKst5mem_5	6
pbNKst5lic_7	4	pbNKst5lic_12	6
pbNKst5mem_1	2	pbNKst5mem_6	6
pbNKst5mem_2	2	pbNKst5nonl_3	6
pbNKst5mem_3	2	pbNKst5lic_11	7
pbNKst5lic_8	2		

To check this hypothesis, ANOVA analysis was done again excluding the five abovementioned samples. Indeed, the results in this case were rather different. Among the top 100 genes, five were expressed higher in the nonlicensed group, than in the licensed group (*TGFBI, MS4A6A, PLD4, NAPSB, KCTD12*, and *CD4*), nine genes were differentially expressed between nonlicensed and memory cells (*DEFA3, LTF, MGAM2, SELENBP1, MPO, ADGRA3, LAG3, RP11-706O15.3,* and *DEFA4*), and 25 genes were differentially expressed between licensed and memory cells (*LTF, SELENBP1, G0S2,*

LAG3, ADGRA3, SERPINA1, S100A9, MPO, CDA, HLA-DQA2, CXCL8, FCN1, HBA1, RXRG, RP11-873E20.1, IL1B, HBA2, CCDC144A, CD4, RP11-706O15.3, LYZ, TNFAIP2, HCK, DEFA4, and *DEFA3*).

However, the higher difference between licensed and memory group, than between nonlicenced and memory, as well as between nonlicensed and licensed groups was still unexpected. Furthermore, when ANOVA pairwise comparison was performed based on top 400 genes, the ratio between gene numbers remained comparable: 14 genes were differentially expressed between licensed and nonlicensed cells, 27 genes were differentially expressed between nonlicensed and memory cells, and 74 genes were differentially expressed between licensed and memory cells (Appendix 9). Also when HC was performed using the same sample set, no clear gene expression patterns specific for a certain functional group was observed, even though in a PCA plot memory-like samples were grouped together and rather separately from other samples (Figure 15).

When the expression levels of the abovementioned genes were examined, it became clear that most of the top 100 genes were expressed relatively low in all stage 5 samples, e.g. *DEFA3* with 1.5 fpkm in nonlicensed samples and 287 fpkm in memory samples or *ADGRA3* with average levels 205 fpkm licensed cells, 47.9 fpkm in nonlicensed and 7.8 in memory cell group. Mathematically, the fold change between them is big enough, however, the biological significance of such difference is questionable. Such low expressed genes might play role, when rather homogenic group of samples is examined, or when differences within compared groups are relatively big. As an outcome, no biologically significant difference can be detected.

Considering all observations mentioned above, it was suggested, that besides tissue origin and functional state, also individual differences between stage 5 NK cells derived from different donors influence their expression profiles; and these influence is big enough to blur other differences.

3.3.3.3 Donor-specific differences

To check a role of donor-derived difference, one more round of analysis was performed. As it was known, from which donor every sample was derived, 21 samples with known licensing status were grouped according to their origin (Table 5). Two samples were excluded form the analysis in comparison to a previous analysis (pbNKst5lic_3 and pbNKst5lic_11), as these were the only samples from a particular donors, so no ANOVA analysis could be performed on them.

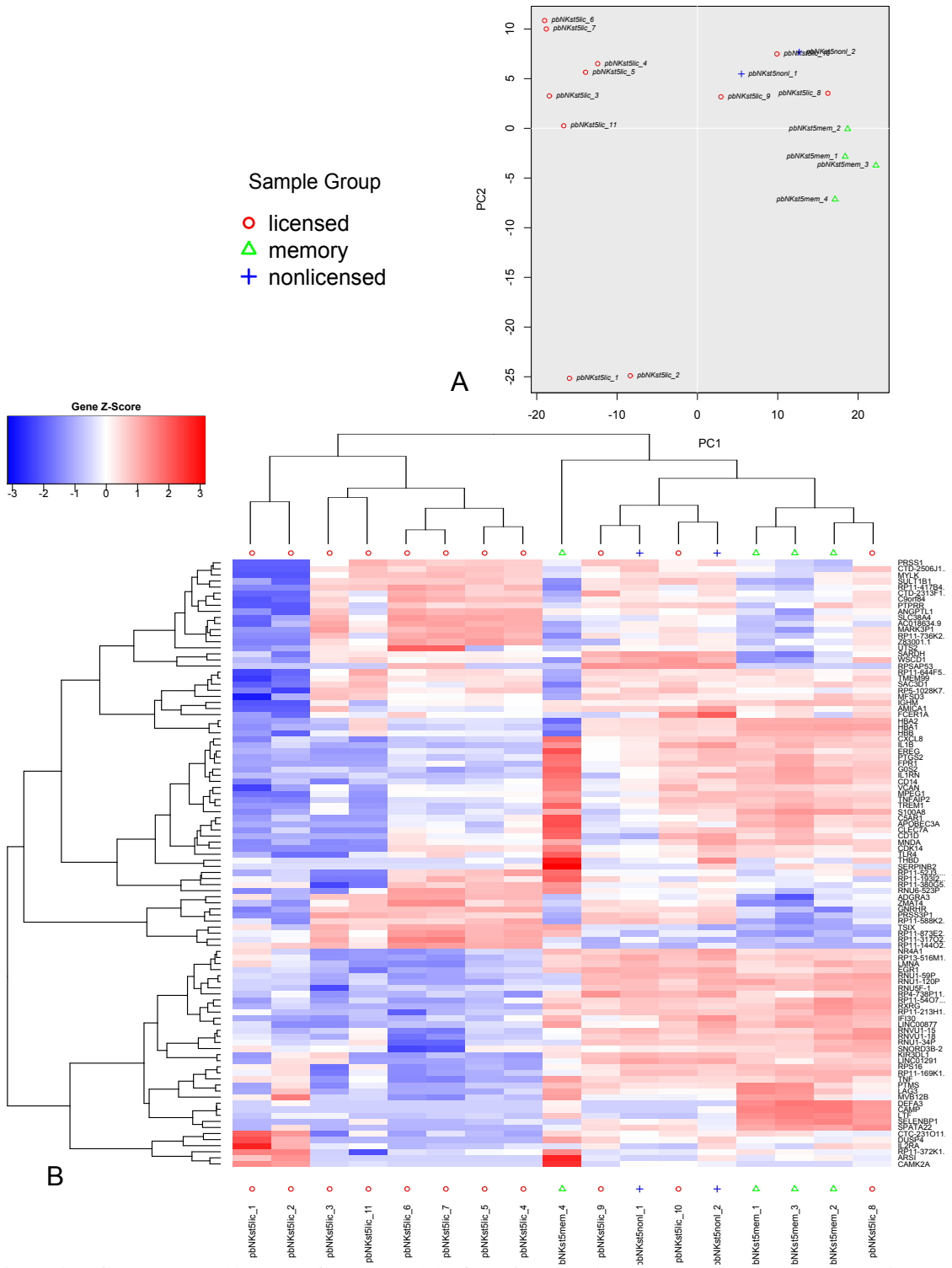

Figure 15 PCA score plot (A) and HC heatmap (B) of nonlicensed, licensed and memory stage 5 *ex vivo* NK samples (excluding samples from donor 6) based on expression levels of the top 100 differentially expressed genes

Genes were selected from the total number of 21903, LoD=1.

As it could be seen in the PCA plot, the donor origin indeed influenced the localization of samples (Figure 16A). Both in PCA plot and in hierarchical clusters (Figure

16B) all five samples from donor 6 clearly clustered together, as well as both samples from donor 1, both samples from donor 4, and 4 out of 5 samples from donor 2.

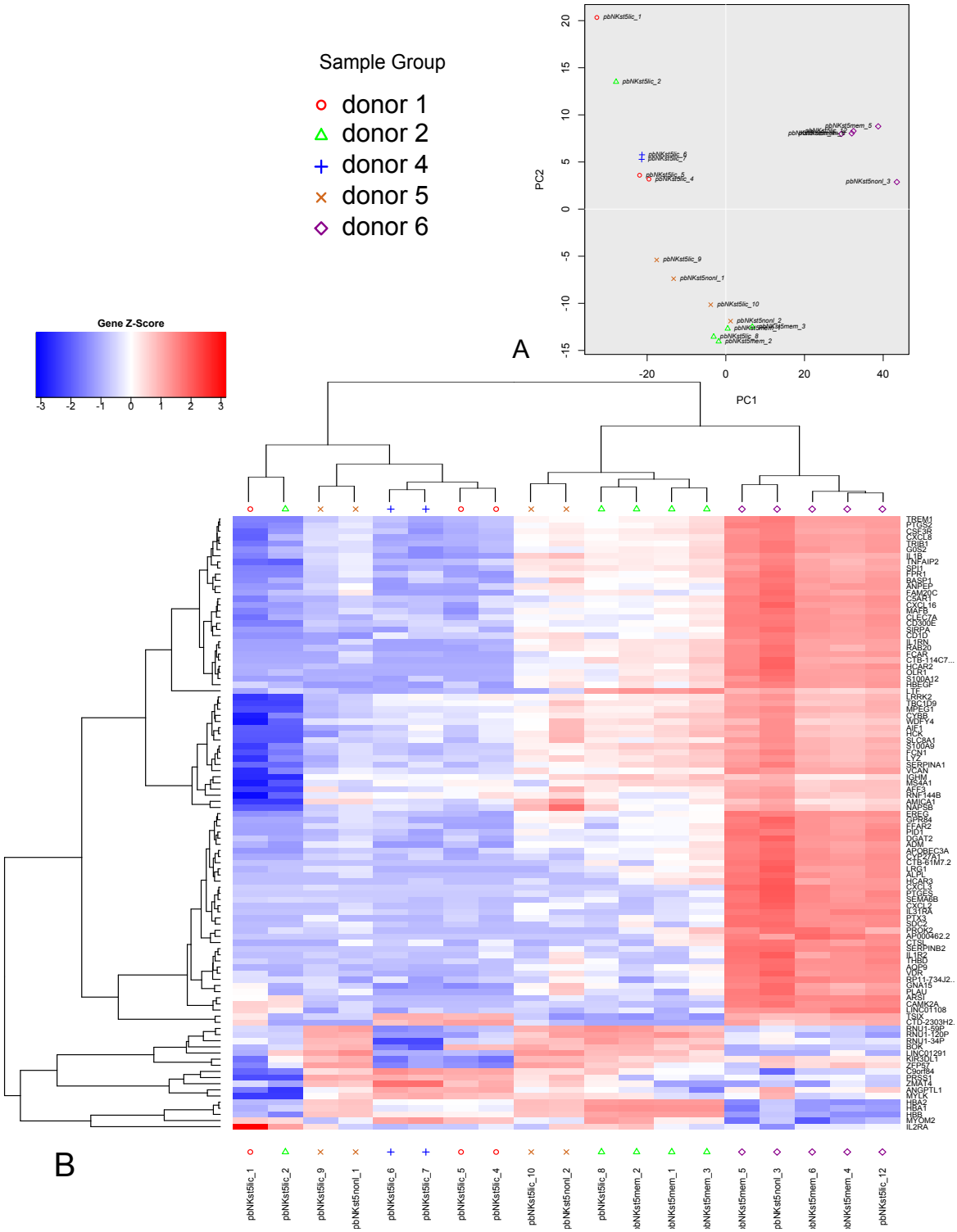

Figure 16 PCA score plot (A) and HC heatmap (B) of stage 5 *ex vivo* NK samples grouped according to donor origin, based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 21903, LoD=1.

Moreover, when ANOVA pairwise comparison was performed for these donor-based groups, it was shown that up to 85 genes among the top 100 genes were differently

expressed ($p<0.05$) between different groups (Table 6). Even though most of the diversity was obviously derived from the donor 6-originated group, significant difference in expression from 2 to 22 genes was also observed between other groups.

Table 6 Number of genes among top 100 differentially expressed between *ex vivo* stage 5 NK sample groups derived from different donors

Donor 2	22			
Donor 4	2	17		
Donor 5	15	4	15	
Donor 6	74	72	68	75
	Donor 1	**Donor 2**	**Donor 4**	**Donor 5**

Taken together, all these results suggest, that expression landscapes of stage 5 *ex vivo* NK cell populations depend on all three studied factors: tissue specificity, licensing status and donor origin. Individual differences between donors contribute not less or even more than functional specificity to the sample diversity, so that transcriptional changes upon licensing could not be identified by the given methods.

3.3.3.4 Correlation network of gene expression within nonlicensed, licensed and memory-like stage 5 NK cells

As it was stated above, studying of transcriptional changes upon cell licensing is complicated due to multiple factors that influence the transcription landscape of stage 5 NK cells. Instead, to investigate functional changes in gene expression network, the correlation between gene expression levels among nonlicensed, licensed and memory-like stage 5 NK samples was performed using Spearman's method.

The list of genes selected for the analysis included NK-cell specific genes identified in the analysis of all samples (chapter 3.1) and stage 5-specific genes identified in the analysis of NK differentiation stages (chapter 3.3.1). In addition, 100 genes were selected from the complete sequencing data that were shown to be highly and specifically expressed in stage 5 NK cells and are of relevance for the immune system. The total number of genes in the assay accounted to 192. Spearman's correlation coefficient **r** was counted for each gene pair, correlations with value of correlation coefficient $r>0.7$ or $r<-0.7$ ($p<0.05$) were considered as significantly strong and were visualized in networks of positive (Figure 17) and negative (Figure 18) correlations respectively. More precise analysis was than narrowed down to genes known being related to licensing and memory acquisition were analyzed.

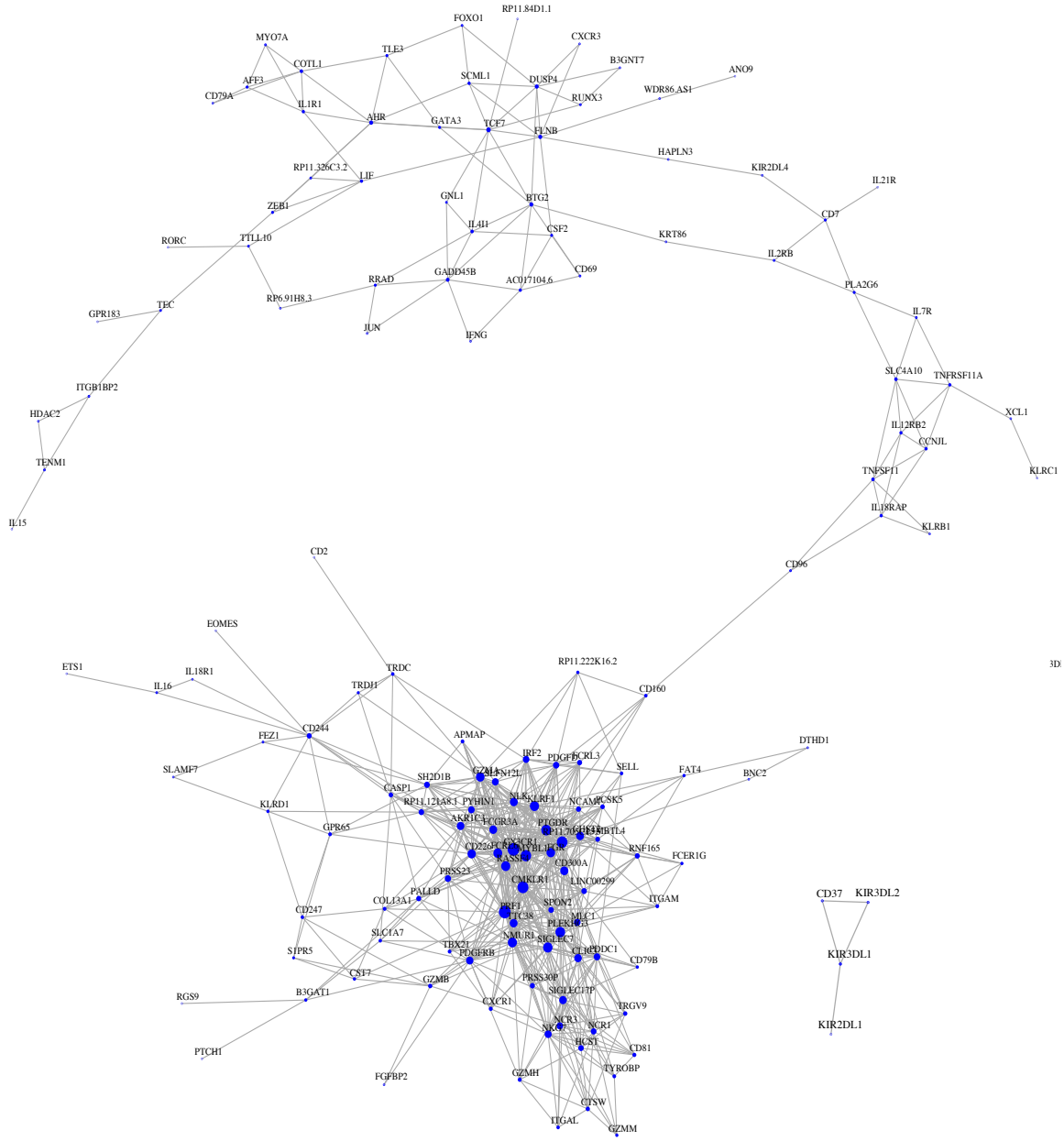

Figure 17 Correlation network of gene expression within stage 5 *ex vivo* NK samples
Only positive correlations with Spearman's coefficient r>0.7, p<0.05 are shown.

Even though it was claimed that transcription profiles of licensed and nonlicensed NK cells almost do not differ (Anfossi et al., 2006), two cell receptors, KLRD1 (CD94) and CD226 (DNAM-1), were previously shown to be differentially expressed between these groups and their expression levels correlate with licensed status of NKs (Anfossi et al., 2006). For this reason, correlations in which these genes were involved where analyzed specifically (Table 8).

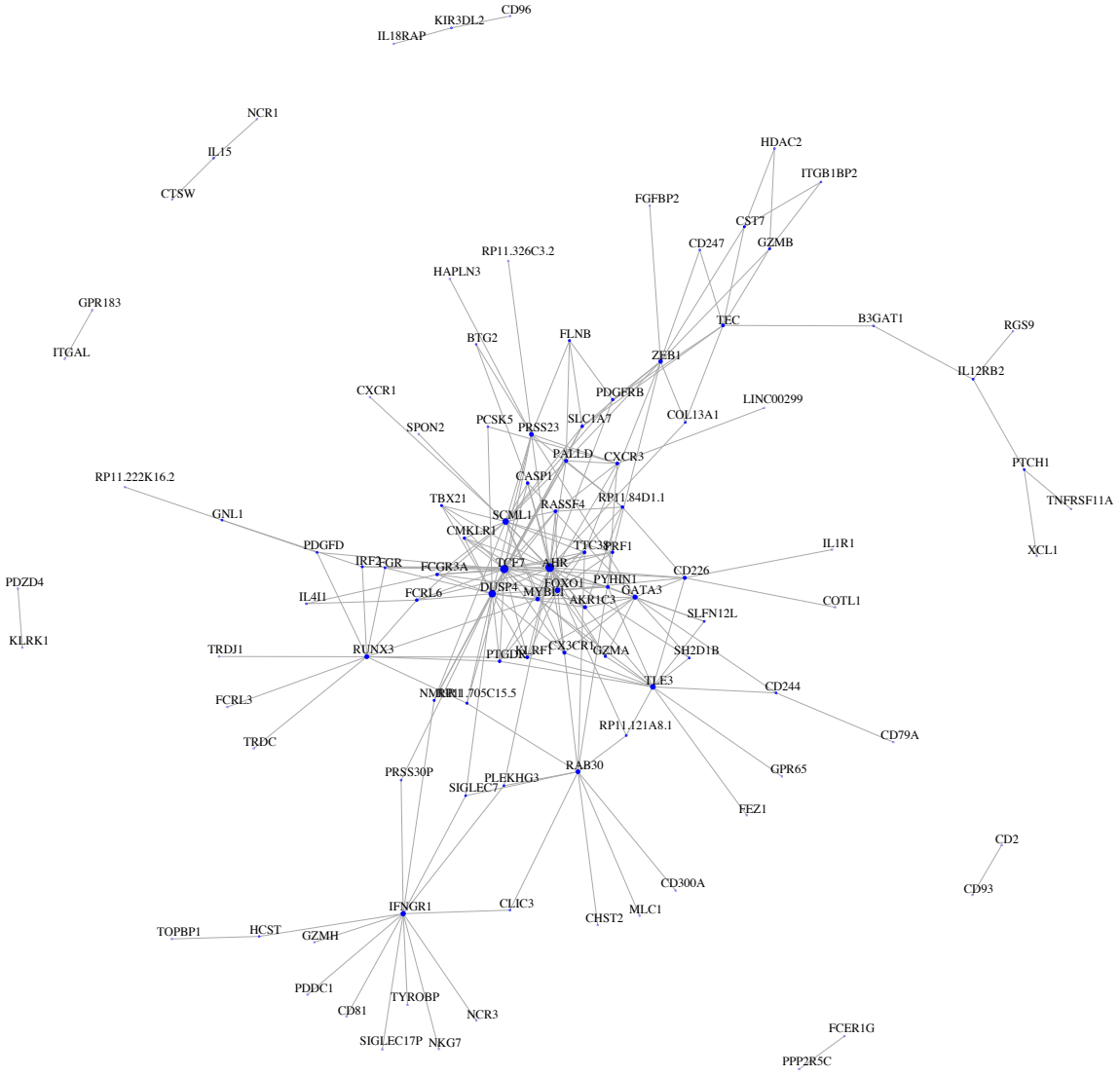

Figure 18 Correlation network of gene expression within stage 5 *ex vivo* NK samples
Only negative correlations with Spearman's coefficient r<-0.7, p<0.05 are shown.

Eight positive correlations with coefficients >0.8 were identified for the *CD226* gene and include correlations with genes associated with NK cell effector function and activation *PRF1* (0.81), *KLRF1* (0.8), *AKR1C3* (0.86), and *SH2D1B* (0.87), genes involved in cell migration *CX3CR1* (0.83) and *COL13A1* (0.84), transcription activator *MYBL1* (0.82), as well as non-protein coding gene *RP11-121A8.1* (0.83). Strong negative correlations (r<-0.8) of *CD226* were observed with *AHR* (coefficient -0.88), which encodes a transcription factor important for ILC3 development; and between *CD226* and *TLE3* (-0.84) that encodes a transcriptional co-repressor protein. Also multiple correlations with coefficients between 0.7 and 0.8 were identified, including correlations to genes typically

expressed in mature NK cells (*FCGR3A, FCRL6, GZMA, CMKLR1, GPR65, PRSS23, CHST2*), genes involved in cell cycle regulation (*PYHIN1, NLK, SLFN12L*), immunologically-related genes (*SPON2, PALLD, PTGDR*), and other genes whose functions are not so well defined (*RASSF4, TTC38, MLC1*, and lncRNA gene *RP11-705C15.5*) (Tan et al., 2012). Negative correlations with coefficients between -0.8 and -0.7 were observed between *CD226* and *COTL1, FOXO1, GATA3, IL1R1, TCF7*, and non-protein coding gene *RP11-84D1.1* (Table 7).

Table 7 Correlations involving markers of licensed (*CD226* and *KLRD1*) and memory-like (*SELL*) NK cells. Only significantly strong correlations are shown

Gene name	CD226	KLRD1	SELL	Gene name	CD226	KLRD1	SELL
AHR	-0.88			PALLD	0.74		
AKR1C3	0.86			PDDC1			0.71
CD244		0.72		PDGFD			0.72
CD247		0.72		PDGFRB			
CD300A			0.71	PRF1	0.82		
CHST2	0.72			PRSS23	0.75		
CMKLR1	0.76			PTCH1			
COL13A1	0.84			PTGDR	0.71		
COTL1	-0.71			PYHIN1	0.80		
CX3CR1	0.83			RASSF4	0.77		
FCGR3A	0.76			RGS9			
FCRL6	0.70			RP11.121A8.1	0.83	0.70	
FGR			0.70	RP11.222K16.2			0.75
FOXO1	-0.71			RP11.705C15.5	0.71		
GATA3	-0.77			RP11.84D1.1	-0.75		
GPR65	0.72	0.77		SH2D1B	0.87		
GZMA	0.79			SLAMF7		0.74	
IL12RB2				SLC1A7			
IL1R1	-0.74			SLFN12L	0.72		
IRF2			0.71	SPON2	0.76		
KLRF1	0.81			TCF7	-0.73		
MLC1	0.73			TEC			
MYBL1	0.82			TLE3	-0.84		
NLK	0.75			TTC38	0.73		

The *KLRD1* gene expression showed three positive correlations with coefficients >0.7: with expression of two activating receptors *CD244* (alias *2B4*) and *CD247* (alias *CD3z*), with those of G-protein coupled receptor *GPR65, SLAMF7* (activating/inibitory

NK receptor) and with those of non-protein coding gene *RP11-121A8.1* (Kim et al., 2013; Krämer et al., 2013).

Among the genes that are known to be more strongly expressed in memory-like NK cells are CXCR6, CD62L (SELL), and activating receptors (in particular NKG2C in humans) (Juelke et al., 2010; Lopez-Verges et al., 2010; Sun et al., 2009). Therefore, correlations in which these genes were involved were analyzed specifically.

Six correlations with coefficient >0.7 were identified for the *SELL* (*CD62L*) gene, including *CD300A, FGR, IRF2, PDDC1, PDGFD*, and non-protein coding *RP11-222K16.2* (Table 7).

No strong correlations were, however, observed, in which *NKG2C* and *CXCR6* genes were involved (not shown).

3.4 Comparison of expression profiles of *ex vivo* and *in vitro* NK developmental stages

To examine whether the *in vitro* differentiation assay is a good model for studying NK development, their expression profiles were compared to profiles of *ex vivo* samples from respective stages.

3.4.1 Stage 3

Within stage 3 two sample groups were formed: the *ex vivo* group included 11 samples described above; the *in vitro* sample group included three stage 3 samples produced by the NK differentiation assay. ANOVA and PCA were performed using 100 of the most differentially expressed genes; HC heatmap and violin plots were created to visualize differences in expression of the selected genes.

Despite heterogeneity between the *ex vivo* samples when PCA was performed, samples from the differentiation assay formed a clearly separated group (Figure 19A). In a HC heatmap as well, a branch containing all three *in vitro* samples was observed (Figure 19B). A higher expression of 11 genes was observed in the *in vitro* samples in comparison to the *ex vivo* sample group: *RP11-1094M14.5, E2F3P1, RP11-1094M14.8, KLRC2, TRAT1, KLRC1, IL26, GZMA, LZTS1, NTRK2*, and *PEG10*.

Further, to specify genes that were significantly differently expressed between *ex vivo* and *in vitro* stage 3 sample groups, ANOVA analysis was performed. This analysis showed that expression levels of 26 out of top 100 genes differed significantly ($p<0.05$) between the two groups (Table 8). Besides the gene group mentioned above, *in vitro* samples also expressed higher levels of *NCAM1, TMEM200A, B4GALNT1, TOMM20P2, PVRL3* and *PBK* (marked in green in Table 8); while *ex vivo* samples expressed higher

levels of *ADGRE1, AKR1C3, BTG1P1, CACNA1D, CDH1, CH25H, DTX1, IGHG4, IL5RA, KRT17P8*.

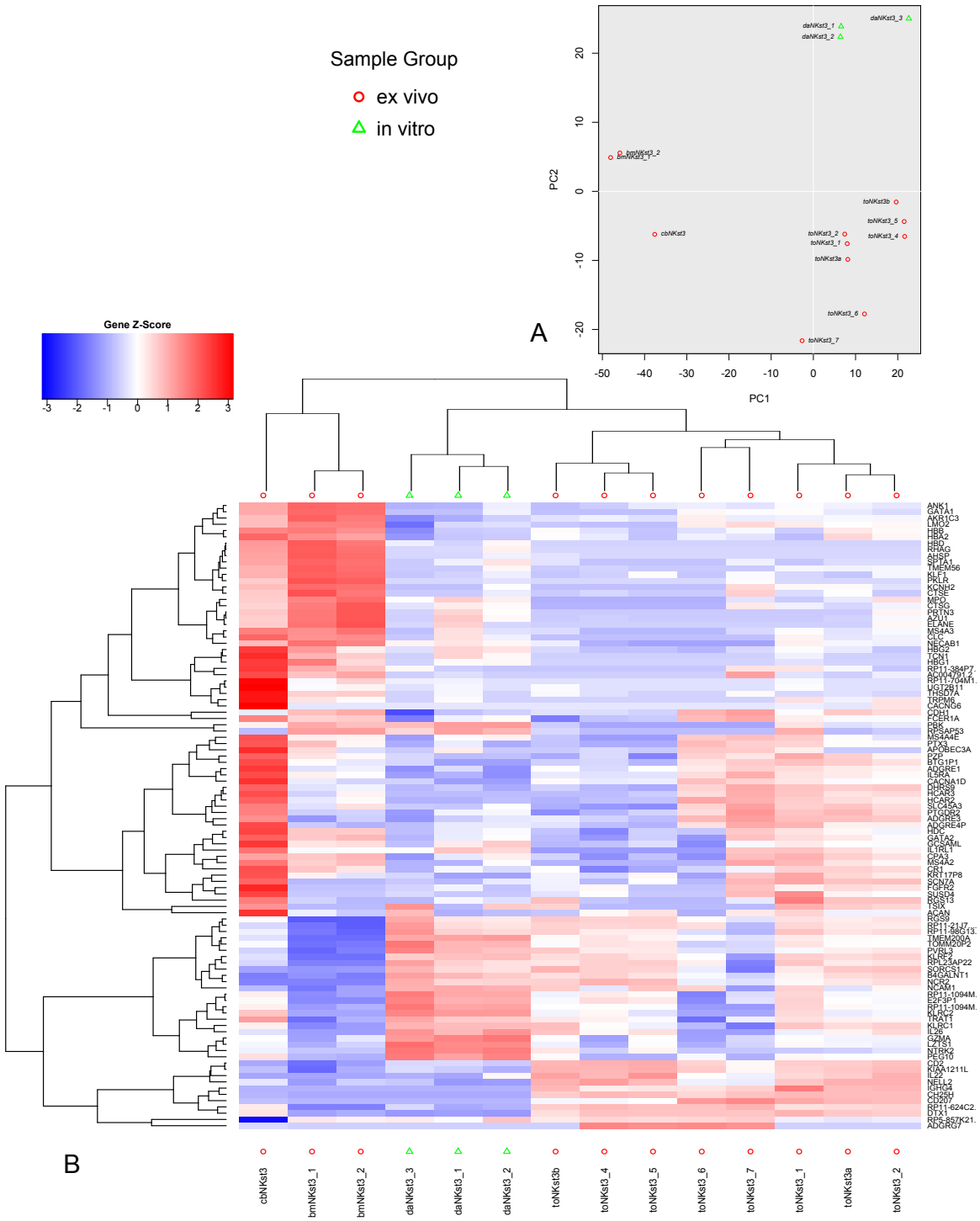

Figure 19 PCA score plot (A) and HC heatmap (B) of *ex vivo* and *in vitro* stage 3 NK samples, based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 20520, LoD=1.

Notably, such genes as *KLRC1, KLRC2, GZMA*, and *PBK* are normally expressed in mature NK cells, and *IL26* is expressed in NK- and T-cells upon activation.

In order to confirm this tendency, ANOVA analysis was performed again using the top 400 genes. In that case, 115 out of 400 genes (28,75%) were expressed differently between the groups (Appendix 10); and besides already mentioned above, genes with higher expression in *in vitro* NKs included *KLRC3, KLRC4, KIR2DL4, SLFN12L*, which are normally expressed at later differentiation stages.

Table 8 Genes differentially expressed between *ex vivo* and *in vitro* stage 3 NK populations identified from top 100 genes using ANOVA analysis. Expression levels are indicated in Log_2 domain.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Ex vivo	In vitro			Ex vivo	In vitro
B4GALNT1	0.049607825	5.10	9.38	TMEM200A	0.007448746	5.14	10.23
E2F3P1	0.007612725	3.84	8.87	TOMM20P2	0.007000877	4.27	8.76
GZMA	5.22E-05	3.03	10.59	ADGRE1	0.017518438	4.48	0.85
IL26	0.006537708	2.70	7.32	AKR1C3	0.021980398	5.93	1.85
KLRC2	0.001137854	2.56	8.32	BTG1P1	0.039814816	3.19	0.49
LZTS1	4.95E-06	2.17	9.51	CACNA1D	0.031633645	3.22	0.49
NCAM1	0.027386606	6.02	11.53	CDH1	0.006806712	8.24	3.64
NTRK2	8.18E-07	1.12	9.27	CH25H	0.019515034	4.64	0.00
PBK	0.01290095	3.03	8.33	DTX1	0.003035575	6.60	0.10
PEG10	0.000166902	2.28	9.31	IGHG4	0.036326526	5.29	0.00
PVRL3	0.019278926	4.58	8.40	IL5RA	0.005842729	4.15	0.52
RP11-1094M14.5	0.001449569	3.31	9.54	KRT17P8	0.042201187	3.27	0.49
RP11-1094M14.8	0.002847264	4.08	9.47	NELL2	0.018406665	7.03	1.19

3.4.2 Stage 4

Within stage 4, nine abovementioned *ex vivo* samples formed one group and the *in vitro* differentiation assay group consisted of three samples.

Hundred of the most differentially expressed genes were selected from the total number of 22794 genes using LoD=1 for PCA. HC heatmap and violin plots were created to visualize differences in expression of selected genes.

The PCA plot showed that the *ex vivo* samples do not form a tight cluster as opposed to the *in vitro* differentiation samples (Figure 20A). Also in the clustered heatmap (besides gene clusters specific for the outlying bone marrow and peripheral blood samples), a group of 16 genes specifically expressed in *in vitro* samples can be observed and include: *PTGIS, AK4, LIN28B, SIX4, CDH17, LINC00599, DEPDC1, NDFIP2,*

TGFBI, NUF2, ITGA1, AFAP1L2, RP11-98G13.1, GPR15, DLGAP5, and *PR11-677N16.1* (Figure 20B).

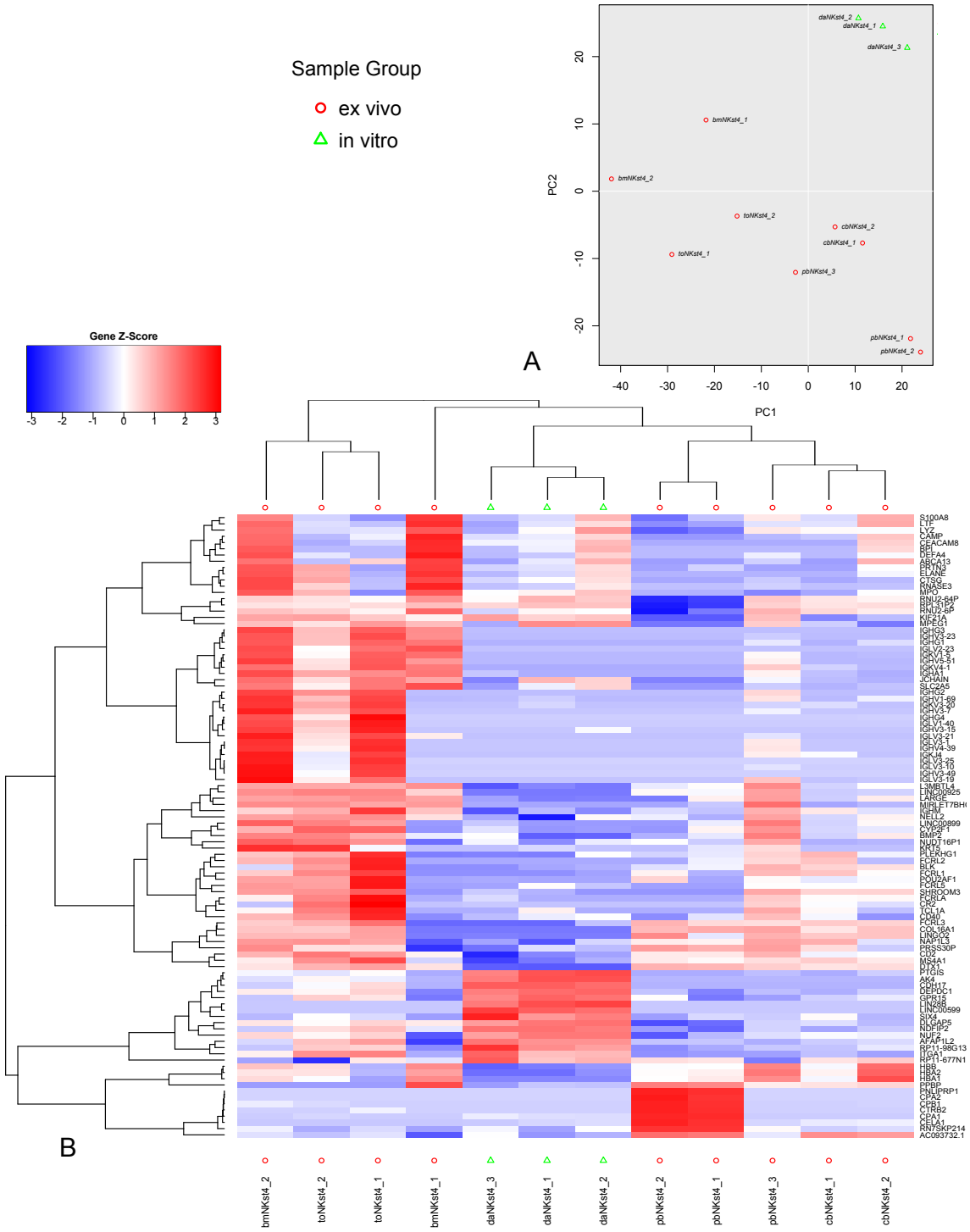

Figure 20 PCA score plot (A) and HC heatmap (B) of *ex vivo* and *in vitro* stage 4 NK samples, based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 22794, LoD=1.

In order to specify for which genes the expression levels significantly differ between *ex vivo* and *in vitro* groups, ANOVA analysis was performed. This analysis

showed that expression levels of 32 of the top 100 (Table 9) genes differ between analyzed groups with a significance of $p<0.05$. Among them, 15 genes were significantly higher expressed in the *in vitro* sample group and corresponded to the gene cluster described above, excluding *TGFB1* (marked in green in the Table 9), while the other 17 genes were higher expressed in the *ex vivo* sample group.

Table 9 Genes differentially expressed between *ex vivo* and *in vitro* stage 4 NK populations identified from top 100 genes using ANOVA analysis. Expression levels are indicated in Log_2 domain.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Ex vivo	In vitro			Ex vivo	In vitro
AFAP1L2	0.010273428	5.07	10.27	CD2	0.000342337	10.46	5.80
AK4	1.08E-06	0.73	8.49	COL16A1	0.001670428	4.48	0.00
CDH17	7.51E-06	0.77	7.61	DTX1	6.40E-08	8.64	0.03
DEPDC1	0.000398382	2.47	8.54	FCRL1	0.029697826	4.36	0.18
DLGAP5	0.005140536	4.43	9.88	FCRL3	0.008851847	7.92	1.28
GPR15	0.000980635	2.85	8.78	HBA1	0.008359766	6.14	0.04
ITGA1	0.04783988	5.31	10.58	HBA2	0.001886636	7.43	0.45
LIN28B	1.81E-10	0.00	8.21	HBB	0.002754031	10.16	2.27
LINC00599	3.08E-11	0.00	8.11	IGHM	0.009239485	7.30	2.63
NDFIP2	0.000994266	4.06	10.93	L3MBTL4	0.000493433	7.14	0.92
NUF2	0.007174789	3.96	9.03	LARGE	0.011181527	5.04	0.41
PTGIS	2.48E-06	1.00	8.23	LINC00925	0.001923865	4.70	0.00
RP11-677N16.1	0.040827167	5.00	8.52	LINGO2	0.00335457	4.23	0.03
RP11-98G13.1	0.001003641	2.25	8.23	MS4A1	0.001711099	7.22	1.67
SIX4	4.26E-05	1.60	8.57	NAP1L3	0.039197562	4.45	1.49
BMP2	0.030305944	5.10	1.44	NELL2	0.032183167	7.43	3.56

Furthermore, among the top 400 genes, 135 genes were differentially expressed between groups (Appendix 11), which are more than the highest difference within *ex vivo* samples (72 genes). These data suggest that besides high heterogeneity between different tissues, the *in vitro* differentiation assay displays important differences in gene expressions compared to analyzed *ex vivo* stage 4 NK cells.

3.4.3 Stage 5

To analyze differences between *ex vivo* and *in vitro* samples of stage 5 NK cells, a group of 30 *ex vivo* samples was compared to two stage 5 samples derived from the differentiation assays. PCA and ANOVA analysis were performed based on 100 of the

most differentially expressed genes, selected from a total number of 22010 genes (LoD=1). The HC heatmap was used to visualize differences in gene expression profiles.

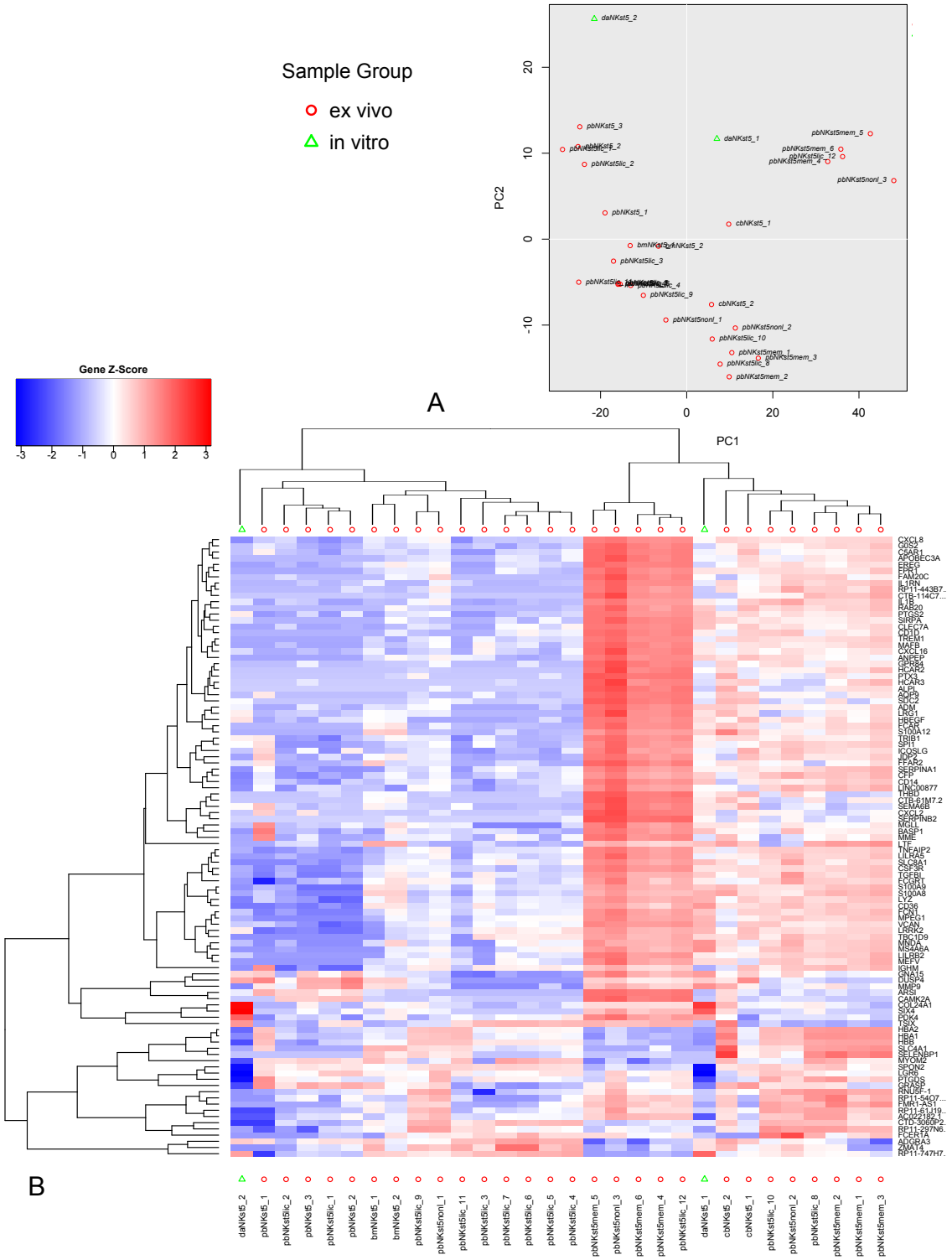

Figure 21 PCA score plot (A) and HC heatmap (B) of *ex vivo* and *in vitro* stage 5 NK samples, based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 21864, LoD=1.

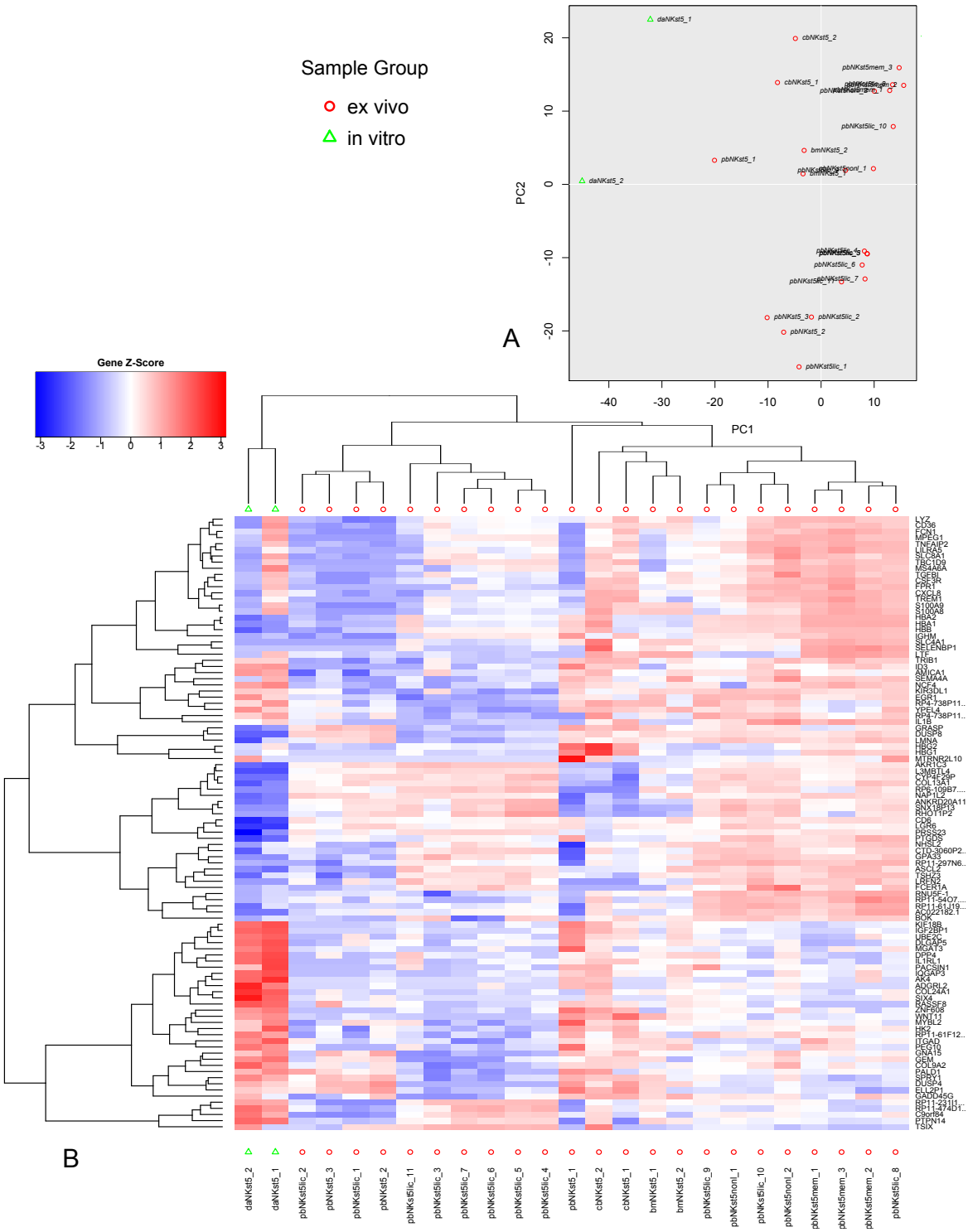

Figure 22 PCA score plot (A) and HC heatmap (B) of *ex vivo* and *in vitro* stage 5 NK samples (excluding samples from the donor 6), based on expression levels of the top 100 differentially expressed genes
Genes were selected from the total number of 21619, LoD=1.

As it is shown in Figure 21, both in PCA plot and HC heatmap the *in vitro* derived samples were not grouped together and no specific gene expression patterns could be identified. However, according to ANOVA analyzes, expression of 19 of the top 100 genes was significantly different ($p<0.05$) between *ex vivo* and *in vitro* sample groups, namely *AC022182.1, GRASP, HBB, RP11-474D14.2, LMNA, RP11-372K14.2, PDK4, HBA2, RHOU, RP11-747H7.3, HBA1, MYOM2, RP11-61J19.5, CTD-3060P21.1, IGHM, IFNLR1, SIX4, COL24A1*, and *DUSP8*, indicating that there is indeed a significant difference between compared groups.

As mentioned above, five samples from donor 6 were the main source of deviation within the *ex vivo* samples. Therefore, an additional analogical comparison between *ex vivo* and *in vitro* groups was done, but excluding five abovementioned samples. In this case, both in PCA plot and clustered heatmap samples from differentiation assay were clustered together and clearly separated from all *ex vivo* samples (Figure 22A).

Similarly, also specific patterns of gene expression were observed: a cluster of 50 genes with higher expression in the *in vitro* samples and a cluster of 11 genes, the expression of which was higher in *ex vivo* samples (Figure 22B). When ANOVA analysis was performed with these samples, 64 were defined as differentially expressed ($p<0.05$) between groups (Table 10). Among them 42 genes were expressed at higher levels in the *in vitro* samples (marked in green in the Table 10), and 21 genes in the *ex vivo* samples.

Table 10 Genes differentially expressed between *ex vivo* (excluding samples derived from the donor 6) and *in vitro* stage 5 NK populations identified from top 100 genes using ANOVA analysis. Expression levels are indicated in Log_2 domain.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Ex vivo	In vitro			Ex vivo	In vitro
KLRF2	1.10E-09	0.55	8.67	COL9A2	0.00597699	3.64	7.24
SIX4	5.66E-11	0.72	9.37	RP11-291B21.2	0.001434989	3.65	9.74
AK4	2.87E-08	0.83	7.88	RP11-231I16.1	0.041534398	3.75	6.90
ADGRL2	3.53E-07	0.87	7.46	C9orf84	0.010192002	4.26	8.57
IGF2BP1	2.06E-05	1.09	6.97	PALD1	0.028360153	3.85	7.02
ROBO1	7.30E-06	1.22	7.01	PTPN14	0.000132141	3.90	9.26
ENPP1	4.91E-06	1.42	8.07	HK2	0.010709339	3.92	7.49
LIF	0.000199058	1.44	7.83	GNA15	0.047382482	3.99	7.17
KIF18B	1.34E-05	1.54	7.69	SPRY1	0.032588834	4.88	8.84
MGAT3	0.000524308	1.66	6.86	MYC	0.025849789	4.89	8.19
E2F8	1.68E-05	1.67	7.57	SPRY2	0.029095291	5.12	8.95

CDCP1	1.35E-07	1.67	8.79	RHOT1P2	0.030321807	3.77	0.00
IQGAP3	5.45E-06	1.76	7.46	RP6-109B7.4	0.003713525	4.09	0.00
WNT11	0.010568965	1.79	5.46	PRICKLE2	0.00165742	4.65	0.00
COL24A1	6.23E-07	1.84	9.42	AC022182.1	0.00046285	4.77	0.00
ZNF608	0.005984163	1.92	6.09	SNX18P13	0.013579534	4.77	0.00
BIRC5	0.000394588	2.18	6.91	RP11-61J19.5	0.038777487	4.78	1.81
UBE2C	0.000122274	2.25	8.26	GPA33	0.00527239	5.56	1.62
PACSIN1	0.003658902	2.27	7.12	CTD-3060P21.1	0.024689696	5.89	2.39
RASSF8	2.63E-05	2.42	9.88	DUSP8	4.25E-05	6.21	0.00
IL1RL1	8.42E-06	2.50	9.79	COL13A1	0.000757249	6.63	0.96
PEG10	0.001368468	2.52	7.84	HBA1	0.00412863	7.03	0.25
ITGA1	4.42E-05	2.63	10.92	CYP4F29P	0.000265404	7.17	0.00
RP11-61F12.1	0.01077182	2.77	6.16	SLC1A7	0.000143426	7.19	0.00
DPP4	5.10E-05	2.79	10.46	ANKRD20A11P	0.000789528	8.09	0.46
DLGAP5	2.55E-05	2.83	9.70	L3MBTL4	4.57E-08	8.51	0.00
ID3	0.026581915	2.84	6.56	HBA2	0.001617629	8.78	1.31
RP11-474D14.2	0.003149095	2.88	6.87	AKR1C3	1.42E-05	8.91	3.02
NCF4	0.013730906	3.28	7.08	GRASP	0.002182389	8.95	2.67
SPTSSB	0.01647313	3.34	8.28	LMNA	0.007507693	9.03	4.46
GEM	0.003847056	3.37	7.81	PRSS23	2.04E-07	10.66	2.85
STYK1	0.000184234	3.46	10.24	HBB	0.000419536	11.64	1.84

In summary, these data suggest that also in NK developmental stage 5 there was a significant difference between *ex vivo* samples and samples derived from the *in vitro* differentiation assays. However, this difference was not higher, that within *ex vivo* group.

3.5 Differential expression of protein-coding and non-protein coding genes

The above-mentioned analyses clearly demonstrated that among the genes specifically expressed in certain cell types and in specific developmental stages both protein-coding and non protein-coding genes were identified. To investigate further how the expression landscapes of protein-coding and non protein-coding genes differ between NK developmental stages, these gene groups were analyzed separately. The data set for this analysis was the same as for chapter 3.3.1 and included 56 *ex vivo* NK samples from developmental stages 1 to 5.

Genes were sorted into protein-coding and non-protein coding based on their Ensemble description from the complete transcriptome data. Using LoD=1, 14449 protein-coding genes were selected for this analysis. The group of non-protein coding genes included lncRNAs, miRNAs, snoRNAs, overlapping antisense and intron transcripts, processed and unprocessed pseudogenes. Using LoD=1, 6218 non-protein coding genes were included in the analysis.

For both protein-coding and non-protein coding genes PCA, ANOVA analysis and HC were performed based on the top 100 differentially expressed genes (Appendices 12 and 13 respectively).

Both PCA and hierarchical heatmap (Figures 23 and 24) showed that successful distinguishing between developmental stages is possible based on expression of both protein-coding and non-protein coding genes.

A clear difference could be observed between expression profiles of early developmental stages (stage 1 and 2), committed NK cells (stage 3) and functional NK cells (stages 4 and 5). While in PCA plot based on protein-coding genes groups of stages 4 and 5 samples were separated with just a single exception (Figure 23A), in the plot based on non-protein coding genes these groups overlapped (Figure 23B). Similar to the analysis with the complete transcriptome (chapter 3.3.2), in the analysis based on only protein-coding genes the stage 3 samples from bone marrow clustered with earlier developmental stages, while the stage 3 sample from cord blood and the stage 2 sample from tonsils clustered in between these groups (Figure 23A).

Considering heatmaps of gene expression (Figure 24A), the samples clustered not only according to developmental stages, but partly according to the sample origin (for example stages 1 to 3 from bone marrow or stages 2 and 3 from tonsils).

When non-protein coding genes were used, samples from stages 1 and 2 formed a clear cluster that also included both stage 3 bone marrow samples (Cluster 1, Figure 24B). This cluster had a high expression of 22 genes (Table 11). The other stage 3 samples formed a separate cluster, which also included stage 2 sample from tonsils (Cluster 2, Figure 24B). These data are in accord with published data that both stage 2 and 3 cells from tonsils include ILC3 cells (Crellin et al., 2010b; Mebius et al., 1997). The samples of Cluster 2 partly shared gene expression patterns with stage 4 and 5 samples, but also had a big specific gene cluster (Figure 24B).

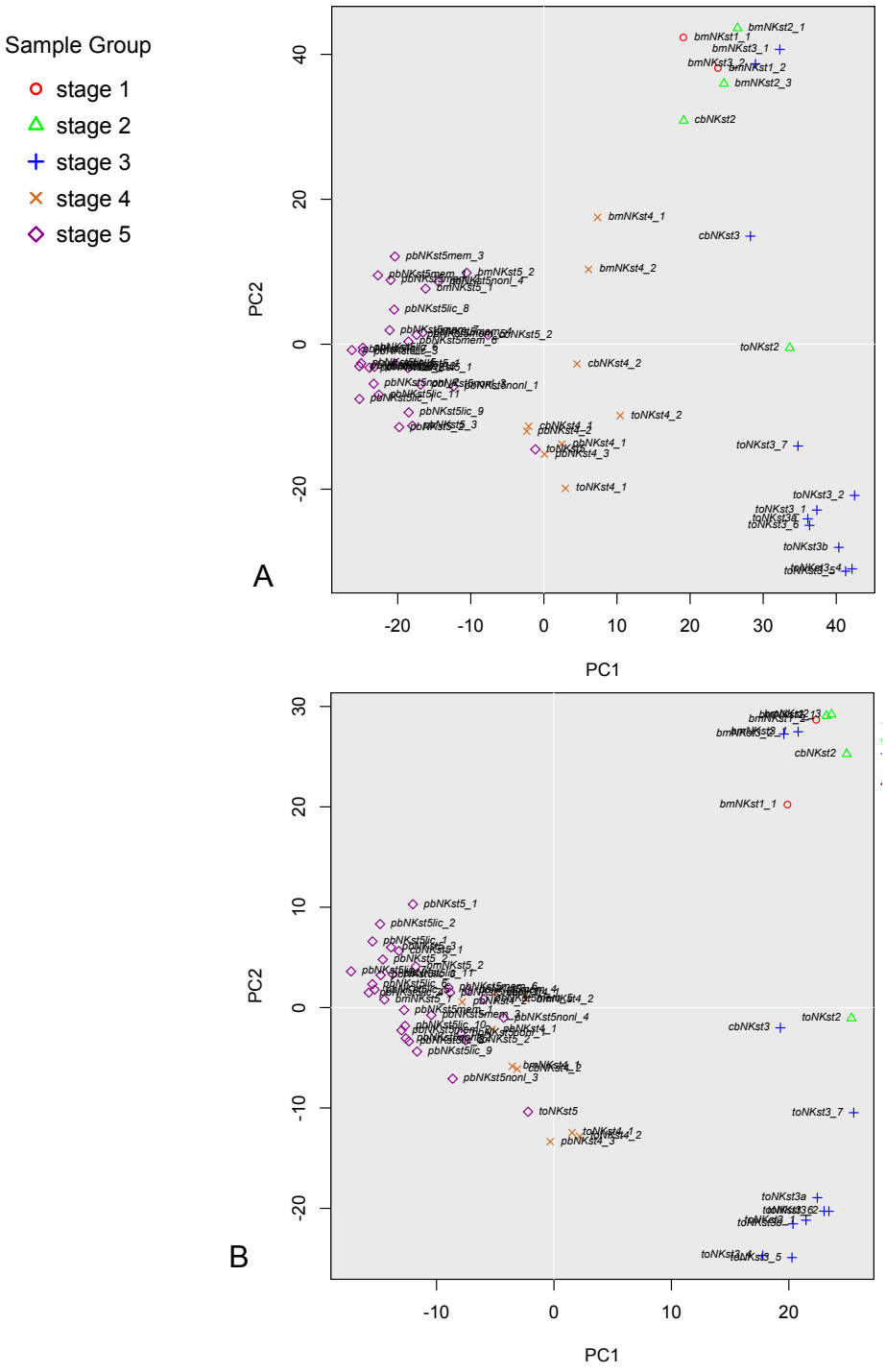

Figure 23 PCA score plots of *ex vivo* NK samples based on top 100 protein coding genes (A) and top 100 non-protein coding genes (B)

Protein-coding genes were selected from the total number of 14177 and non-protein coding – from the total number of 6613, LoD=1.

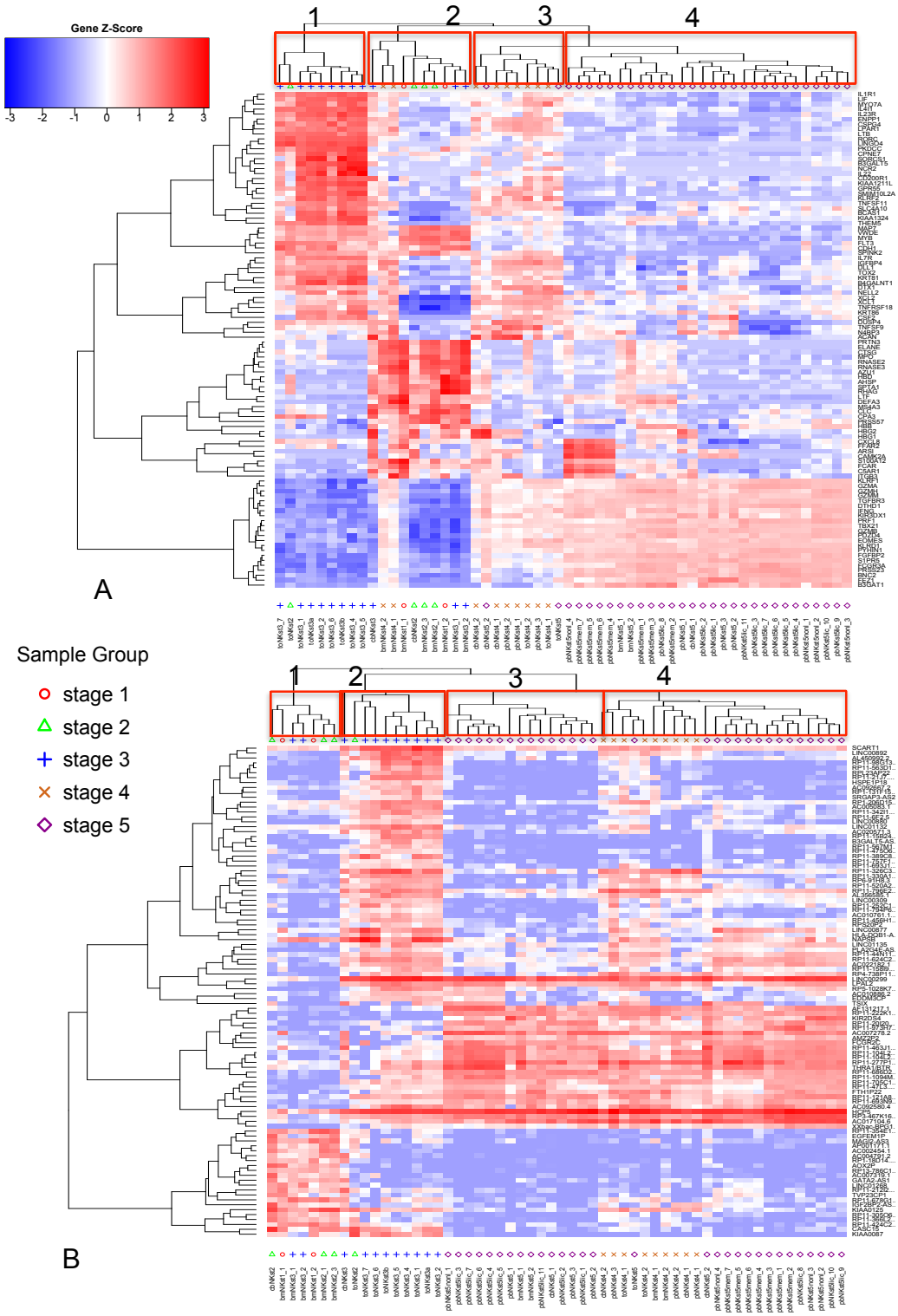

Figure 24 HC heatmaps of *ex vivo* NK samples based on top 100 protein coding genes (A) and top 100 non-protein coding genes (B)
Protein-coding genes were selected from the total number of 14177 and non-protein coding – from the total number of 6613, LoD=1.

While some genes from this cluster were also expressed at lower levels among stage 4 samples, at least 7 genes were expressed exclusively in non-bone marrow stage 3 samples: *RP11-15B24.5, B3GALT5-AS1, RP11-567M16.1, RP11-475O6.1, RP11-389C8.2, RP11-757F18.5, RP11-693J15.5*. Consistent with the sample distribution observed in the

PCA plot (Figure 23B), stage 5 samples were grouped in two clusters (Clusters 3 and 4, Figure 24B) with stage 4 samples intermingled in Cluster 4, suggesting that the gene expression differences in non-protein coding genes between stages 4 and 5 is smaller than in protein-coding genes.

Table 11 Expression patterns of non-protein coding genes among *ex vivo* NK samples

Cluster 1	**Cluster 2**		**Clusters 2-4**	**Clusters 3-4**
AC002454.1	AC005083.1	RP11-330A16.1	AC017104.6	AC007278.2
AC004791.2	AC010761.14	RP11-342I1.2	HCP5	AC010886.2
AC007319.1	AC020571.3	RP11-389C8.2	LINC00299	AC092580.4
AOX2P	AC022182.1	RP11-44N11.3	LPAL2	AF131217.1
AP001171.1	AC092667.2	RP11-456H18.2	RP3-467K16.4	AMZ2P2
CASC15	AL356585.1	RP11-475O6.1	XXbac-BPG181B23.6	EDDM3CP
EGFEM1P	AL450992.2	RP11-520A21.1		FCGR2C
GATA2-AS1	B3GALT5-AS1	RP11-563D10.1		FTH1P22
IGF2BP2-AS1	HLA-DQB1-AS1	RP11-567M16.1		RP11-104L21.2
KIAA0087	LINC00309	RP11-624C23.1		RP11-104L21.3
KIAA0125	LINC00877	RP11-693J15.5		RP11-1094M14.5
LINC01268	LINC00880	RP11-6F2.5		RP11-121A8.1
MAGI2-AS3	LINC00892	RP11-757F18.5		RP11-20I20.4
RP1-18D14.7	LINC01132	RP11-794P6.1		RP11-222K16.2
RP11-212I21.4	LINC01135	RP11-796E2.4		RP11-277P12.20
RP11-305O6.3	NAPSB	RP11-98G13.1		RP11-463J10.2
RP11-354E11.2	PLA2G4E-AS1	RP4-738P11.		RP11-47L3.1
RP11-366L20.2	RP1-131F15.2	RP6-91H8.3		RP11-686D22.10
RP11-424C20.2	RP1-206D15.6	RPL23AP22		RP11-693N9.2
RP11-678G14.3	RP11-158I9.5	RPS20P2		RP11-705C15.5
RP13-786C16.1	RP11-15B24.5	SCART1		RP11-973H7.1
TVP23CP1	RP11-252C15.1	SRGAP3-AS2		RP5-1028K7.2
	RP11-326C3.2			THRA1/BTR

As already seen in the PCA plot (Figure 23A), the protein-coding gene clustering appeared more diffuse (Figure 24A). While most of stage 3, 4 and 5 samples were clustered according to corresponding developmental stages. Cluster 2 was mixed and included samples from stages 1 to 4. This Cluster contained samples from bone marrow and one stage 2 sample from cord blood. Specific gene expression patterns of observed clusters are summarized in the Table 12. Thus, different from the non-protein coding genes, the protein-coding genes of stages 1 to 3 of the bone marrow samples do not cluster according to their developmental stages. This indicates that the expression of the non-protein coding genes more strictly follows the developmental pathway and suggests a role of these genes in developmental processes.

Table 12 Expression patterns of non-protein coding genes among *ex vivo* NK samples

Cluster 1		Cluster 2	Cluster 3	Cluster 4
B3GALT5	LIF	ACAN	DUSP4	B3GAT1
B4GALNT1	LINGO4	AHSP	N4BP3	BNC2
BCAS1	LPAR1	AZU1	TNFSF9	DTHD1
CD200R1	LTB	CLC	**Shared with cluster 1**	EOMES
CDH1	MAP7	CPA3	B4GALNT1	FCGR3A
CPNE7	MYB	CTSG	DLL1	FEZ1
CSPG4	MYO7A	DEFA3	DTX1	FGFBP2
DLL1	NCR2	ELANE	IGFBP4	GZMA
DTX1	NELL2	HBB	IL7R	GZMB
ENPP1	PKDCC	HBD	KRT81	GZMH
FLT3	RORC	HBG2	KRT86	GZMM
GPR55	SLC4A10	LTF	NELL2	IFNG
IGFBP4	SMIM10L2A	MPO	SPINK2	KIR3DX1
IL1R1	SORCS1	MS4A3	TNFRSF18	KLRD1
IL22	SPINK2	PRSS57	TOX2	KLRF1
IL23R	THEM5	PRTN3	XCL1	PDZD4
IL4I1	TNFRSF18	RHAG	XCL2	PRF1
IL7R	TNFSF11	RNASE2	**Shared with cluster 4**	PRSS23
KIAA1211L	TOX2	RNASE3	DTHD1	PYHIN1
KIAA1324	VWDE	SPTA1	EOMES	S1PR5
KLRF2	XCL1		GZMA	TBX21
KRT81	XCL2		GZMB	TGFBR3
KRT86			GZMH	
			GZMM	
			IFNG	
			KIR3DX1	
			KLRD1	
			KLRF1	
			PDZD4	
			PRF1	
			PYHIN1	
			TBX21	
			TGFBR3	

Of the protein-coding genes, Cluster 1 (Figure 24A) includes stages 2 and 3 samples of tonsil origin contains 46 genes. It included genes typical for early hematopoietic stages (e.g. *LIF, MYB*), as well as typical ILC3-related genes (e.g. *IL1R1, IL23R, LTB, IL22, RORC*). The abovementioned mixed cluster (Cluster 2, Figure 24A) included stage 1 to 4 samples from bone marrow, as well as stage 2 to 3 samples from cord blood. All of them expressed 19 genes at high level, most of which were associated with early hematopoietic stages (e.g. *ELANE, MPO, RNASE2, RNASE3, AZU1, LTF* etc.). Most

of stage 4 samples besides two of bone marrow origin clustered together (Cluster 3, Figure 24A) and expressed high levels of *DUSP4, N4BP3*, and *TNFSF9*. Besides this they shared gene expressions with both Cluster 1 and Cluster 4, however both patterns were expressed on lower levels. Finally, Cluster 4 included only stage 5 samples and expressed mainly genes typical for mature NKs (e.g. *KLRF1, GZMA, GZMB, GZMH, GZMM, IFNG, KIR3DX1, PRF1, TBX21* etc.).

Pairwise correlation of gene expression between stages was performed based on the top 100 protein-coding and non-protein coding genes, respectively (Figures 25). Similar to the analysis of the complete transcriptome, both for protein-coding and non-protein coding genes the highest correlation was detected between stages following each other in development. The strongest correlations were observed between stages 1 and 2, and between 4 and 5. Correlation coefficients amounted to 0.744 and 0.69 respectively when protein-coding genes were analyzed, and were even higher (0.83 and 0.74 respectively) in non-protein coding genes. Also positive, but weaker correlation was observed between stages 2 and 3, and between stages 3 and 4. However, in that case a clear difference was observed between protein-coding and non-protein coding gene. For non-protein coding genes the correlation between stages 2 and 3 and between stages 3 and 4 was beyond the threshold (correlation coefficients 0.285 and 0.402 respectively). For protein-coding genes in contrast the correlation was relatively strong between stages 2 and 3 (correlation coefficient 0.62) and dramatically weaker between stages 3 and 4 (correlation coefficient 0.046).

Notably, rather strong negative correlations were detected between stages 1 and 5, as well as between 2 and 5 (-0.4 and -0.55) when non-protein coding genes were analyzed, which was not the case upon analysis based on protein-coding genes. In the latter, however, the strongest negative correlation was detected between stages 3 and 5 (-0.43), which is in accord with the analogical analysis based on the complete transcriptome.

Higher correlations for expression of non-protein coding genes in comparison to protein-coding genes (Figure 26) might be due to smaller expression changes of non-protein coding genes during development. Further, as non-protein coding genes are thought to be involved in the expression regulation of (protein-coding) genes, their expression would be expected to be more tightly regulated during either very early or during very late developmental stages. If an outstanding position of stage 3 is due to presence of ILC3s, one can suppose, that differences between stage 3 NK cells and ILC3s are more pronounced in expression of protein-coding, than in non-protein coding genes.

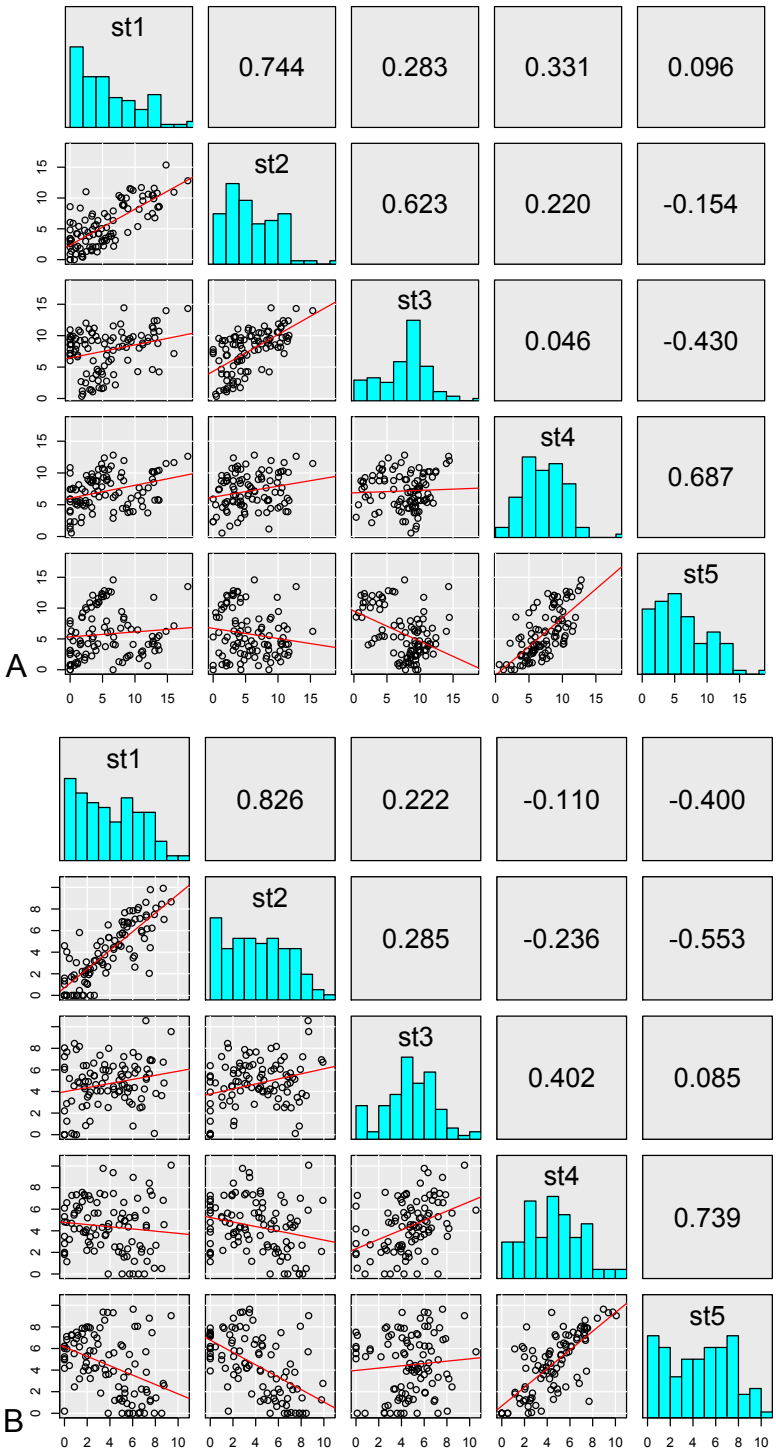

Figure 25 Pairwise comparison of gene expression levels within *ex vivo* NK samples based on top 100 protein coding genes (A) and top 100 non-protein coding genes (B)
Protein-coding genes were selected from the total number of 14177 and non-protein coding – from the total number of 6613, LoD=1.

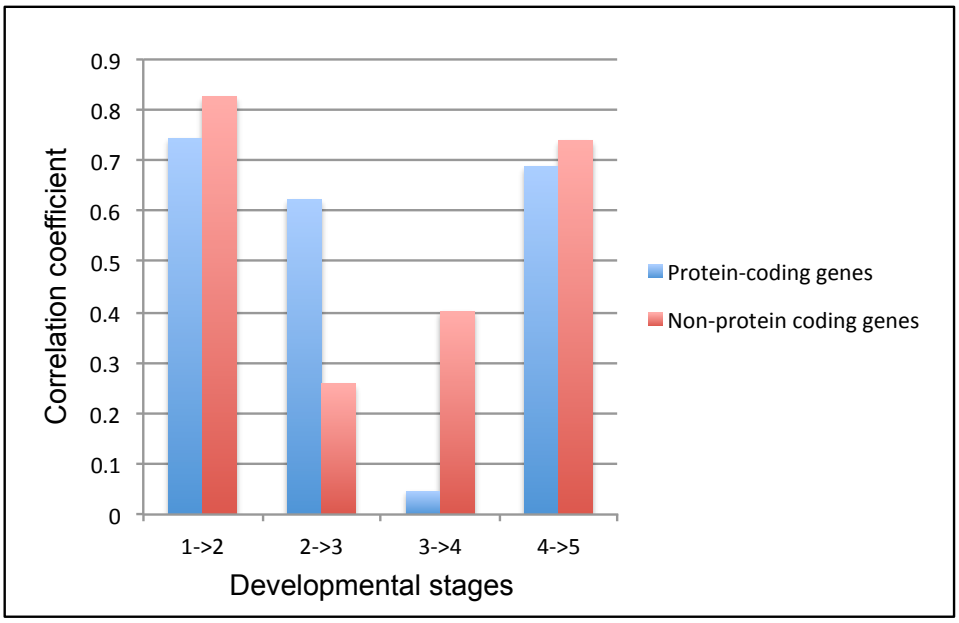

Figure 26 Correlation of gene expression levels between consequent developmental stages based on expression of protein-coding and non-protein coding genes

3.6 Regulatory potential of non-protein coding RNAs

As many of the non-protein coding genes among the studied genes were specifically expressed in certain lymphocyte subgroups and among certain developmental stages, we supposed that they might have regulatory potential for lymphocyte development and function.

3.6.1. Correlation between expression levels of protein-coding and non-protein coding genes during NK development

As described in chapter 3.5 the expression of both protein-coding and non-protein coding genes changes during NK cell development from stage 1 to stage 5. To investigate further relations between these two gene groups, a correlation analysis of their expression levels was performed. The sample set was the same as in chapter 3.5 and included 56 *ex vivo* NK samples from stages 1 to 5; the gene set included the top 100 differentially expressed protein-coding genes and the top 100 non-protein coding genes, described for this sample set above. Only correlations between protein-coding and non-protein coding genes were taken into account. Correlations with the value of coefficient >0.8 or <-0.8 and $p<0.05$ were considered to be significantly strong and are summarized in the Appendix 14.

Out of the top 100 non-protein coding genes, 26 showed a significant strong positive (Figure 27) or negative (Figure 28) correlation to at least one of the top 100 protein-coding genes; 42 out of the top 100 protein-coding genes were involved in these correlations (Appendix 14).

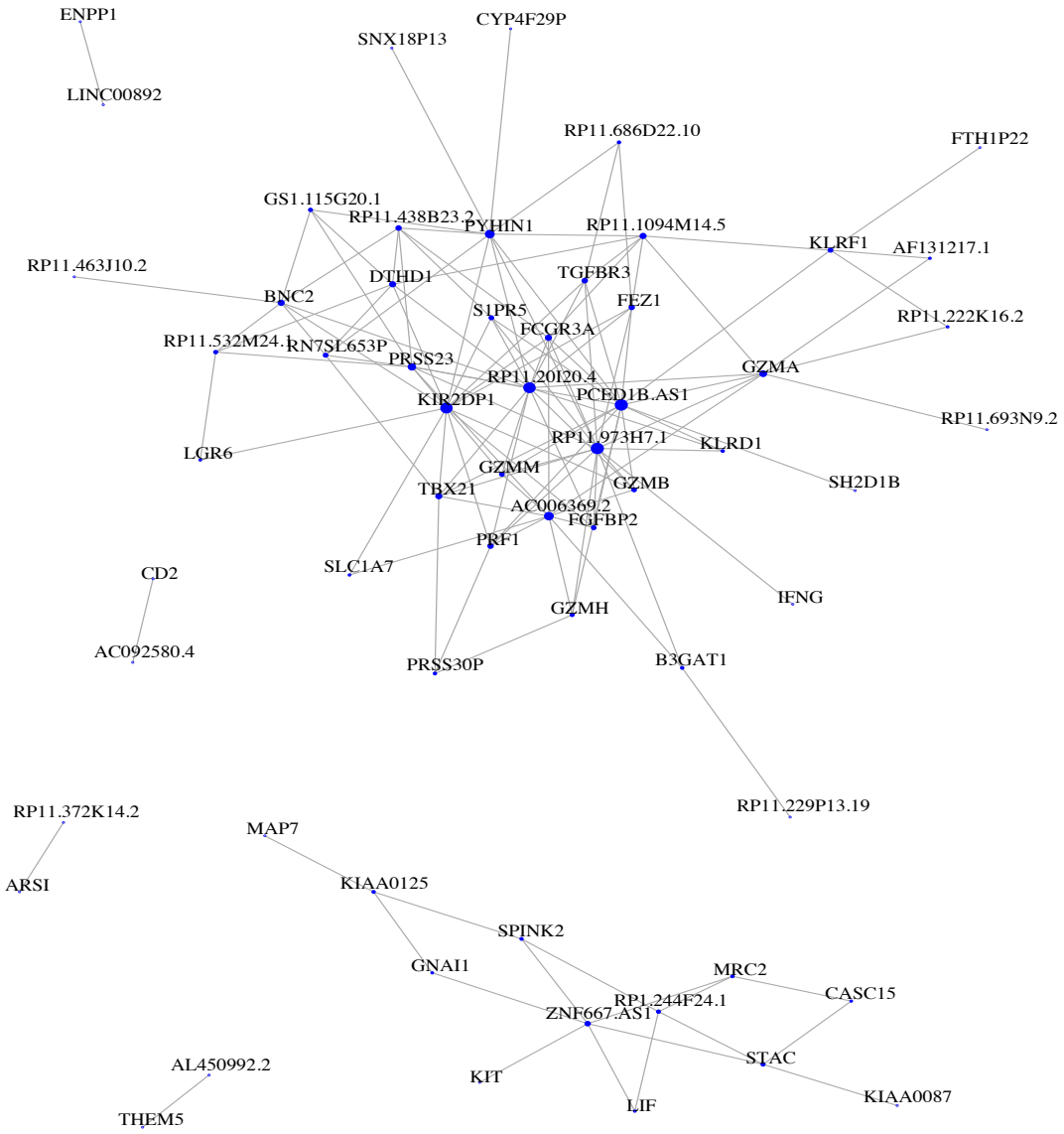

Figure 27 Correlation network of gene expression between protein-coding and non-protein coding genes within *ex vivo* NK samples
Only correlations with Spearman's coefficient>0.8, p<0.05 are shown.

Two genes among these 26 genes, *RP11-973H7.1* and *RP11-20I20.4*, are expressed specifically in stages 4 and 5 and had the highest number of correlations: 17 and 20 respectively. Both of them correlated to following genes: *S1PR5* (correlation coefficients 0.85 and 0.91 respectively), *TGFBR3* (0.85 for both), *GZMA* (0.84 and 0.8), *GZMB* (0.92 for both), *GZMM* (0.82 for both), *FGFBP2* (0.86 and 0.82 respectively), *KLRD1* (0.82 and 0.81), *TBX21* (0.82 and 0.83), *PYHIN* (0.87 and 0.86), *PDZD4* (0.86 and 0.84), *PRSS23* (0.81 and 0.8), *FCGR3A* (0.87 and 0.84), *FEZ1* (0.88 for both), and *PRF1* (correlation

coefficients 0.88 and 0.87 respectively). Consistently, all of these protein-coding genes were also expressed by samples from stage 4 and/or 5.

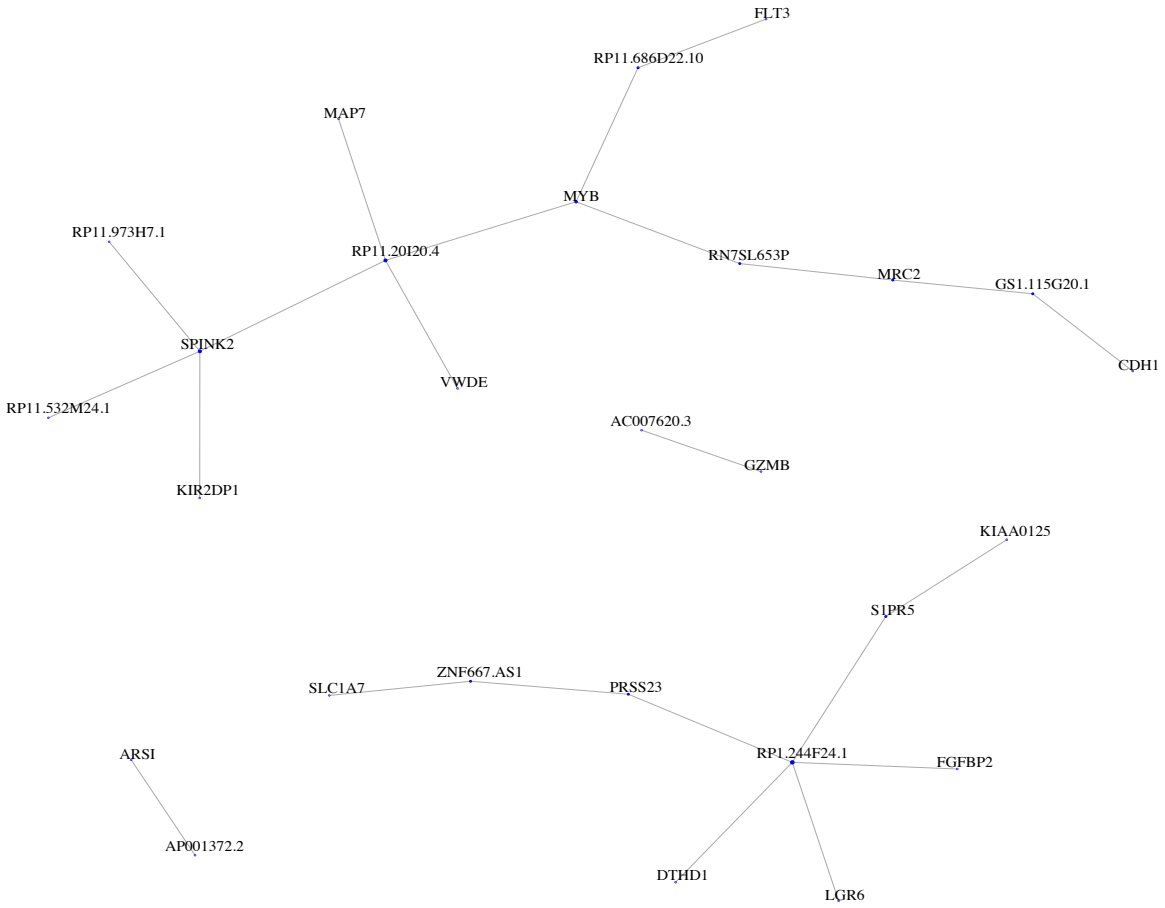

Figure 28 Correlation network of gene expression between protein-coding and non-protein coding genes within *ex vivo* NK samples
Only correlations with Spearman's coefficient<-0.8, p<0.05 are shown.

The expression of 3 more genes typical for mature NK cells, *IFNG, GZMH* and *B3GAT1*, correlated with the expression of *RP11-973H7.1* (with coefficients 0.84, 0.8 and 0.81 respectively), but not with *RP11-20I20.4*. At the same time, expression of *RP11-20I20.4*, but not of *RP11-973H7.1*, strongly correlated with the expression of *DTHD1* and *BNC2* (coefficients 0.8 and 0.82 respectively). In addition, expression of *RP11-20I20.4* was negatively correlated with the expression of *SPINK2* (coefficient -0.82), *MAP7* (-0.83), *MYB* (0.82) and *VWDE* (-0.8). Notably, while *MAP7, MYB* and *VWDE* were expressed by stages 1 to 3, *SPINK2* gene was expressed mainly by stage 4 NKs.

RP11-1094M14.5, the gene with the third highest number of correlations, was also expressed within stage 4 and 5 samples, but also in samples from stage 3. Strong positive correlations were observed for genes *TGFBR3* (correlation coefficient 0.8), *DTHD1* (0.84),

PYHIN1 (0.9), *FCGR3A* (0.82), *KLRF1* (0.83), *GZMA* (0.85), *BNC2* (0.8), and *FEZ1* (0.8); all these genes were also highly expressed by mature NK cells of stages 4 and 5. The same was the case for another non-protein coding gene expressed within stages 4 and 5, *RP3-467K16.4*. Its correlation also correlated to those of mature NK cell markers: *GZMB* (coefficient 0.83), *GZMM* (0.82), *IFNG* (0.82), *PDZD4* (0.81), *PRF1* (0.8).

Three more lncRNA genes, *RP11-330A16.1* (expressed mainly in stage 4 samples), *RP11-563D10.1* (expressed in stage 3 samples from tonsils and cord blood) and *RP1-206D15.6* (expressed in stage 3 samples, but also among some samples from stage 4) possessed four positive correlations each. Expression levels of *RP11-330A16.1* correlated with expression of *ENPP1* (correlation coefficient 0.82), *CSPG4* (0.83), *IL7R* (0.82), and *LTB* (0.82). *RP11-563D10.1* expression correlated expression levels of *ENPP1* (correlation coefficient 0.82), *LIF* (0.84), *RORC* (0.82), and *KLRF2* (0.8). Expression of *RP1-206D15.6* correlated with expression levels of *LIF* (correlation coefficient 0.83), *IL1R1* (0.8), *LPAR1* (0.83), and *MYO7A* (0.83). Expression of *AC005083.1,* a gene encoding processed transcript and expressed also mainly in stage 3 samples, correlated positively to expression of *IL1R1* (correlation coefficient 0.82) and *MYO7A* (0.81) and negatively to *S1PR5* (-0.81).

KIAA0125 (expressed in stages 1 to 4, but not in any stage 5 sample) and *RP11-686D22.10* (expressed mainly in stages 4 and 5, but also stage 3 samples from tonsils) had four correlation connections each as well. The expression of *KIAA0125* positively correlated with that of *SPINK2* (correlation coefficient 0.84) and *MAP7* (0,81), and correlated negatively with two genes expressed within stages 4 and 5, *S1PR5* and *GZMB* (-0.8 for both). The expression of *RP11-686D22.10* positively correlated to that of *TGFBR3, PYHIN1* and *FEZ1* (correlation coefficients 0.81, 0.86 and 0.82 respectively) and negatively with that of *FLT3* (-0.82).

Genes *RP6-91H8.3, RP11-796E2.4* and *RP11-326C3.2* were expressed in stages 3 and 4, while the latter two of them were also expressed in stage 5, but at lower levels. The expression of all three genes correlated significantly with that of *LTB* (correlation coefficient 0.8, 0.81 and 0.82 respectively), while *RP6-91H8.3* expression also correlated with that of *MYO7A* (0.83), the expression of *RP11-796E2.4* with that of *XCL1* (0.85), and expression of *RP11-326C3.2* correlated with that of *DLL1* (0.84). The expression of *RP11-222K16.2*, a gene expressed strictly in stage 4 and 5, correlated with that of typical mature NK marker genes, *KLRF1* and *GZMA* (correlation coefficient 0.84 and 0.86 respectively).

Finally, the other eight non-protein coding genes correlated to only one gene each. Two genes, *LINC00892* and *RP11-520A21.1*, were expressed by samples from stages 3 and 4. While *LINC00892* expression correlated with that of *ENPP1* (correlation coefficient 0.84), *RP11-520A21.1* expression correlated with *IL4I1* (0.86). One more gene, *LINC00309*, which was also expressed at stage 3 and at lower level in stages 4 and 5, correlated as well with *IL4I1* expression (0.84).

From another three genes expressed mainly at stages 4 and 5, but also partly at stage 3, the expression of *FTH1P22* correlated significantly with that of *KLRF1* (correlation coefficient 0.82), *RP11.121A8.1* correlated with *PRF1* (0.83), and *RP11.463J10.2* correlated with *BNC2* (0.84). In addition, *RP11-693N9.2* was expressed in stages 3 to 5 and correlated with *PRF1* (correlation coefficient 0.83).

The last identified non-protein coding gene, *SCART1*, was expressed ubiquitously within all analyzed samples and its expression correlated with that of *ENPP1* (correlation coefficient 0.82).

3.6.2. Cis-regulation potential

As many non-protein coding genes regulate the expression of neighboring genes, the cis-regulatory potential of the top 100 non-protein coding genes was studied. First of all, their neighboring protein-coding genes were identified based on the genes' localization (Table 13). To narrow down the number of potential target genes, only genes localized within up to 5% of the chromosome length from the given non-protein coding genes were considered. The result of this analysis is shown in Table 13. For the sake of clarity, the further analysis was restricted to 28 non-protein coding genes (marked green in Table 13) that are located in the neighborhood of protein-coding genes and have a known function in the immune system, in particularly play a role in the function and development of NK, T, and B lymphocytes.

Table 13 Characteristic of top 100 non-protein coding genes differentially expressed within ex vivo NK samples

GeneID	Localization	Gene type	Neighboring genes of interest
AC002454.1	7q21.2	antisense	CDK-6, GATAD1, FAM133B
AC004791.2	19q13.12	lincRNAS	CYP4F2, OR10H4
AC005083.1	7p21.1	processed transcript	ITGB8, MACC1
AC007278.2	2q14.3	sense intronic	IL18RAP, IL18A1, IL1RL1
AC007319.1	2q32.1	antisense	CALCRL, TFP1
AC010761.14	17q11.2	antisense	PRL23A, RAB34, TLCD1
AC010886.2	2q35	processed pseudogene	DIRC3, TNS1
AC017104.6	2q37.1	processed transcript	B3GNT7, NCL

AC020571.3	2q32.3	antisense	STK17B, HGNC
AC022182.1	8q12.2	lincRNA	CHD7, CLVS1
AC092580.4	2p25.1	lincRNA	ID2
AC092667.2	2q11.2	antisense	NMS, LOMRF2, CHST10, AFF3
AF131217.1	21q21.3	lincRNA	BCL11B
AL356585.1	13	miRNA	FAM230C, AL356585.2
AL450992.2	1q21.3	antisense	C2CD4D, THEM5, RORC, S100A10
AMZ2P2	6q25.3	processed pseudogene	SYTL3, EZR
AOX2P	2q33.1	pseudogene	AOX1, BUW1, NIF3L1, CLK1, PPIL3
AP001171.1	21q21.1	lincRNA	-
B3GALT5-AS1	21q22.2	antisense	SH3BGR, IGSF5, HMGN1, B3GALT5
CASC15	6q22.3	linc	SRY
EDDM3CP	2q21.3	processed pseudogene	MGAT5
EGFEM1P	3q26.2	pseudogene	GOLIM4
FCGR2C	1q23.3	pseudogene	FCGR2C, FCRLA, FCRLB, DUSP12, FCGR2A, SDHC
FTH1P22	1p13.1	processed pseudogene	CD2, CD101, CD58, IGSF3, PTGFRN
GATA2-AS1	3q21.3	antisense	GATA2, RAB7A, ACAD9, DNSJB, RPN1
HCP5	6p21.33	sense overlapping	TCF19, AIF1, LTB, LTA, LST1, MICB, TNF, NCR3, BAG6, NFKBIL1, MICA
HLA-DQB1-AS1	6p21.32	antisense	HLA-DQB1A, BTNL2
HSPE1P18	11q23.3	processed pseudogene	IL10RA, CD3E, CD3G, CD3D, KMT2A, TMPRSS4, AMICA1, CXADR
IGF2BP2-AS1	3q27.2	antisense	IGF2BP2, SENP2, TRA2B
KIAA0087	7p15.2	lincRNA	SNX10, SKAP2
KIAA0125	14q32.33	lincRNA	IG locus
LINC00299	2p25.1	lincRNA	ID2
LINC00309	2p14	lincRNA	PELI1, LGALSL1
LINC00877	3q13	lincRNA	PROK2, EIF4E3, GPR227
LINC00880	3q35.21	lincRNA	TIPARP, LEKR1, CCNL1
LINC00892	Xq26.3	lincRNA	CD40LG, HTATSF1,
LINC01132	1q42.3	lincRNA	IRF2BP2
LINC01135	1p32.1	lincRNA	JUN, OMA1, MYSM1
LINC01268	6q21	lincRNA	MARCK, HDAC2
LPAL2	6q25.3	unprocessed pseudogene	SLC22A3
MAGI2-AS3	7q21.11	processed transcript	MAGI2
NAPSB	19q13.33	pseudogene	NR1H2, KCNC3, POLD1, SPIB, EMC10, NAPSA
PLA2G4E-AS1	15q15.1	antisense	PLA2G4E, EHD4
PLA2G4E-AS1	15q15.1	antisense	EHD4
RP1-131F15.2	6q23.2	processed pseudogene	ENPP1
RP1-18D14.7	1p33	antisense	TAL1, PDZK1IP1, STIL
RP1-206D15.6	1q24.2	lincRNA	SLC19A2, BLZF1, SELL, NME7, SLC19A2
RP11-104L21.2	1q24.2	sense intronic	CD3Z
RP11-104L21.3	1q24.2	lincRNA	CD3Z
RP11-1094M14.5	17q12	sense intronic	TAF5L, SLFN12L

RP11-121A8.1	7p14.1	lincRNA	TRGV9, TRGC2, TRGV8, TRGVA, TSGA10
RP11-158I9.5	11q23.3	antisense	CXCR5, TRAPPC4, FOXR1, BCL9L
RP11-15B24.5	9q21.32	lincRNA	SPATA31D1
RP11-20I20.4	4p16.3	antisense	SPON2, FGFRL1
RP11-212I21.4	16q	lincRNA	LPCAT2
RP11-21J7.1	1q31.2	lincRNA	CDC73, TROVE2, B3GALT2
RP11-222K16.2	3p24.1	lincRNA	EOMES
RP11-252C15.1	8p23.1	lincRNA	KIAA1456, LONRF1, DLC1
RP11-277P12.20	12p13.2	antisense	KLRC1, KLRC2, KLRC3, KLRC4, KLRK1, KLRD1
RP11-305O6.3	12q14.3	sense intronic	MSRB3, LEMD3
RP11-326C3.2	11p15.5	antisense	AP2A2, IFITM2, ATHL1, NLRP6, RASSF7, IFITM3, IRF7, IFITM5, ANO9
RP11-330A16.1	6p23	lincRNA	CD83
RP11-342I1.2	4q31.1	antisense	RAB33B, SETD7
RP11-354E11.2	10p12.31	antisense	MALRD1, PLXDC2
RP11-366L20.2	12q14.3	antisense	HMGA2,
RP11-389C8.2	5q13.2	sense overlapping	ZNF366, PTCD2
RP11-424C20.2	12p12.2	processed pseudogene	SLCO1C1, PDE3A
RP11-44N11.3	8q24.13	lincRNA	DERL1, ZHX2
RP11-456H18.2	6p21.1	lincRNA	HIVEP1
RP11-463J10.2	14q22.1	antisense	GNG2, NID2
RP11-475O6.1	1p31.1	lincRNA	TTLL7
RP11-47L3.1	17q12	lincRNA	SLFN12L, SLFN5, SLFN11, SLFN12, SLFN13; SLC35G3, NLE1
RP11-520A21.1	3p22.1	lincRNA	ZNF619, ZNF620, ZNF621
RP11-563D10.1	1q31.3	lincRNA	-
RP11-567M16.1	18q23	processed transcript	CTDP1, NFATC1
RP11-624C23.1	8p22	antisense	NEFM, ADAM28
RP11-678G14.3	18p12	lincRNA	ZNF493, ZNF429, ZNF100
RP11-686D22.10	17q12	pseudogene	SLFN12
RP11-693J15.5	12q22	lincRNA	BTG1, CLLU1, CLUU1OS
RP11-693N9.2	11q22.3	unprocessed pseudogene	CASP12, CASP5, CASP4, CASP1, CARD16 CARD17, CARD18
RP11-6F2.5	3q25.31	lincRNA	LEKR1, CCNL1
RP11-705C15.5	12p13.31	processed transcript	KLRB1, KLRF1, KLRF2, CLECL1C, CLEC2D, CD69
RP11-757F18.5	3q13.2	antisense	GCSAM, SLC9C1
RP11-794P6.1	11q23.1	sense intronic	POU2AF1, COLCA1, COLCA2, BTG4
RP11-796E2.4	12q21.33	antisense	BTG1
RP11-973H7.1	18p11.12	lincRNA	PSMG2, PTPN2
RP11-98G13.1	1q31.2	lincRNA	B3GALT2, KCNT2
RP13-786C16.1	11p13	antisense	FBOXO3, CD59, LMO2
RP3-467K16.4	1p36.21	lincRNA	EFHD, DNAJC16, CTRC, CASP9, CELA2A , CELA2B
RP4-738P11.3	1q24.2	processed pseudogene	XCL2, XCL1
RP5-1028K7.2	17q21.2	lincRNA	IGFBP4, CCR7

RP6-91H8.3	14q24.2	lincRNA	PCNX
RPL23AP22	1q31.2	processed pseudogene	-
RPS20P2	16p13.3	processed pseudogene	UBE21, BAIAP3
SCART1	10q26.3	unitary pseudogene	CYP2E1, SPRN
SRGAP3-AS2	3p25.3	antisense	SRGAP3
THRA1/BTR	17q21.32	antisense	SKAP1, NFE2L1, SNX11, CBX1
TVP23CP1	7q21.13	processed pseudogene	CDK14
XXbac-BPG181B23.6	6p21.33	unprocessed pseudogene	TCF19, AIF1, LTB, LTA, LST1, MICB, TNF, NCR3, BAG6, NFKBIL1, MICA

Then, the correlation between the expression of non-protein coding genes and the corresponding potential targets was calculated in the specific lymphocyte populations. Spearman's correlation coefficients were calculated and compared for following sample groups: all available samples (n=96), *ex vivo* NK samples from stages 1 to 5 (n=56), stage 3 *ex vivo* NK samples (n=11), stage 4 *ex vivo* NK samples (n=9), and stage 5 *ex vivo* NK samples (n=30). Only correlation with values of Spearman's coefficient >0.5 or <-0.5, $p<0.05$ were considered as significant and taken into account; correlation with values of Spearman's coefficient >0.7 or <-0.7 were considered as strong and are summarized in the Appendix 15. Expression of some non-protein coding genes also correlated to those of protein-coding genes that are not closely localized; most important of such correlations are mentioned in chapters 3.6.1 and 3.7.4, so they will not be discussed here.

Even though it could be possible that correlation is caused by an opened or closed chromatin state, the close localization of genes does not necessarily lead to expression correlation. For example, there was no significant correlation detected in any of the sample groups (neither positive nor negative) between *LINC011332* and *IRF2BP2*, which are localized about 66 kb apart. Also the expression of *RP1-206D15.6* and *SELL*, of *RP11-693J15.5* and *BTG1*, and of *AOX2P* and *NIF3L1* (localized 124 kb, 289 kb, and 140 kb apart, respectively), is not significantly correlated in any sample group. At the same time, expression of another non-protein coding gene, *RP11-796E2.4*, which is localized less than 1kb from the *BTG1* gene, correlated with the *BTG1* expression in all sample groups and in the *ex vivo* NK sample groups (r=0.57 and 0.56 respectively). *RP5-1028K7.2* is localized less than 26 kb away from the CCR7 gene and less than 20 kb away from the *IGFBP4* gene. However, no correlation was found between the *RP5-1028K7.2* and the *CCR7* expression, while the expression of *RP5-1028K7.2* and *IGFBP4* correlated within stage 3 samples (r=0.79) and weaker among all samples (r=0.52).

When such RNA-seq data are analyzed, one should also take into account the expression levels of the given genes. For example, expression levels of *RP11-158I9.5* and *CXCR5* strongly correlate among all samples (0.8), among *ex vivo* NK samples (0.6), among stage 3 (0.9) and stage 4 (0.7). However, this effect seems to be achieved due to low expression levels of both genes in all these samples, besides B-lymphocytes. This was also the case for *B3GALT5-AS1* (an antisense RNA of *B3GALT5*). Its expression correlated to *B3GALT5* within all sample group (r=0.6), *ex vivo* NK cells (0.64), and at stage 3 (0.9), however, in all samples except for stage 3 the expression of both these genes did not exceed 10 fpkm. Notably, expression of another closely localized gene, *IGSF5*, correlated to *B3GALT5-AS1* (r=0.91) only within stage 3, where *B3GALT5-AS1* expression was relatively high. Between other closely located genes, *HMGN1* and *SH3BGR*, and *B3GALT5-AS1* no significant correlation was detected.

In some cases among several of the protein-coding genes, which are localized in the neighborhood of a given non-protein coding gene, expression of only a certain one correlates. For example, genes of four interferon induced transmembrane proteins (*IFITM1, -2, -3* and *-5*) are localized within 30 kb distance of *RP11-326C3.2; SIGIRR* is localized at the same DNA strand in less that 50 kb. However, expression of only one of them, *IFITIM5*, correlated with *RP11-326C3.2* expression in both group of all samples and *ex vivo* NK cells (0.7 and 0.69 respectively), even though they are localized on opposite DNA strands. Expression of *IFITM1, IFITM2*, and *SIGIRR* correlates instead with *RP11-326C3.2* expression within stage 3 (Spearman's coefficients 0.68, 0.68 and 0.72 respectively), while between expression levels of *RP11-326C3.2* and *IFITM3* there was no significant correlation at all.

The four genes encoding cell surface molecules *CD2, CD58, IGSF3*, and *CD101* (*IGSF2*) are localized in the neighborhood of the *FTH1P22* non-protein coding gene. Among them, only the expression of *CD2* positively correlated with that of *FTH1P22* within all samples group (0.55), *ex vivo* NK group (0.6), among stage 3 samples (0.68), and among stage 5 samples (0.63).

Another example is two T-cell receptor segments, five of which (*TRGV8, TRGVA, TRGC2, TRGV9*, and *TRAG10*) are localized within 70 kb of *RP11-121A8.1*, which was expressed among NK cells from stage 3 to stage 5 with highest expression at stage 5. Among them *TRGV8* and *TRGVA* were expressed in all NK, as well as in T cell samples at low level, however, their expression levels were strongly correlated among all samples groups (correlation coefficients from 0.64 to 0.88) for *TRGV8* and among all groups

besides stage 5 NK cells (0.6 to 0.81) for *TRGVA, TRGC2* and *TRGV9* genes were both expressed within stages 3 to 5 and their expression correlated with that of *RP11-121A8.1* within all sample group (0.75 and 0.8 respectively), *ex vivo* NK samples (0.72 and 0.65 respectively) and weaker within stage 5 samples (0.58 and 0.56 respectively), while for *TRGC2* the a correlation coefficient of 0,65 was also observed within stage 4 samples. The last examined gene from this locus, *TRGA10*, was relatively highly expressed within stage 5 NK and T-cell samples, however, no correlation of the *TRGA10* expression with that of *RP11-121A8.1* was observed in any of studied sample groups.

The same was the case for *HSPE1P18*: among three genes located close to it, only *CD3E* expression correlated to *HSPE1P18* among stage 3 samples (r=0.8), while for *CD3D* and *IL10RA* no significant correlation was detected.

Vice versa, when two non-protein coding genes localized close to the same protein-coding gene, it is not necessary that their expression correlate in the same way. For example, *LINC00299* and *AC092580.4* are localized in the neighborhood of the *ID2* gene, which encodes an essential factor of ILC and NK cell development (Yokota et al., 1999; Cherrier et al., 2012). *LINC00299* is an NK- and ILC3-specific lncRNA, localized about 270 kb from the ID2 gene; its expression correlated with *ID2* expression levels, but only when all samples were analyzed (r=0.68), and not among *ex vivo* NK developmental stages or within specific stages. *AC092580.4* is localized 207 kb from *LINC00299* and in 870 kb distance from the *ID2* gene, but in contrast to *LINC00299* the expression of *AC092580.4* correlated with that of ID2 not only in all samples (0.75), but also among *ex vivo* NK cells (0.59) and – in the strongest manner – among stage 4 NK cells (0.86).

Two other analyzed non-protein coding genes localized close to each other are *XXbac-BPG181B23.6* and *HCP5*. They are located in the HLA complex, a gene-rich region of chromosome 6, in less than 30 kb from each other. The following protein-coding genes were localized within 200kb from *HCP5* and *XXbac-BPG181B23.6: AIF1, BAG6, HLA-B, HLA-C, LTS1, LTA, LTB, MICA, MICB, NCR3, NFKBIL1*, and *TNF*; and it seems that in most cases correlation between their expression levels did not depend on whether genes were localized on the same or on opposite DNA strands. In particular, expression of *LTB* and *NFKBIL1* (both localized at the same strand with HCP5) correlated with that of both *HCP5* and *XXbac-BPG181B23.6* within the same patterns (only in stages 3 and 4 groups, but not in others) and at comparable levels. Expression levels of *LTB* and *HCP5* correlated with Spearman's coefficients 0.89 and 0.68 respectively, *LTB* and *XXbac-BPG181B23.6* expression levels correlated with coefficients of 0.75 and 0.73; expression

of *NFKBIL1* correlated with that of *HCP5* with coefficients of 0.75 and 0.65, expression of *NFKBIL1* correlated to those of *XXbac-BPG181B23.6* with coefficients of 0,7 and 0, 67. Also expression of *BAG6* correlated with both these non-protein coding genes within the stage 5 sample group (0.67 for *HPC5* and 0.77 for *XXbac-BPG181B23.6*), but not in any other group; *LTA* correlated with both of them within the stage 3 group (0.95 and 0.84 respectively). *NCR3* expression (localized on the same strand as *XXbac-BPG181B23.6*) correlated with both non-protein coding genes among all samples group (0.55 and 0.5 respectively), among *ex vivo* NK sample group (0.67 and 0.7), and among stage 3 NK samples (0.97 and 0.82). NCR3 also weakly correlated with *XXbac-BPG181B23.6* expression among stage 5 (0.5), which was not the case for *HCP5*, while correlation patterns of *AIF1* and *LST1* (both localized on the same strand as *HCP5*) differed between *HCP5* and *XXbac-BPG181B23.6.* While no significant correlation was detected between expression of *AIF1* and *XXbac-BPG181B23.6* as well as between *XXbac-BPG181B23.6* and *LST1*, LST1 correlated positively with *HCP5* among stage 3 samples, and *AIF1* expression correlated negatively with *HCP5* among all sample and *ex vivo* NK sample groups (-0.58 and -0.57 respectively).

However, The expression of both *RP11-109M14.5* and *RP11-686D22.10,* localized in the same region as *SLFN12L*, correlate with the *SLFN12L* expression in the *ex vivo* NK sample group (r=0.94 and 0.88 respectively), among all samples (0.95 and 0.94 respectively), and among stage 5 samples (0.88 and 0.58 respectively). Besides this, *RP11-109M14.5* expression also strongly correlated with that of *SLFN12L* within stage 4 samples (0.91), but not among stage 3, while *RP11-686D22.10* expression vise versa correlated with that of *SLFN12L* among stage 3 samples (0.92), but not among stage 4 samples. Notably, even though *RP11-109M14.5* is a sense intronic RNA-encoding gene located in the TAF5L gene, their expression levels were not correlated in any of studied groups.

A further very interesting correlation with regulatory potential in NK cell development deserves to be described separately: between *RP11-222K16.2* and *EOMES. RP11-222K16.2* is localized on the opposite DNA strand of *EOMES*, in about 2 kb distance from it. *EOMES* is an important transcription factor in NK cell development, and correlation between its expression levels to expression level of *RP11-222K16.2* was observed among all samples and *ex vivo* NK cells (0.87 and 0.8 respectively), but not among stage 3, stage 4 or stage 5 NK groups. These data were consistent with differences in gene expression among developmental stages (chapter 3.3.1) and suggests a possible

positive role of *RP11-222K16.2* in the regulation of *EOMES* expression in early NK cell developmental stages.

The expression of *RP11-567M16.1* - localized on the same DNA strand and in about 131 kb distance from *NFATC1* - strongly correlated with that of *NFATC1* of stage 3 samples (0.98), while correlation within all samples and *ex vivo* NK cells was weaker (0.51 and 0.54 respectively) and completely absent within stages 4 and 5, suggesting that a potential regulation could take place at developmental stage 3, but not later.

Also a rather strong correlation within the stage 3 sample group was detected between expression of *RP4-738P11.3* and two cytokine genes, *XCL1* and *XCL2* (0.75 and 0.7 respectively). Their expressions correlated within all samples (0.84 and 0.87) and within the *ex vivo* NK group (0.71 and 0.79), but not within the stage 4 and 5 groups.

Among other genes of interest was *AC007278.2*, a sense intronic RNA of IL18RAP, and the expression of both genes expectedly correlate, but to different extent in different groups: the correlation coefficient was 0.9 among all samples, 0.8 among *ex vivo* NK cells, 0.6 among stage 3, 0.67 among stage 5, but negatively among stage 4 (-0.6; although not significantly, p=0.07). On the same DNA strand in less than 70 kb from *AC007278.2* are the *IL18R1* and *IL1RL1* genes localized. While there is no significant correlation between *AC007278.2* and *IL1RL1* expression in any of the analyzed groups, *IL18R1* expression correlated with *AC007278.2* in the group of all samples and even stronger in stage 3 NK sample group (0,64 and 0,86 respectively), but not in any other.

AL450992.2 is localized about 20 kb from the *RORC* gene (a highly important marker of ILC3 cells, which is normally not expressed in NK lymphocytes) and 67 kb from the *S100A10* gene (both are localized on the opposite DNA strand of *AL450992.2*). While the expression of *AL450992.2* and *RORC* slightly correlated in both all sample group and *ex vivo* NK cells (0.56 and 0.64 respectively), there was no significant correlation between *AL450992.2* and *S100A10* in either of these groups. Among stage 3 samples, correlation between *AL450992.2* and *RORC* was even higher (0.82), while *AL450992.2* and *S100A10* were strongly negatively correlated in stage 4 samples (-0.76).

Two immunologically relevant genes are localized in the neighborhood of *LINC00892*: *CD40LG* localized on the same DNA strand in less than 6 kb, and *HTATSF1*, which is located on the opposite strand at 77 kb distance of *LICN00892*. While no significant correlation in any groups was detected between *LINC00892* and *HTATSF1*, expression of *LINC00892* and *CD40LG* correlated within stage 5 group (r=0.57) and even stronger within stage 3 group (0.88).

Expression of lncRNA *RP11-330A16.1* and that of the *CD83* gene, localized in 44 kb distance on the opposite DNA strand, were negatively correlated among stage 4 samples (-0.67), while no other correlations were observed within other groups.

Among three genes localized close to *GATA2-AS1*, only *GATA2* expression itself correlated with that of *GATA2-AS1* among all samples (0.64) and *ex vivo* NK samples (0.91). Expression of the other two genes, *ACAD9* (localized on the same DNA strand in 315 kb distance) and *RAB7A* (localized also on the same DNA strand in 195 kb distance) did not correlate with that of *GATA2-AS1* in any of those sample groups. However, the expression of *DNAJB8*, (localized on the opposite DNA strand of *GATA2-AS1* in about 20 kb distance) correlated with that of *GATA-AS1* among *ex vivo* NK samples (0.66) and among stage 5 samples (0.55).

The *RP11-277P12.20* non-protein coding gene is localized in the locus of NK cell receptors, including *KLRC1, -C2, -C3, -C4, -K1*, and *-D1*. Its expression correlated with that of all mentioned KLR genes among all samples (correlation coefficients from 0.69 to 0.92), and among stage 3 samples it correlated with that of all *KLR* genes except for *KLRD1* (0.72 to 0.97). However, among *ex vivo* NK samples its expression strongly correlated with that of *KLRC4, KLRK1, KLRC2,* and *KLRC3* (coefficients 0.9, 0.87, 0.77, 0.84 respectively), and weaker with that of *KLRD1* (0.53), but not with that of *KLRC1*. Among stage 4 samples, the expression of *RP11-277P12.20* was strongly correlated with that of *KLRC4* and *KLRC1* (0.7 and 0.9 respectively), and among stage 5 samples with that of *KLRC3, KLRC4*, and *KLRK1* (0.6, 0.83, 0.66 respectively).

One more non-protein coding genes localized in the *KLR* locus was *RP11-705C15.5,* while the distance to above-mentioned *KLR* genes was more than 400 kb, *KLRB1, -F1* were localized in less than 100 kb distance from it, and *KLRF2* in about 150 kb. The expression of both *KLRB1* (localized on the opposite DNA strand of *RP11-705C15.5*) and *KLRF1* (localized on the same DNA strand) correlated with that of *RP11-705C15.5* among all sample group and *ex vivo* NK group (0.72 and 0.53 respectively for *KLRB1*, 0.7 and 0.71 for *KLRF1*), while the expression of *KLRF1* also correlated strongly with that of *RP11-705C15.5* among stage 5 with a coefficient of 0.85. At the same time, no significant correlation between *KLRF2* and *RP11-705C15.5* expression was detected. Another immunologically relevant gene in that locus was *CD69* (expressed by NK cells upon activation), localized less than 30 kb away from *RP11-705C15.5* on the opposite DNA strand, but no significant correlation was detected.

3.7. Transcriptome analysis of NK lymphocytes at the single cell level

Sequencing libraries were prepared from 18 intact cells obtained from NK-cell enriched peripheral blood for single cell transcriptome analysis. Similar to bulk RNA sequencing, library preparation and Illumina sequencing was performed in the Transcriptome Analysis Laboratory (UMG).

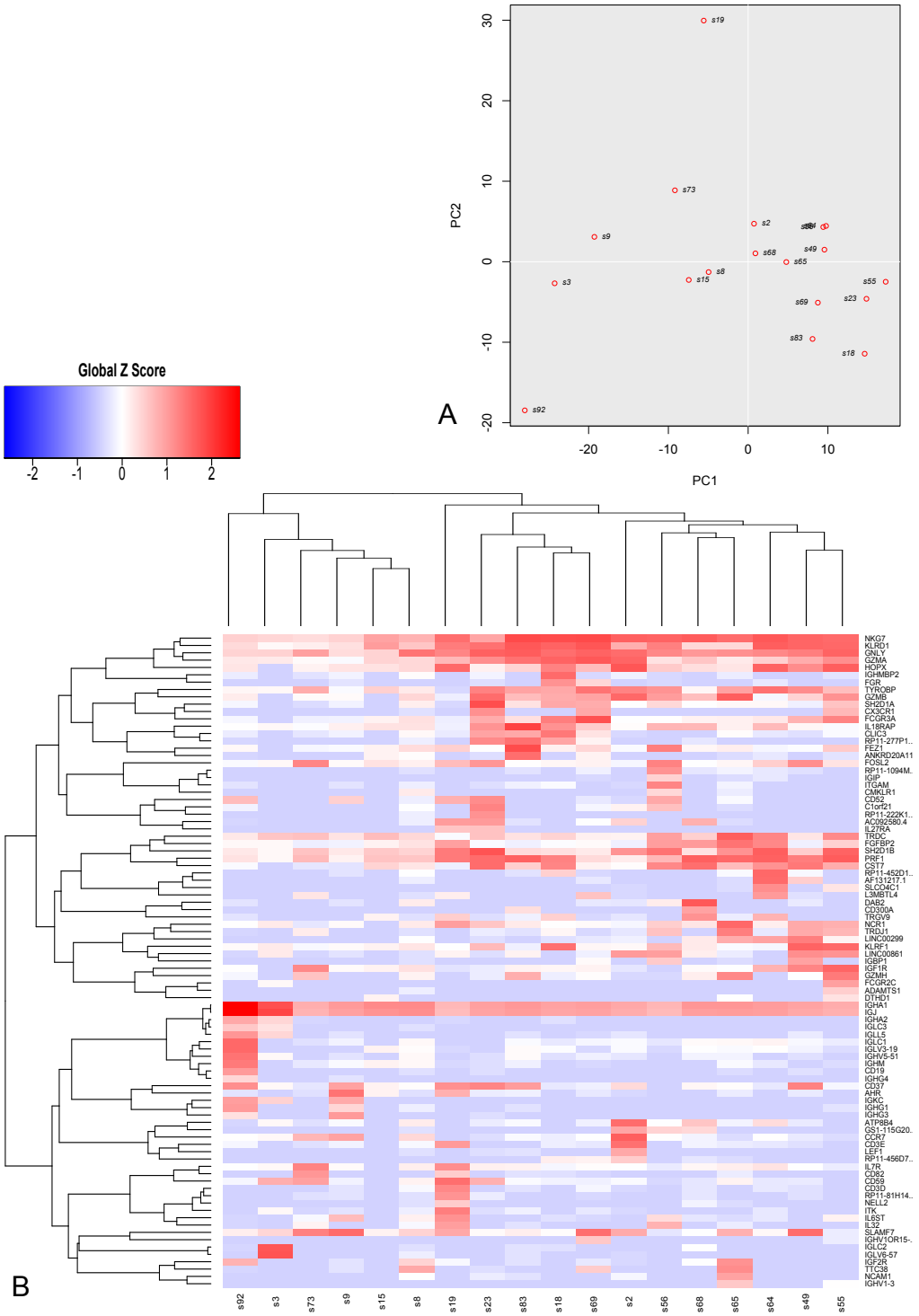

Figure 29 PCA score plot (A) and HC heatmap (B) of single-cell NK samples
Analysis was based on 89 genes selected using LoD=1.

The gene set used for analysis included protein-coding and non-protein coding genes that were identified as specifically expressed within mature NK cells (chapter 3.2). The number of genes from this set that were expressed in single cell samples counted to 86 (Appendix 16).

3.7.1. Identification of outliers

As the cells used for the experiment were not sorted by FACS, but enriched by MACS-sorting, the obtained libraries were checked for the presence of contaminating T and B lymphocytes. For this purpose, typical B cell markers (e.g. *CD19*, and multiple immunoglobulin-genes), as well as T cell markers (e.g. *CD3D, CD3E, NELL2, CD82, IL32, IL6ST, IL7R, IL27RA*) were added to the gene set described above. The total number of genes included in the analysis was 171; PCA and HC were performed with the top 89 differentially expressed genes that acceded LoD=1 threshold (Appendix 17).

When PCA and clustered heatmaps were performed using this gene set, two samples (s19 and s92) were identified as outliers (Figure 29A,B). Sample s92 expressed high levels of genes encoding immunoglobulin chains and B cell marker *CD19*, while sample s19 expressed high levels of *CD3E, CD3D, IL7R, ITK, IL32, IL6ST*. Furthermore, both of them expressed low levels of typical NK cell markers such as *TYROBP, GZMB, CX3CR1, FCGR3A* and *PRF1*. Thus, samples s19 and s92 were suggested to be contaminating T and B cells, respectively, and were excluded from further analysis.

3.7.2. Heterogeneity of gene expression in single NK cells

Upon exclusion of the outlying samples s19 and s92, the analysis was performed using the set of NK-cell specific genes (n=86); 52 of these genes exceeded the threshold of LoD=1 and were used for PCA and HC. Gene expression levels were visualized using violin plots and the Spearman's correlation analysis was performed using all 86 genes.

As can be see on violin plots (Figure 30A), for all 52 selected genes different subpopulations of cells were observed that expressed the given genes on different levels. Consistent with this, also the clustered heatmap (Figure 30B) showed different gene expression patterns. Some genes such as *GNLY, GZMA, NKG7,* and *KLRD1* were expressed ubiquitously in every single cell, while other genes were expressed only in certain cells. For example, genes *PRF1* and *CTS7* were expressed higher in nine cells (s23, s83, s69, s18, s64, s65, s49, s55, s68) than in the other eight, and *TYROBP, SH2D1B* and *GZMB* were expressed in cells s23, s83, s69, s18, s64, s65, s49, s55, s68, s56 and s2, but not in cell s68 (Figure 30B).

To investigate further relations between the analyzed genes, Spearman's correlation analysis was performed. Only correlations with an absolute value of correlation coefficient of > 0.7, $p<0.05$ or with a coefficient of <-0.7, $p<0.05$ were assumed as significantly strong. When the network of strong positive correlations was visualized (Figure 31), four correlation nods and several additional correlations were observed.

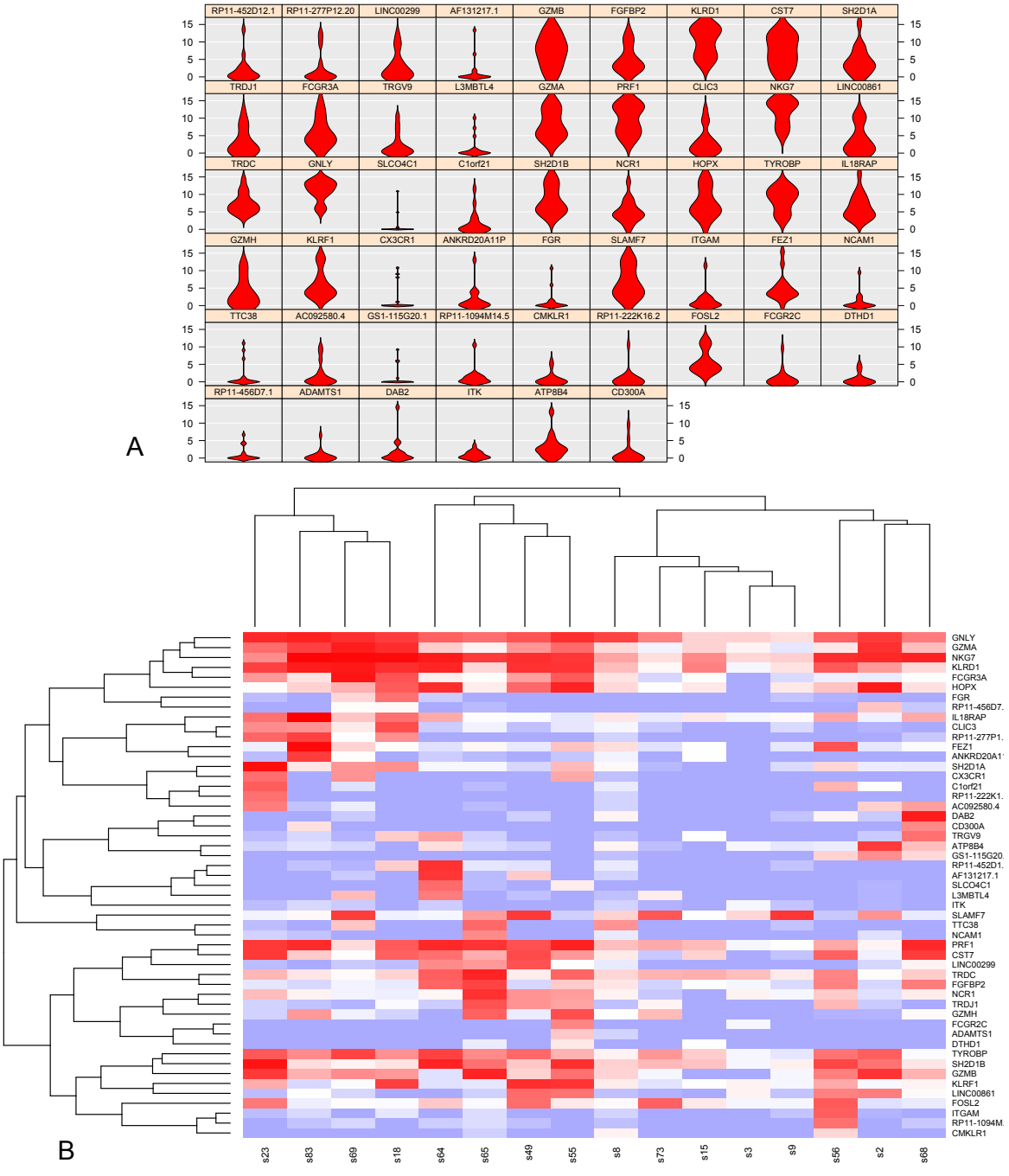

Figure 30 Violin plots (A) and HC heatmap (B) of single-cell NK samples
Analysis was based on 52 NK-specific genes selected using LoD=1.

The biggest nod included 14 genes expressed only or on higher levels in the sample s8: *PDGFD, AC017104.6, SLC1A7, RP13-516M14.2, RP11-305L7.1, RASSF4, SIPR5, RP11-18316.2, PCDH1, PDGFRB, SLC4A4, KIR3DX1, FAT4* and *CMKLR1* genes.

The second nod included mainly genes typical for mature NKs that were expressed in most of samples: *NKG7, GZMA, KLRD1, GNLY, FCGR3A, SH2D1A, CLIC3, and RP11-277P12.20* (Hayashi et al., 2007; Scheiter et al., 2013).

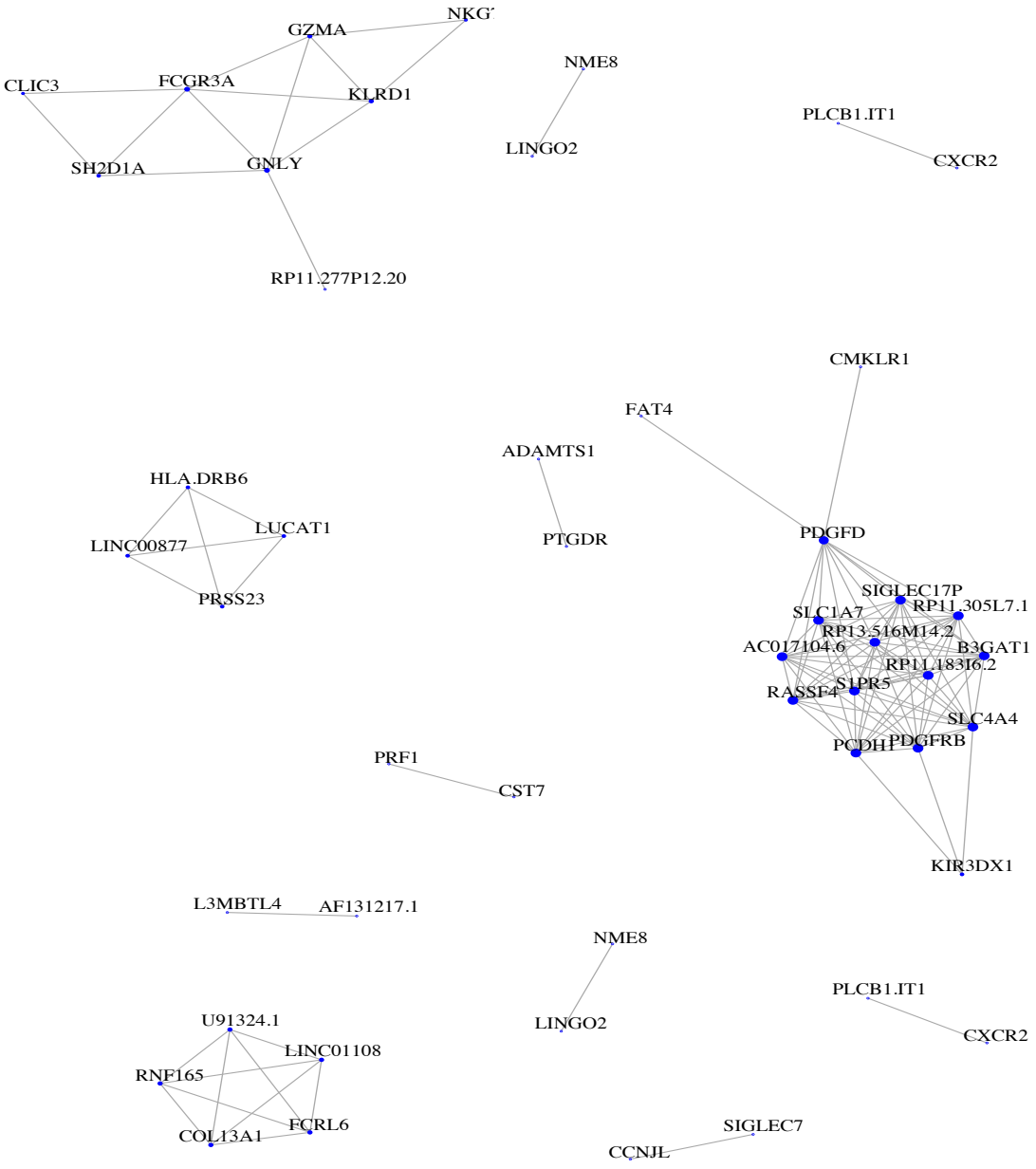

Figure 31 Correlation network of gene expression within single-cell NK samples
Only positive correlations with Spearman's coefficient>0.7, p<0.05 are shown.

The third nod consisted of five genes expressed in the cell s65 (*RNF165, COL13A1, FCRL6, LINC01108, U91324.1*).

In addition, the following correlations were observed:

- between *PRF1* and *CST7* (r=0.75);
- between *L3MBTL4* and *AF131217.1* (r=0.82);
- between *NME8* and *LINGO2* (r=1);
- between *PLCB1-IT1* and *CXCR2* (r=1);
- between *ADAMTS1* and *PTGDR* (r=0.73);
- between *CCNJL* and *SIGLEC7* (r=0.73).

Notably, no strong negative correlations were identified within the given set of genes. However, several correlations were observed with coefficients lower that -0.5 (Figure 32), which were also considered as significant:

- *SLAMF7* was negatively correlated with *FGFBP2, CST7, TRDC* and *PRF1*;
- *DTHD1* was negatively correlated with *ATP8B4* and *ITK*;
- *FCGR2C* was negatively correlated with *TRGV9* and *IL18RAP*.

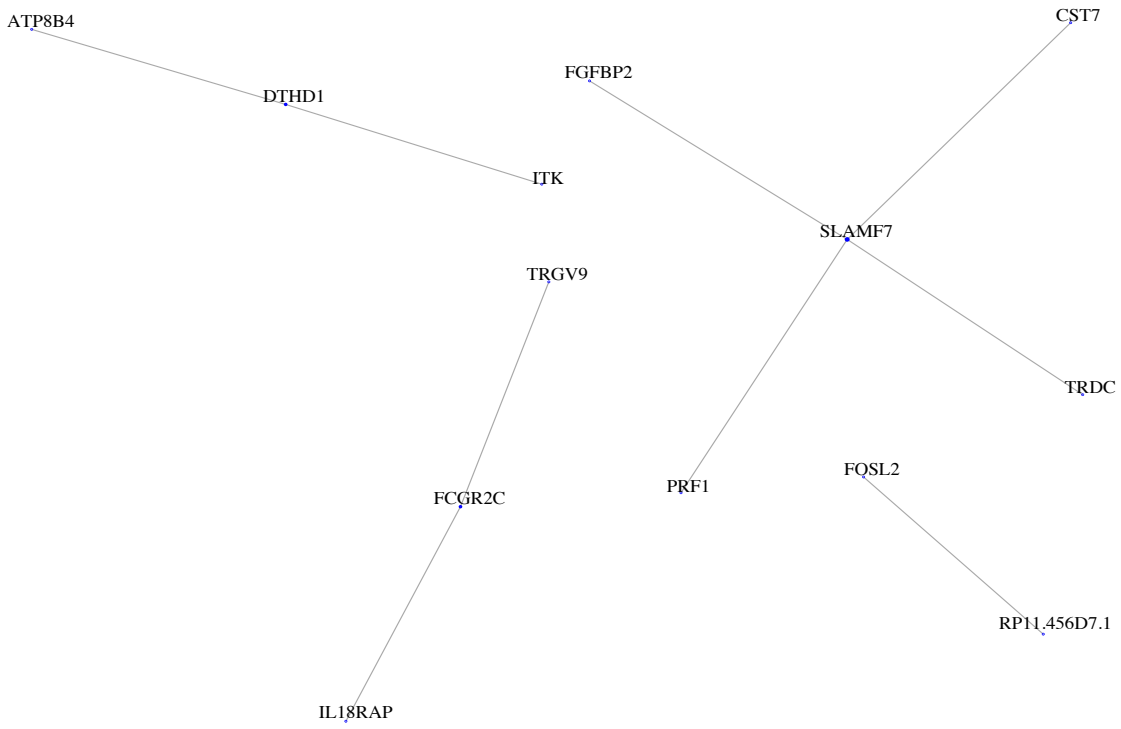

Figure 32 Correlation network of gene expression within single-cell NK samples
Only negative correlations with Spearman's coefficient<-0.7, p<0.05 are shown.

3.7.3. Regulatory potential of non-protein coding RNAs

To further examine the potential regulatory role of non-protein coding genes, the correlation between expression of protein-coding and non-protein coding genes were analyzed at the single cell level. The gene set included the set used for analysis protein-coding and non-protein coding genes identified as specifically expressed in NKs in the chapter 3.2 as well as protein-coding and non-protein coding genes identified as differentially expressed within NK developmental stages in the chapter 3.3.1. From this set only those genes expressed in more than one single cell sample were used for analysis, counting for 69 protein-coding and 26 non-protein coding genes.

Correlations with Spearman's coefficient $r>0.5$ or <-0.5 ($p<0.05$) were considered as significant (Appendix 18) Correlation with values of correlation coefficient >0.7 or <-0.7 ($p<0.05$) were considered as strong. The strongest positive correlations between protein-coding and non-protein coding genes are summarized in Table 15.

Table 15 The strongest positive correlations between protein-coding and non-protein coding genes detected at the single cell level (r>0.7, p<0,05)

Gene name	CAMK2A	CMKLR1	GNLY	KIR3DX1	L3MBTL4	NELL2	PDGFD	VWDE	XCL1
AF131217.1					0.82				
CASC15				0.77					
GS1.115G20.1									0.75
LINC01108	0.72								
MAGI2.AS3	0.84	0.72				0.72	1.00	1.00	
RP11.277P12.20			0.71						
RP11.693N9.2	0.84	0.72				0.72	0.99	0.99	

When the network of strong positive correlations was visualized (Figure 33), five groups of correlations were observed. Two non-protein coding genes were in the center of the constellation-like structure: pseudogene *RP11-693N9.2* and *MAGI2-AS3* antisense RNA. Both of them correlated with five protein-coding genes: *CAMK2A* (regulator of cell cycle; correlation coefficients 0.84 for both), *PDGFD* (mitogenic factor; 1 and 0.99 respectively), *CMKLR1* (0.72 for both), *VWDE* (1 and 0.99), and *NELL2* (cell growth regulation; 0.72 for both). Besides this, the *CAMK2A* gene expression also correlated with that of *LINC01108*. Interestingly, three of these genes, *CAMK2A, PDGFD* and *NELL2*, are involved in cell cycle regulation, and *CMKLR1* encodes a receptor of the chemoattractant chemerin, the activation of which leads to various effects including cell differentiation and

reduction of immune responses (Kaur et al., 2010; Matsumoto et al., 2002). Notably, *RP11-693N9.2* was also detected in the analysis of bulk sequencing samples (chapter 3.6.1), where it was ubiquitously expressed among NK samples from stages 3 to 5 and its expression strongly correlated with that of *PRF1*.

Besides this, the following correlations were observed:

- between *GC1-115G20.1* and *XCL1* (r=0.75);
- between *RP11-277P12.20* and *GNLY* (r=0.71);
- between *AF131217.1* and *L3MBTL4* (r=0.82);
- between *CASC15* and *KIR3DX1* (r=0.77).

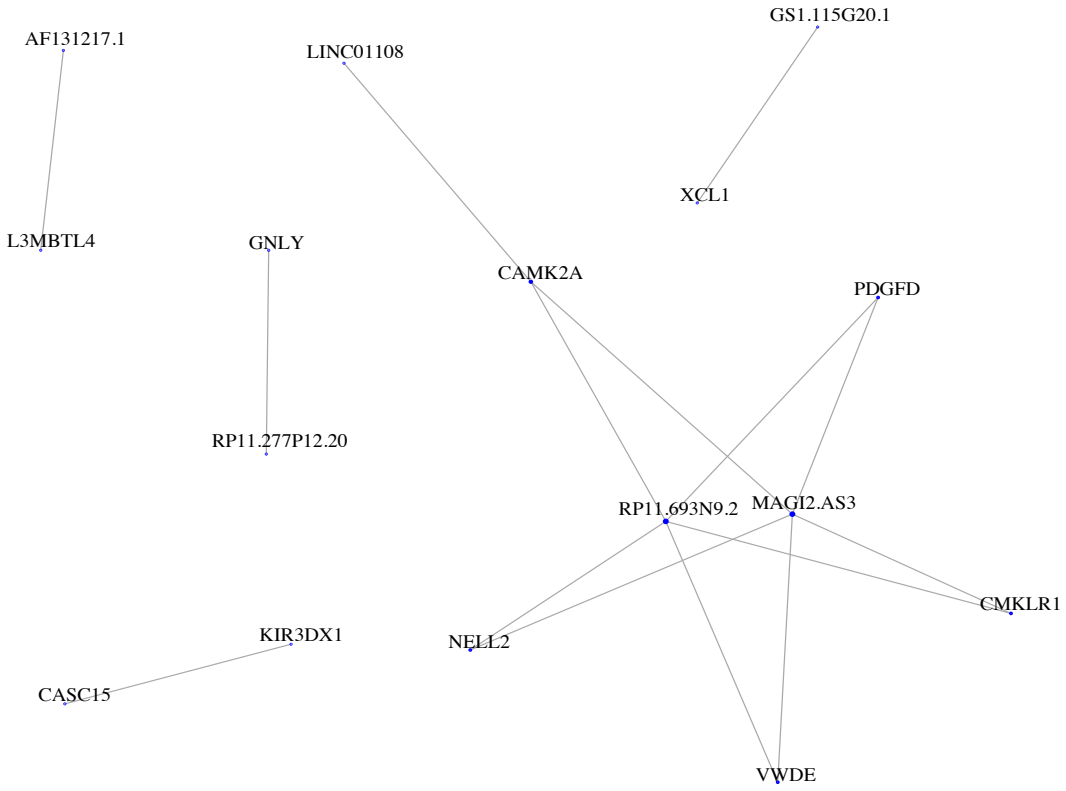

Figure 33 Correlation network of gene expression between protein-coding and non-protein coding genes within single-cell NK samples
Only positive correlations with Spearman's coefficient>0.7, p<0.05 are shown.

Also weaker positive correlations (with coefficients 0.5> r <0.7, p<0,05) were detected (Appendix 18), that were also considered being significant. In particular, the expression of *LINC00299* lncRNA (expressed in NK cells starting from stage 3 and in ILC3 stage 3 cells) correlated with that of other mature NK cell markers such as *CST7, PRF1, TGFBR3, XCL1* and *XCL2* (coefficients 0.51, 0.52, 0.63 and 0.52 respectively).

The *RP11-222K16.2* gene expression correlated with that of typical NK cell genes such as *CXCR1* and *NCAM1* (0.57 and 0.54 respectively), as well as with that of *Corf21* (0.52). Interestingly, although the expressions of *RP11-222K16.2* and *EOMES* did not correlate in the single cell data (but was the case for bulk sequencing data), the expression of *RP11-222K16.2* correlated with that of *NFIL3* (coefficient 0.59) at the single cell level. NFIL3 is required for the formation of EOMES-expressing NK cells (Seillet et al., 2014), suggesting that *RP11-222K16.2* might indeed have a role in NK development.

The expression of *RP11-277P12.20*, besides *GNLY* mentioned above, also correlated with that of genes typical for mature NK cells such as *CX3CR1, FCGR3A, FEZ1, GZMA, KLRD1*, and *SH2D1A* (coefficients 0.52, 0.69, 0.51, 0.59, 0.52, 0.65), as well as *CLIC3* (0.65), which is known to be tightly regulated in NK cells, but its function remains unclear (Kaur et al., 2010; Liu et al., 2002; Matsumoto et al., 2002;). Interestingly, as it was shown in the chapter 3.6.2, *RP11-277P12.20* gene is localized in the same locus as the *KLR* genes and according to the analysis of bulk sequencing data the expression of this non-protein coding gene was correlated not only with that of KLRD1, but also to some extend of *KLRC1, -C2, -C3, -C4,* and *–K1*.

The expression of both *MAGI2-AS3* and *RP11-693N9.2*, in addition to the abovementioned genes *CAMK2A, PDGFD, CMKLR1, VWDE*, and *NELL2*, correlated with that of three more genes: *NCAM1* (0.69 and 0.7 respectively), *TBX21* (0.61 for both) and *TTC38* (0.62 for both). NCAM1 (alias CD56) is a very important NK cell marker and TBX21 is an essential transcription factor of T_H1, ILC1 and NK cells (Montaldo et al., 2015).

Although no strong negative correlations were observed, several genes showed correlation coefficients of <-0.5 ($p<0.05$), which were also considered to be significant (Figure 34 and Table 16). In particular, the expression of the *FCGR2C* (mentioned above in the chapter 3.7.2) non-protein coding gene was negatively correlated with that of protein-coding genes *IL18RAP* and *TRGV9* (coefficients -0.69 and -0.53), while *TRGV9* was also correlated with *AC022182.1* (stage 5, differentially expressed between *ex vivo* and *in vitro*) (coefficient -0.52). Besides this, the expression of *RP11-456D7.1* (antisense RNA of *SKAP1* gene) correlated negatively with that of *FOSL2* (-0.56), which encodes a component of transcription factor AP-1 that is involved in regulation of cell proliferation, differentiation, and transformation (Bamberger et al., 2001).

Table 15 The strongest negative correlations between protein-coding and non-protein coding genes detected at the single cell level (r<-0.5, p<0,05)

Gene name	FOSL2	IL18RAP	TRGV9
AC022182.1			-0.52
FCGR2C		-0.69	-0.53
RP11.456D7.1	-0.56		

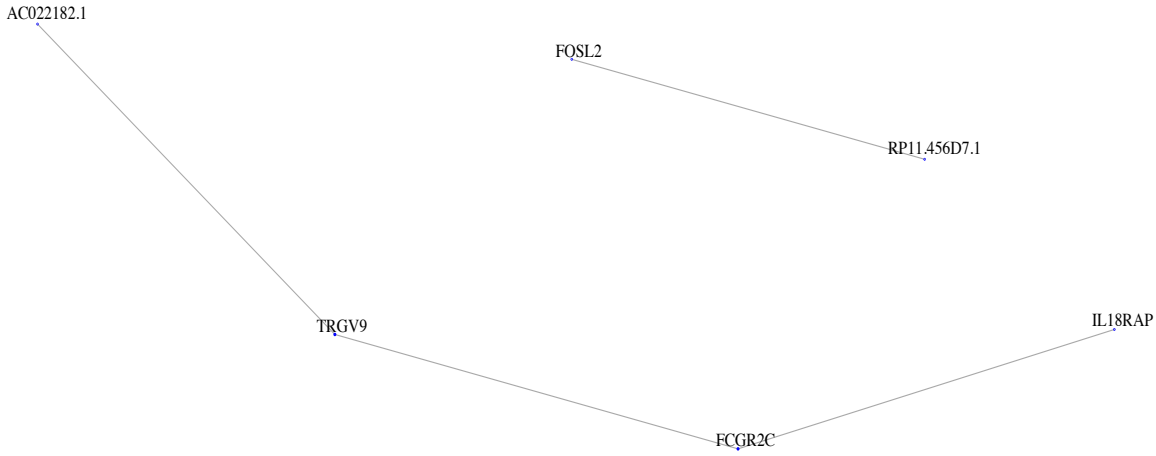

Figure 34 Correlation network of gene expression between protein-coding and non-protein coding genes within single-cell NK samples
Only negetive correlations with Spearman's coefficient<-0.5, p<0.05 are shown.

3.7.4. Comparison to bulk sequencing data

The correlations discovered in the bulk sequencing data (chapters 3.6.1 and 3.6.2) were compared to those of the single cell data. No corresponding relations between genes were identified, besides the one between *RP11-277P12.20* and *KLRD1*. However, this does not appear to be surprising, as correlations detected for populations show rather differences between NK developmental stages, tissues of origin, and eventually between donors, while the single cell data are all from mature NK cells from peripheral blood of a single blood donor.

Instead, the novel correlations were identified in single cells between genes that were not analyzed together before. These newly identified correlations were further tested on the bulk sequencing data. For this purpose, different data sets were used including the whole available data set (described in chapter 3.1), *ex vivo* NK samples from stages 1 to 5 (corresponding to the data set used in chapter 3.3.1), *ex vivo* NK samples from peripheral blood stages 4 and 5 (n=28), *ex vivo* from stage 5 in general, as well as separated nonlicensed, licensed and memory samples (all corresponding to data sets described in

chapter 3.3.3). Only correlations with values of Spearman's coefficients >0.5 or <-0.5, $p<0.05$ were considered as significant; correlations with values of Spearman's coefficients >0.7 or <-0.7 were considered to be strong (Appendix 19).

Some of the corresponding correlations could be found only among the group of all samples, suggesting that particular non-protein coding genes might be important for NK cells, and correlations between them and corresponding protein genes differ between NK cells and other samples. In particular, correlations identified between *AC092580.4* and *GZMB* (correlation coefficient 0.61), as well as between *LIN00299* and *XCL1* or *XCL2* (0.69 and 0.67 respectively).

In some other cases correlations between gene expression could be also identified in the *ex vivo* NK sample group, but were still higher in the group of all samples, for example the correlation between *AF131217.1* and *CCNJL* (coefficients 0.62 and 0.73 respectively), the correlation between *RP11-222K16.2* and *NCAM1* (0.69 and 0.76 respectively), correlation between *RP11-277P12.20* and *KLRD1* (0.53 and 0.71 respectively), as well as between *RP11-277P12.20* and *GNLY* (0.61 and 0.73 respectively).

In other cases, the correlations between particular genes were observed in all or almost all examined samples groups. In particular, *LINC00861* expression level correlated with that of *NCR1* among all samples (0.65), *ex vivo* NK sample group (0.73), stage 5 samples (0.63) and at the highest level among nonlicensed and memory samples (0.83 and 0.89 respectively). Unexpectedly, however, the correlation between them was lower when only NK samples from peripheral blood were analyzed (0.51) and completely absent among licensed samples.

Strong correlations between another lncRNA gene, *RP11-81H14.2*, and *CMKLR1* was detected among all samples (0.73), *ex vivo* NK samples (0.79), peripheral blood NKs (0.7), licensed NKs (0.77) and memory NKs (0.86), while it was lower among all stage 5 samples (0.52) and absent among nonlicensed NKs. The expression of *AF131217.1* correlated with that of *L3MBTL4* within all analyzed groups (coefficients between 0.62 and 0.83), besides licensed and memory NKs, with the highest correlation within nonlicensed stage 5 samples ($r=0.83$). Strong correlation was also observed between the genes *LINC01108* and *CAMK2A* within *ex vivo* NK cells and peripheral blood NK cell sample groups (0.78 and 0.76 respectively).

In many other observed cases corresponding correlations were identified only for specific stage 5 NK groups or where much higher than among broader groups (for example the correlation between *CASC15* and *PCDH1* (coefficient 0.86), observed only within

nonlicensed NK samples), suggesting that such non-protein coding genes might play a role during licensing or functioning of mature NKs. The expression of *AC092580.4* strongly correlated with that of *GZMA, NCAM1, PYHIN1* and *C1orf21* among memory samples (coefficients 0.93, 0.93, 0.78 and 0.82 respectively), and with *GNLY* among nonlicensed samples (0.83). Analogically, HCP5 expression correlates with both *CMKLR1* and *ITGAM* expression only among memory NK samples (coefficients 0.82 and 0.86 respectively).

The expression of *ANKRD20A11P* correlated with that of *CLIC3* and *KLRD1* expression in the strongest manner in memory NK samples (0.86 and 0.96 respectively). Interestingly, the expression of another typical NK gene, *FCGR3A* (alias *CD16*), correlated with that of *ANKRD20A11P* only in broader sample groups, especially high in *ex vivo* NK samples and peripheral blood NK samples (0.72 and 0.7). This likely corresponds to the known increase of *FCGR3A* expression during transition from stage 4 to stage 5 and suggests a possible role of *ANKRD20A11P* in NK cell maturation.

The expression of *RP11.456D7.1* was negatively correlated with that of FOSL2 in the *ex vivo* NK samples (coefficient -0.6) and stronger in the nonlicensed and licensed sample groups (-0.94 and -0.79), and this in accord with a negative correlation between them observed at the single cell level. Within the memory NK cell group, however, correlation between *RP11-456D7.1* and *FOSL2* was strongly positive (0.78), suggesting that *RP11-456D7.1* might play a role in the memory status of NK cells or in their proliferation.

Expression of the abovementioned *LINC00299* correlates with that of *CST7* and *TGFBR3* only in the memory NK cell samples, and with that of *PRF1* in peripheral blood NK cell samples, within stage 5 and in the stronger manner in memory samples (0.75, 0.71 and 0.96 respectively), but not within nonlicensed and licensed samples.

Interestingly, two non-protein coding genes *RP11-693N9.2* and *MAGI2-AS3* that were so closely related in single cell data, showed different correlation profiles in the cell populations. While *RP11-693N9.2* correlated with *CMKLR1, PDGFD, TBX21*, and *TTC38* in the strongest manner within *ex vivo* NK samples (coefficients 0.75 to 0.77) and weaker in all samples and peripheral blood, *MAGI2-AS3* correlated in a strong manner only with *CAMK2A* in memory samples (0.78) and with *PDGFRB* in licensed samples (0.71), and weaker with *VWDE* in *ex vivo* NK cell samples (0.51).

4. DISCUSSION

4.1. Proof of the methodology and data quality

The large RNA-seq data set and the analyses described in this work is, to our knowledge, the most complete that was yet performed on human NK cell populations. The data are reproducible and sufficient for successful distinguishing between the studied sample groups, namely NK cells, B and T lymphocytes, lymphocyte precursors as well as monocytes. Furthermore, genes identified as differentially expressed between groups corresponded to genes known to be involved in the immune response and to be expressed in corresponding lymphocyte populations, such as classical markers of NK cells (*NCR1, NCR3, NCAM1, KLRC1*) or lymphocyte precursors (*MYB, CPA3, FLT3, LMO2, MPO*), or classical T cell markers (*CD28, CD5, TRAT1*) (Eissens et al., 2012; Illario et al., 2008; Surmiak et al., 2012; Yang et al., 2002). The high reliability of the data set and the consistency of the observed expression patterns of known genes that are described in the literature allowed the identification of expression patterns not only of well-known genes but also of novel genes, whose function and/or expression in particular cell populations was not yet known and will be described below.

The only limitation of the applied analysis methods is the low detection level of genes that are expressed only in small subset of samples in comparison to the whole data set. In particular, genes expressed exclusively in B cell or monocyte samples were not observed due to the low number (n=2) of samples for each population. A plausible reason for this is the analysis algorithm according to which genes expressed in a small number of probes in comparison to the whole sample set might be underrepresented. First of all, the bigger the number of samples within a sample group, the more significant are conclusions about their common expression profile. This means that for sample groups with small numbers of samples, less genes can be significantly identified as specifically expressed. In addition, the bigger the group size, the higher the number of genes contained in the section of top selected genes that are differentially expressed between groups. However, this effect decreased when more homogeneous sample groups (such as separate NK cell developmental stages) were analyzed.

4.2. The specific NK cell expression profile

Previous studies on human NK cell expression profiles were performed using microarrays (Watkins et al., 2009), which limited the analysis to already known (protein-coding) genes. The RNA-seq analyses in this thesis used complete transcriptome data

instead, allowing for an unbiased analysis of all genes (protein-coding and non-protein coding) that are expressed in a certain cell population.

Among the 100 most differentially expressed genes, 41 genes turned out to have a significantly higher expression in mature NK cells (= stage 5 sample group) than in all other studied cell types, namely T and B lymphocytes as well as monocytes. Classical NK cell markers and genes described as being expressed by NK cells composed indeed the larger portion of this gene group (28 of 41) and included the well-known NK cell marker genes *GZMA, GZMB, GZMH, KIR3DX1, KLRD1, KLRF1, NCAM1, NCR1, NKG7, TYROBP, PRF1, GNLY, ADGRG1, B3GAT1, CST7, FCRL6, FGFBP2, S1PR5, SH2D1B, BNC2, CLIC3, COL13A1, PRSS23, FEZ1, LINGO2, PTGDR, HOPX,* and *MLC1* (Peng et al., 2011; Della et al.; Chiesa et al., 2010; Lozano et al., 2011; Watkins et al., 2009).

The expression profile and/or function of the other 13 genes were not yet defined in NK cells. Three of these 13 genes, *TRDC, TRDJ1*, and *TRGV9*, encode T cell receptor chains. Indeed, rearrangement and expression of T cell receptor genes is known for NK cells, although they were shown not to play a role in NK cell function Leiden et al., 1988). Seven further protein-coding genes, *PCDH1, RNF165, FAT4, SLC1A7, LGR6, SLCO4C1*, and *TTC38*, were described in immune-related studies before. However, their expression was never strictly connected to NK cells. *PCDH1*, a gene encoding a membrane protein found at cell-cell boundaries, was previously described to be ubiquitously expressed among human tissues and was suggested to be involved in neuron development (AI et al., 2004). In addition, it was shown to be associated with autoimmune pathologies such as bronchial hyperresponsiveness (Koppelman et al., 2009) and asthma (Toncheva et al., 2012). Ring finger protein 165 (RNF165) was detected as one of the genes whose expression decreases in the course of the Gulf War Illness (GWI). GWI is a syndrome characterized by a variety of symptoms including fatigue, musculoskeletal discomfort, skin rashes, and cognitive dysfunction, and recently was shown to be associated with impaired immune function, in particular with the inability of NK cells and CD8 T cells to respond to "stress-mediated activation" (Whistler et al., 2009). Together with other NK cell-related genes such as *GZMB* and *CCL4, RNF165* was shown to be downregulated in GWI patients in comparison to healthy donors, suggesting its possible role in NK cell activation. The expression of the *SLC1A7* and *TTC38* genes were previously shown to be significantly lower in PBMCs upon chronic antibody-mediated rejection after kidney transplantation, compared to kidney transplant recipients with normal graft function and histology (Rascio et al., 2015). Furthermore, *SLC1A7*, together with other NK cell-associated genes such as

FCGR3A, CST7, NCAM1, KLRK1, GZMA, was shown to be lower expressed in PBMCs of smokers (Charlesworth et al., 2010).

Besides these protein-coding genes, three non-protein coding genes, the long non-coding RNA gene *LINC00299* and the pseudogenes *SIGLEC17P* as well as *PRSS30P* completed the NK cell-specific transcription signature. The *LINC00299* non-coding RNA gene is of particular interest as it was found to be disrupted in a developmentally delayed patient and, thus, is thought to be involved in developmental processes (Talkowski et al. 2012). *LINC00299* was also demonstrated by these authors to be expressed in lymphocytes. The RNA-seq data presented here showed that *LINC00299* is specifically expressed in stage 3 ILC3 precursors and in $CD56^{bright}$ (= stage 4) and $CD56^{dim}$ (= stage 5) NK cells. Thus, *LINC00299* might play a role in the development of innate lymphoid cells of the ILC3 and NK cell type. Support for this hypothesis comes from the correlation data, as the expression of *LINC00299* is positively correlated with that of the *ID2* gene, a well-known factor involved in the development of these lymphocytes (Cherrier et al. 2012; Yokota et al. 1999;) but also in brain development (Park et al. 2013).

4.3. Changes in gene expression during NK cell development

A strong correlation of gene expression between neighboring stages of NK cell development was observed, indicating sequential changes of the gene expressions during development of these lymphocytes. The only exception was developmental stage 3, which did not correlate to the stage 4 sample group and correlated only weakly with the stage 2 sample group. Two explanations of this phenomenon are suggested: (1) the individual samples within stage 3 differ substantially (technical variation) and (2) the phenotype ($CD34^{-}CD117^{+}CD94^{-}$) used to sort stage 3 cells detects different cell types. The first hypothesis was rejected because the results of the Spearman's correlation analysis between expression profiles of individual samples indicated a high level of correlation (correlation coefficients between 0.85 and 0.93, $p<0.05$) and, thus, homogeneity of replicated experiments. Indeed, the second hypothesis was supported by previously published studies showing that the $CD34^{-}CD117^{+}CD94^{-}$ cells represent a heterogeneous population (Ahn et al., 2013; Crellin et al., 2010a; Mjosberg et al., 2011). It is known, that these cell population can give rise to T lymphocytes, DCs, and ILC3s (Eissens et al., 2012); furthermore, ILC3 precursors represent a specific subpopulation with specific characteristics within $CD34^{-}CD117^{+}CD94^{-}$ cells (Ahn et al., 2013; Hughes et al., 2010). Our further analysis also supported this explanation due to the specific expression of ILC3 markers (e.g. RORC, IL23R, IL22) within stage 3 samples from tonsils and cord blood

(known to contain also ILC3s) as compared to bone marrow stage 3 samples. It is already known that bone marrow does not include ILC3s (Ahn et al., 2013; Cupedo et al., 2009; Hughes et al., 2010; Killig et al., 2014) and, therefore, $CD34^{-}CD117^{+}CD94^{-}$ cells in the bone marrow should represent stage 3 NK cell progenitors.

NK cells develop from hematopoietic stem cells and become more and more restricted to NK cell development upon progression through the various developmental stages. Stage 1 and 2 cells still can give rise to the erythroid lineage and various granulocytes and cells of stage 3 are also able to give rise to T lymphocytes and DCs. According to Wang and colleagues, hematopoietic pluripotent cells selectively loose their pluripotency together with the expression of erythroid and granulomonocytic lineage genes and start to selectively express NK cell and T cell commited genes (Wang et al., 2015). This corresponded to the expression patterns described here (see chapter 3.3.1). While cells from stages 1 and 2 expressed a broad variety of genes related to the myeloid linage, such as *MMP8, LTF, MPO, AZU1*, and *DEFA3*, the expression of these genes decreased in stage 3. Indeed, the expression of myeloid-lineage related genes was completely lost in samples originated from tonsils, while it was still retained in cells of stage 3 samples from bone marrow and cord blood. For some of these genes the expression was retained even in cells of stage 4 samples from bone marrow. Cells of stage 3 samples from tonsils, instead, partly share the expression pattern with cells of stage 4 samples, in particular genes known to be expressed by cytotoxic NK cells, e.g. *XCL1, XCL2, CXCL8,* and *TOX2*. All in all, this leads to the conclusion that cord blood and bone marrow cells have an immature phenotype in comparison to the tonsil samples, and, at least for bone marrow, this difference remains until stage 4.

Two samples differed from the rest of their developmental group: the tonsil-originated sample from stage 2 and the cord blood-originated stage 3 sample. Both showed an intermediate position between the group of lymphocyte precursors (including also stage 3 samples from bone marrow) and stage 3 samples from tonsils. A possible explanation could be either that their expression phenotypes are intermediate between early progenitors (stages 1 and 2) and stage 3 NK cell precursors, or cells included in them represent a mixture of several cell types. A possible explanation for the position of the tonsil-derived sample was given by Luetke-Eversloh and colleagues, who identified ILC3 precursor cells among stage 2 and stage 3 cells of tonsil samples (Luetke-Eversloh et al., 2013). This also corresponds to the high expression of ILC3-related genes, such as *IL23R* and *RORC*, by this sample. For the cord blood sample both explanations could be true. On the one hand,

ILC3 cells are known to be present among stage 3 cells from the cord blood as well as tonsil-derived samples (Yu et al., 2013); on the other hand the sample from the cord blood expressed genes typical for earlier developmental stages (including such as *MPO, ELANE, RNASE2*, and *RNASE3*), suggesting that it had a less mature phenotype in comparison to tonsil-derived stage 3 samples.

4.4. Differences in expression profiles between stage 5 functional groups

Unlike cells of stages 3 and 4, no significant difference between samples from different organs was identified for cells of stage 5. Instead, a combination of at least three factors might influence the gene expression profiles: the tissue of origin, the licensing status, and the donor origin. This variability made it more or less impossible to identify exact changes that happen during licensing and education of NK cells with the available data set. This finding is in accord with previous findings and suggests that there are no major transcriptomic differences between these populations (Anfossi et al., 2006; Guia et al., 2011), showing that licensed and nonlicensed cells differ rather in their organization of cellular receptors than in their transcriptome signatures.

It is known however, that some genes (or protein epitopes such as CD57) are more strongly expressed in licensed NK cells, namely CD226 (alias DNAM-1) and KLRD1 (alias CD94) as well as in memory-like NK cells, namely CD57 and SELL (alias CD62L) (Anfossi et al., 2006; Enqvist et al., 2015; Juelke et al., 2010; Lopez-Verges et al., 2010; Sun et al., 2009). Using these genes, correlation networks of gene expression within stage 5 NK cells were build to check whether other genes show co-expression and might play a role in the licensing and memory phenotype of NK cells. Among the genes that are significantly ($p<0.05$) co-expressed with either *CD226* or *CD94* were other effector NK cell genes such as *PRF1* and *GZMA* as well as other genes typically expressed by mature NK cells such as activating receptors *CD244 (2B4), CD247 (CD3z), FCGR3A* (*CD16*), and *SLAMF7* and the cytokine receptor *CX3CR1* (Kim and Lanier, 2013; Krämer et al., 2013). In addition, among the genes that correlated with *CD226* expression were proliferation regulators *PYHIN, NKL*, and *SLFN12L*, and the transcription activator *MYBL1* (Tan et al., 2012), while the expression of the transcription co-repressor TLE3 was negatively correlated with that of *CD226*. Furthermore, the expression of GPR65 (G-protein coupled receptor) and lncRNA *RNARP11-121A8.1* correlated with the expression of both, *CD226* and *CD94*.

Taken into account that the molecular mechanism of NK cell licensing is not yet known, these observations might be used as a basis for future functional studies to analyze,

whether silencing or over-expression of any of the above mentioned genes influences the function of NK cells.

4.5. Specific expression profiles of *in vitro* differentiated NK cells

As it is impossible to study human lymphocytes *in vivo* and the sample source for *ex vivo* studies is limited (e.g. stage 2 or stage 3 cells), an established *in vitro* differentiation assay is commonly used as a model system for NK cell development and function. However, the extent of differences in gene expression between *in vitro* differentiated NK cells and *ex vivo* NK cells was unknown so far.

As shown in chapter 3.1, the transcription profiles of cells from *in vitro* samples corresponded to those of cells from the *ex vivo* samples of the same developmental stages. This indicates that the *in vitro* differentiation assay generally produces cells that correspond to *ex vivo* progenitors and mature NK cells. However, upon specific comparison of *ex vivo* and *in vitro* samples it became obvious that there are significant differences for earlier developmental stages. In particular for stages 3 and 4 the number of genes differentially expressed between cells of *ex vivo* and *in vitro* stage 5 samples was lower than the number of genes differentially expressed between the most different *ex vivo* stage 5 sample groups. Furthermore, among the genes that were highly expressed by cells of the *in vitro* samples from stage 3 and partly stage 4 were such genes as *KLRC1, KLRC2, KLRC3, KLRC4, GZMA, NCAM1, KIR2DL4, SLFN12L*, and *PBK*, which correspond to later NK cell developmental stages, and even T lymphocytes (such as *AK4, TRAT1, DEPDC1*). An obvious reason for such differences is the cytokine cocktail that is necessary for proper growth and development of cells during the *in vitro* differentiation assay. Indeed, to grow and maintain NK lymphocytes in culture, a cocktail of cytokines including stem cell factor, granulocyte-macrophage colony-stimulating factor, thrombopoetin and various Interleukins is used, which is critical for NK cell development from $CD34^+$ precursors and their maintenance (Kao et al., 2007). Although these growth factors correspond to those that NK cells face upon normal *in vivo* development, their concentrations are substantially higher in the culture medium than in the natural NK cell developmental niche. This cytokine treatment might lead to partial activation of immature NK cells and their more mature phenotype in comparison to their *in vivo* counterparts. The fact that culturing of NK cells with various cytokines leads to their activation is well known and the effect of different cytokines on NK cells alone and in combination was tested before, in particular to find cytokine combinations that would allow culturing of NK

cells without their excessive activation (Agaugué et al., 2008; Guven et al., 2003; Hromadnikova et al., 2013; Kuribayashi et al., 1981).

Nevertheless, the data clearly show the consistencies and differences between *in vitro* differentiated NK cells and their *ex vivo* counterparts and form a highly important basis for future experiments to study the role of certain genes - either protein-coding or non-protein coding - in human NK cell development.

4.6. Identification of novel genes with potential medical relevance

The data described here identified also some novel genes for which a specific function or even expression in NK lymphocytes is not yet known. The potential importance of non-protein coding genes will be described below. However, also novel protein-coding genes are of interest. Description of their specific expression during NK cell development or in mature NK cells may for example shed light on the etiology of autoimmune diseases. Among the novel candidates that are expressed in NK cells were genes with known medical relevance in systemic sclerosis, rheumatoid arthritis, asthma, psoriasis, and even in schizophrenia.

The *G0S2* (G0/G1 switch gene 2) gene is reported to be over-expressed in psoriasis (Koczan et al., 2005) and turned out in the study presented here to be also highly expressed in KIR-positive T lymphocytes and most of stage 5 NK cell samples, and on lower level in $CD8^+$ T lymphocytes. G0S2 was originally identified in PBMCs and is involved in cell cycle regulation. Initially it was believed to mediate re-entry of cells from the G0 to G1 phase of the cell cycle. However, recently its pro-apoptotic function by antagonizing BCL-2 became clear (Welch et al., 2009). Silencing of the *G0S2* gene by DNA methylation is thought to be an important feature of various cancer types, such as squamous lung cancer (Kusakabe et al., 2009). In particular, it is known to be highly expressed in proliferating lymphocytes in bone marrow, and is suggested to be one of the key regulators of the cell cycle in T lymphocytes (Kusakabe et al., 2009; Sivozhelezov et al., 2006; Yamada et al., 2012). However, its specific expression in mature NK lymphocytes from peripheral blood as well as the differential expression in different T cell subpopulations was not yet described. Even though the role of activated lymphocytes and monocytes in the peripheral blood in the pathogenesis of psoriasis is known (Abrams et al., 2000), the exact etiology remains unclear. The fact that two genes, over-expressed in psoriasis, are specifically expressed in KIR-positive lymphocytes might be helpful for a better understanding of the pathogenesis of psoriasis.

Several genes, known to participate in the regulation of the cell cycle or cellular growth control in other cell types (not known for NK cells), were also identified in the current analysis. For example *CAMK2A*, which in the current analysis was shown being expressed in mature NK cells and was also detected in NK cells at the single cell level. The product of this gene is a protein involved in the regulation of the G1 phase of the cell cycle as well as in G1/S transition; its expression and critical role in cell cycle regulation was demonstrated for brain tissue, and its presence was declared in PBMCs, however the specific role in NK cells was not described (Esteras et al., 2012; Uhlen et al., 2015;).

The *FOSL2* gene encodes a leucine zipper protein that can dimerize with proteins of the JUN family and, thereby, forming the transcription factor complex AP-1, which participates in the regulation of cell proliferation, differentiation, and transformation. *FOSL2* was previously shown to be expressed in PBMCs and to be downregulated in autoimmune diseases such as multiple sclerosis and systemic lupus erythematosus (Mandel et al., 2004). The current study allowed specifying its expression pattern, as the highest expression was seen in mature NK cells (both, in bulk RNA and single cell sequencing) in comparison to earlier developmental stages as well as to other PBMCs. A further gene implicated in an autoimmune disease is defensin alpha 4 (*DEFA4*). It was shown before to be up-regulated in peripheral blood of patients with systemic sclerosis in comparison to healthy donors and was suggested to be involved in the IFN-γ response (Bos et al., 2009). According the current analysis, it was specifically expressed in bone marrow NK cell samples in all stages, suggesting its specific and precise regulation.

In summary, specific expression patterns of various genes were newly described, in particular, genes relevant for immune pathologies. This new knowledge of gene expression may shed more light on the respective pathogenesis of the diseases.

4.7. The potential regulatory role of non-protein coding genes

It is currently clear, that non-protein coding genes such as long non-coding RNAs (lncRNA) play a distinct role in the regulation of the expression of other genes (Fatica and Bozzoni, 2014). A role of lncRNAs in T and B cell development (Casero, 2015) as well as in monocyte and DC function (Carpenter et al., 2013) was demonstrated. In NK cells, however, only the role of microRNAs was so far described (Bezman et al., 2010; Fehniger et al., 2010), while the role of lncRNAs in NK cell development and function was not known. Here, we broadened our studies to all non-protein coding genes, including lncRNAs, antisense RNAs, and pseudogenes, suggesting that if their transcription levels

are high within specific patterns, there is a high probability that they have a regulatory function in NK cells.

Both, negative and positive regulation of gene expression has been described for lncRNAs (Fatica and Bozzoni, 2014). For example lncRNAs are involved in histone modifications or may recruite corresponding regulatory proteins to specific DNA loci (Carpenter, 2013; Keller, 2013). Also natural antisense transcripts can increase the levels of the target mRNA by stabilizing its molecules (Matsui, 2008). Therefore, in the current study both, positive and negative correlations between expression of non-protein coding genes and protein-coding genes with immunological relevance were evaluated.

As shown in chapter 3.5, expression levels of non-protein coding genes change during NK cell development. This generally indicates that the expression of non-protein coding genes is also highly regulated. Unlike protein-coding genes, the expression of which is rarely limited to only one of the different stages, most of the described non-protein coding genes were expressed either during early development (stages 1, 2, and partly 3) or in mature NK cells (stages 4 and 5) and with lower expression among stage 3 samples. Smaller differences in gene expression between neighboring developmental stages were also supported by a higher pairwise correlation of gene expression profiles in comparison to expression profiles of protein-coding genes in most cases. The specific expression profile of stage 3, however, is due to the heterogeneous mixture of cells of different type as discussed above.

The strict expression profiles of non-protein coding genes in the various developmental stages suggest, that they play an important role during the cells' development. For example, non-protein coding genes expressed during early stages (e.g. *KIAA0087, RP11-212I21.4*, or *RP11-395O6.3*) might be important for turning the development towards NK cell development. In contrast, lncRNAs that are constantly expressed within committed NK cells (e.g. *LINC00299* and *RP11-222K16.2*) might be important for the function of mature NK cells. This is also consistent with a known role of lncRNAs in B and T cell development (Casero et al., 2015).

Taking into account the big variety of the examined non-protein coding genes as well as multiple potential mechanisms of their regulatory influence, different explanations can be suggested for the observed correlations in the expression of protein-coding and non-protein coding genes. Some of the identified non-protein coding genes might be important at particular developmental stages, e.g. *RP11-973H7.1* and *RP11-20I20.4* that are expressed in stages 4 and 5, and which correlated with multiple genes of the 'mature NK

cell' signature (e.g. *TBX21, KLRD1, FCGR3A, PRF1*, and granzymes). Other genes correlated more specifically to one or two protein-coding genes, e.g. *ANKRD20A11P*, which is expressed in mature NK cells and correlated with stage 5 marker *FCGR3A* (*CD16*), when their expression in all developmental stages are compared, i.e. during acquisition of FCGR3A by NK cells. Furthermore, their expression was also correlated in mature NK cells in the single cell analysis, supporting the co-expression of these genes.

Other correlations observed upon a closer look within stages could play a role in the function of NK cells, the same as was shown for activation of monocytes and DCs (Carpenter et al., 2013). In particular, an interesting candidate for such a role is the non-protein coding *gene RP11-456D7.1*, the expression of which is negatively correlated with the cell cycle regulator gene *FOSL2* based on single cell data as well as in cells of the stage 5 sample group; however, only within non-licensed and licensed samples their correlation was negative, while within memory samples it changed to be strongly positive. Taking into account that memory-like NK cells have a dramatically longer lifespan than naïve NK cells and the ability of self-renewing, it could be supposed that a specific regulation of the cell cycle with contribution of *RP11-456D7.1* and *FOSL2* might play a certain role in their specific properties (Reeves et al., 2015; Sun et al., 2009). A plausible mechanism could be that *RP11-456D7.1* is directly involved in the negative regulation of *FOSL2* expression. This could easily explain the positive regulation in naïve NK cells as well as the negative regulation in memory NK cells.

Another potential regulator is the non-protein coding gene *RP11-121A8.1*, the expression of which strongly correlated with that of *KLRD1* and *CD226* in cells of the stage 5 samples. As both, *KLRD1* and *CD226*, were previously shown to be more strongly expressed in licensed than in non-licensed cells (Anfossi et al., 2006), it is suggested that *RP11-121A8.1* might be associated with the licensing status of NK cells. Unfortunately, the expression of this gene was not detected in the single cell samples, so that its role in individual cells could not yet be checked. A possible reason for this lack of detection in single cells is the generally low level of expression of non-protein coding genes compared to protein-coding genes (Djebali et al., 2012; Fatica and Bozzoni, 2014). Similarly, the expression of *RP11-222K16.2* correlated with that of the *SELL* (alias *CD62L*). This correlation might be important for the function of mature NK cells, in particular for the function of memory-like NK cells. Notably, *RP11-222K16.2* arose in the analysis independently several times, which marks it as one of the most potent regulators identified in this research. Besides *SELL*, its expression was also correlated with such important NK

cells markers as *NCAM1* (alias *CD56*), *KLRF1, GZMA*, and *EOMES*. At the single cell level the correlation with NCAM1 was confirmed, and also the correlation with another important transcription factor, *NFIL3*, was shown.

Finally, in the case of closely localized genes such as for example *RP11-277P12.20* that maps to the locus of *KLR* genes, it could be possible that their co-expression is caused by an opened conformation of the DNA locus, particularly, if the genes are localized on the same DNA strand. However, it was also suggested by Carpenter and colleagues, that expression of lncRNAs in regions where also high expression of immune genes is encountered might be caused by their co-regulation (Carpenter et al., 2013). The fact that the expression of *RP11-277P12.20* correlates with that of only certain genes from this locus (*KLRC2, -C3, -C4* and *-K1*, but not with *KLRC1* in *ex vivo* NK cells) supports the hypothesis of co-regulation, rather than simple accidental co-expression due to open conformation of chromatin.

The next logical step in this project is to define the biological significance of the identified correlations. In future studies, the impact of over-expression and silencing of the respective lncRNAs on NK cell development and function should be performed. The first step to check the real influence of the lncRNAs is the transcription analysis at the single cell level either by RNA-seq or qRT-PCR, which allows for a more precise analysis of the correlations in individual cells and not biased by averaging within the cell population.

4.8. Single cell transcriptome analysis of lymphocytes: perspectives and limitations

Transcriptome sequencing on the population level is a good tool to determine typical expression profiles of cell groups and major differences between cell populations, as it was demonstrated here for mature NK lymphocyte populations and distinct stages of NK cell development. However, it does not allow for the analysis of variations inside the analyzed cell population, for example in cases of cell heterogeneity such as the stage 2 and stage 3 samples from tonsils and cord blood that contained NK cell and ILC3 precursor cells. Furthermore, in those cases where the average expression levels are biased by the presence of cells with very high or low expression of certain genes, the average expression level of those genes do not represent the true expression in individual cells. Therefore, RNA-seq of single cells is becoming more and more popular.

During this study single cell expression analysis of mature human NK lymphocytes using the microfluidic Fluidigm© technique was established and a pilot experiment was performed. Although the single cell RNA sequencing technique exists since several years

and was successfully used for transcriptome analysis of various cell types including hematopoietic stem cells, T lymphocytes and lymphocytic leukemia cells (Arsenio et al., 2014; Benavides-Garcia et al., 2014; Landau et al., 2014; Li et al., 2013), it was not used for the analysis of NK lymphocytes before. The main complication for this application is the small size of resting lymphocytes in general and for NK cells (5-6µm in average) and the corresponding low amount of mRNA in these cells. In the current study, NK cells, freshly isolated from peripheral blood by MACS-sorting, were directly used for the analysis, avoiding FACS-sorting or any additional staining to preserve the most natural transcription profile. It was shown, that RNA-seq of single NK cells is indeed possible, and the expression of genes associated with a typical 'mature NK cell' signature (e.g. *GNLY, GZMA, GZMB, PRF1, CTS7, NKG7, KLRD1*, and *TYROBP*), was confirmed. However, there was also a high heterogeneity of gene expression levels among single cells and this interfered with the correlation studies, in particular the correlation of expression of non-protein coding genes and protein-coding genes expressed by NK cells (in particular *LIN00299, RP11-222K16.2*, and *RP11-2777P12.20* mentioned above).

A clear disadvantage of using MACS-NK cell isolation in comparison to FACS sorting is the incomplete elimination of other cell types and, therefore, the presence of contaminating cells among the sequenced libraries. Taking this into account, the next step should be to use FACS-sorted cells and to specifically analyze certain NK cell populations: not only mature NK cells, but also earlier developmental stages, as well as a separate analysis of licensed and non-licensed cells. Furthermore, the new high-throughput dynamic arrays are expected to come to the market in the next months, which allows simultaneous analysis of up to 800 single cells in a single run. In future, the combination of these advanced techniques and the established knowledge will allow a fast acquisition of information about NK cell subpopulations and gene regulatory networks within them. This will not only foster basic research with human NK cells, but also studies using NK cells as therapeutic tools in the treatment of various diseases, in particular treatment of cancer, where NK cells are already used in clinical trials (Miller, 2013; Specht et al., 2015;).

REFERENCES

Abrams, J.R., Kelley, S.L., Hayes, E., Kikuchib, T., Browna, M.J., Kangc, S., Lebwohl M.G. et al. (2000). Blockade of T lymphocyte costimulation with cytotoxic T lymphocyte-associated antigen4- immunoglobulin (CTLA4Ig) reverses the cellular pathology of psoriatic plaques, including the activation of keratinocytes, dendritic cells, and endothelial cells. J. Exp. Med. 192, 681-94.

Agaugué, S., Marcenaro, E., Ferranti, B., Moretta, L., Moretta, A. (2008). Human natural killer cells exposed to IL-2, IL-12, IL-18, or IL-4 differently modulate priming of naive T cells by monocyte-derived dendritic cells. Blood 112, 1776 – 1783.

Ahmad, A. and Menezes, J. (1996). Antibody-dependent cellular cytotoxicity in HIV infections. FASEB J. 10, 258-266.

Ahn,Y.O., Blazar, B.R., Miller, J.S., Verneris, M.R. (2013). Lineage relationships of human interleukin-22-producing CD56+ RORγt+ innate lymphoid cells and conventional natural killer cells. Blood 121, 2234–2243.

Andersson, K.E., Williams, G.S., Davis, D.M., Höglund, P. (2007). Quantifying the reduction in accessibility of the inhibitory NK cell receptor Ly49A caused by binding MHC class I proteins in cis. Eur. J. Immunol. 37, 516–527.

Anfossi, N., Andre´, P., Guia, S., Falk, C.S., Roetynck, S., Stewart, C.A., Breso, V. et al. (2006). Human NK cell education by inhibitory receptors for MHC class I. Immunity 25, 331–342.

Arai, E., Sakamoto, H., Ichikawa, H., Totsuka, H., Gotoh, M., Mori, T., Ohnami, S. et al. (2013). Multilayer-omics (whole-exome, methylome and transcriptome) analysis identifies the Wnt/β-catenin pathway as a key player in the development of renal cell carcinoma. Cancer Res. 73, 2013-2022.

Arsenio, J., Kakaradov, B., Metz, P.J., Kim, S.H., Yeo, G.W., Chang, J.T. (2014). Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single cell gene expression analyses. Nat. Immunol. 15, 365-372.

Bamberger, A.M., Milde-Langosch, K., Rössing, E., Goemann, C., Löning, T. (2001). Expression pattern of the AP-1 family in endometrial cancer: correlations with cell cycle regulators. J. Cancer Res. Clin. Oncol 127, 545-550.

Barber, D.F., Long, E.O. (2003). Coexpression of CD58 or CD48 with intercellular adhesion molecule 1 on target cells enhances adhesion of resting NK cells. J. Immunol. 170, 294–299.

Benavides-Garcia, R., Joachim, R., Pina, N.A., Mutoji, K.N., Reilley, M.A., Hermann, B.P (2014). Granulocyte colony-stimulating factor prevents loss of spermatogenesis after sterilizing busulfan chemotherapy. Fertility and Sterility 103, 270–280.

Bezman, N.A., Cedars, E., Steiner, D.F., Blelloch, R., Hesslein, D.G.T., Lanier, L.L. (2010). Distinct requirements of microRNAs in NK cell activation, survival, and function. J Immunol 185, 3835-3846.

Bezman, N.A., Kim, C.C., Sun, J.C., Min-Oo, G., Hendricks, D.W., Kamimura, Y., Best, J.A. et al. (2012). Molecular definition of the identity and activation of natural killer cells. Nat. Immunol. 10, 1000–1009.

Biron, C.A. (1997). Activation and function of natural killer cell responses during viral infections. Curr. Opin. Immunol. 9, 24–34.

Bos, C.L., van Baarsen, L.G.M., Timmer, T.C.G., Overbeek, M.J .N., Basoski, M., Rustenburg, F., Baggen, J.M.C. et al. (2009). Molecular subtypes of systemic sclerosis in association with anti-centromere antibodies and digital ulcers. Genes and Immunity 10, 210–218.

Bottcher, A., Ostwald, J., Koczan, D., Knecht, R., Kramp, B. and Dom- merich, S. (2013). Gene expression profiling of circulating natural killer cells in head and neck squamous cell carcinoma. Cancer Genom. Proteomics. 10, 197–207.

Bottino, C., Augugliaro, R., Castriconi, R., Nanni, M., Biassoni, R., Moretta, L., Moretta A. (2000). Analysis of the molecular mechanism involved in 2B4-mediated NK cell activation: evidence that human 2B4 is physically and functionally associated with the linker for activation of T cells. Eur. J. Immunol. 2000 30, 3718–3722.

Brodin, P., Karre, K., Höglund, P. (2009a). NK cell education: not an on–off switch but a tunable rheostat. Trends Immunol. 30, 143–149.

Brodin, P., Lakshmikanth, T., Johansson, S., Kärre, K., Höglund, P. (2009b). The strength of inhibitory input during education quantitatively tunes the functional responsiveness of individual natural killer cells. Blood 113, 2434-2441.

Bryceson, Y.T., Ljunggren, H.G., Long, E.O. (2009). Minimal requirement for induction of natural cytotoxicity and intersection of activation signals by inhibitory receptors. Blood 114, 2657–2666.

Bryceson, Y.T., March, M.E., Barber, D.F., Ljunggren, H.G., Long, E.O. (2005). Cytolytic granule polarization and degranulation controlled by different receptors in resting NK cells. J. Exp Med. 202, 1001–1012.

Bryceson, Y.T., March, M.E., Ljunggren, H.G., Long, E.O. (2006). Synergy among receptors on resting NK cells for the activation of natural cytotoxicity and cytokine secretion. Blood 107, 159–166.

Burshtyn, D.N., Scharenberg, A.M., Wagtmann, N., Sumati Rajagopalan1, Berrada, K., Yi, T., Kinet, J-P., O'Long. E. (1996). Recruitment of tyrosine phosphatase HCP by the killer cell inhibitor receptor. Immunity 4, 77–85.

Campbell, J.J., Qin, S., Unutmaz, D., Soler, D., Murphy, K.E., Hodge, M.R., Wu, L., Butcher, E.C. (2001). Unique subpopulations of CD56 NK and NK-T peripheral blood lymphocytes identified by chemokine receptor expression rep- ertoire. J. Immunol. 166, 6477-6482.

Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M. et al. (2013). A long noncoding RNA mediates both activation and repression of immune response genes. Science 16, 789-792.

Casero, D., Sandoval, S., Seet, C.S., Scholes, J., Zhu, Y., Ha, V.L., Luong, A. et al. (2015). Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. Nat. Immunol. 16, 1282-1291.

Cella, M. Fuchs, A., Vermi, W., Facchetti, F., Otero, K., Lennerz, J.K.M., Doherty, J.M. et al. (2009). A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. Nature 457, 722–725.

Chan, A., Hong, D-L., Atzberger, A., Kollnberger, S., Filer, A.D., Buckley, C.D., McMichael, A. et al. (2007). CD56bright human NK cells differentiate into CD56dim cells: role of contact with peripheral fibroblasts. J. Immunol. 179, 89–94.

Charlesworth, J.C., Curran, J.E., Johnson, M.P., Göring, H.H.H., Dyer, T.D., Diego, V.P., Kent Jr, J.W. (2010). Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. BMC Med. Genom. 3, 29.

Chen, L., Kostadima, M., Martens, J.H.A., Canu, G., Garcia, S.P., Turro, E. et al. (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science 345, 1251033.

Cherrier, M., Sawa, S., Eberl, G. (2012). Notch, Id2, and RORγt sequentially orchestrate the fetal development of lymphoid tissue inducer cells. JEM 209, 729-740.

Chetyrkin, S.V., Belyaeva, O.V., Gough, W.H., ,Kedishvili, N.Y. (2001). Characterization of a novel type of human microsomal 3alpha -hydroxysteroid dehydrogenase: unique tissue distribution and catalytic properties. J. Biol. Chem. 276, 22278–22286.

Choi, Y. L., Makishima, H., Ohashi, J., Yamashita, Y., Ohki, R., Koinuma, K., Ota, J. et al., (2004). DNA microarray analysis of natural killer cell-type lymphoproliferative disease of granular lymphocytes with purified CD3- CD56+ fractions. Leukemia. 18 556–565.

Cooley, S., Xiao, F., Pitt, M., Gleason, M., McCullar, V., Bergemann, T.L., McQueen, K.L. et al. (2007). A subpopulation of human peripheral blood NK cells that lacks inhibitory receptors for self-MHC is developmentally immature. Blood 110, 578-586.

Cooper, M.A. Fehniger, T.A., Turner, S.C., Chen, K.S., Ghaheri, B.A., Ghayur, T., Carson, W.E.,. Caligiuri, M.A. (2001). Human natural killer cells: a unique innate immunoregulatory role for the CD56 (bright) subset. Blood 97, 3146-3151.

Cooper, M.A. Elliotta, J.M., Keyela, P.A., Yanga, L., Carreroc, J.A., Yokoyamaa, W.M. (2009). Cytokine-induced memory-like natural killer cells. Proc. Natl Acad. Sci. USA 106, 1915–1919.

Crellin, N.K., Trifari, S., Kaplan, C.D., Satoh-Takayama, N., Di Santo, J.P. Spits, H. (2010a). Regulation of cytokine secretion in human CD127(+) LTi-like innate lymphoid cells by Toll-like receptor 2. Immunity 33, 752–764.

Crellin, N.K., Trifari, S., Kaplan, C.D., Cupedo, T., Spits, H. (2010b). Human NKp44+IL-22+ cells and LTi-like cells constitute a stable RORC+ lineage distinct from conventional natural killer cells. J. Exp. Med. 207, 281–90.

Cruz-Munoz, M-E. and Veillette, A. (2010). Do NK cells always need a license to kill? Nat. Immunol. 11, 279-280.

Cupedo,T., Crellin, N.K., Papazian, N., Rombouts, E.J., Weijer, K., Grogan, J.L., Fibbe, W.E., Cornelissen, J.J., Spits, H. (2009). Human fetal lymphoid tissue-inducer cells are interleukin 17-producing precursors to RORC+ CD127+ natural killer-like cells. Nat. Immunol. 10, 66–74.

Djebali, S., Davis, C.A., Merkel, A,. Dobin, A., Lassmann, T., Mortazavi, A., et al. (2013). Landscape of transcription in human cells. Nature. 489, 101–108.

Dorfman, J.R., Zerrahn, J., Coles, M.C., Raulet, D.H. (1997). The basis for self-tolerance of natural killer cells in b2m Δ and TAP-1Δ mice. J Immunol. 159, 5219–5225.

Doucey, M.A. Scarpellino, L., Zimmer, J., Guillaume, P., Luescher, I.F., Bron, C., Held, W. (2004). Cis association of Ly49A with MHC class I restricts natural killer cell inhibition. Nat. Immunol. 5, 328–336.

Du, X., Tang, Y., Xu, H., Lit, L., Walker, W., Ashwood, P., Gregg, J.P. et al. (2006). Genomic profiles for human peripheral blood T cells, B cells, natural killer cells, monocytes, and polymorphonuclear cells: comparisons to ischemic stroke, migraine, and Tourette syndrome. Genomics. 87, 693–703.

Dybkaer, K., Iqbal, J., Zhou, G., Geng, H., Xiao, L., Schmitz, A. et al. (2007). Genome wide transcriptional analysis of resting and IL2 activated human natural killer cells: gene expression signatures indicative of novel molecular signaling pathways. BMC Genomics 8, 230.

Eissens, D.N., Spanholtz, J., van der Meer, A., van Cranenbroek, B., Dolstra, H., Kwekkeboom, J., Preijers, F.W.M.B., Joosten, I. (2012). Defining early human NK cell developmental stages in primary and secondary lymphoid tissues. PLoS ONE 7, e30930.

Enqvist, M., Ask, E.H., Forslund, E., Carlsten, M., Abrahamsen, G., Béziat, V., Andersson, S. et al. (2015). Coordinated Expression of DNAM-1 and LFA-1 in Educated NK Cells. J. Immunol. 94, 4518-4527.

Eriksson, M., Meadows, S.K., Wira, C.R., Sentman, C.L. (2004). Unique phenotype of human uterine NK cells and their regulation by endogenous TGF-beta. J. Leukoc. Biol. 76, 667–675.

Esteras, N., Bartolome´, F., Alque´zar, C., Antequera, D., Mun˜oz, U., Carro, E., Martin-Requero, A. (2012). Altered cell cycle-related gene expression in brain and lymphocytes from a transgenic mouse model of Alzheimer's disease. Europ. J. Neurosc. 36, 2609–2618.

Fatica, A., Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet. 15, 7–21.

Fauriat, C., Ivarsson, M.A., Ljunggren, H-G., Malmberg, K-J., Michae lsson, J. (2010). Education of human natural killer cells by activating killer cell immunoglobulin-like receptors. Blood 115, 1166-1174.

Fehniger, T.A., Cooper, M.A., Nuovo, G.J., Cella, M., Facchetti, F., Colonna, M., Caligiuri, M. A. (2003). CD56bright natural killer cells are present in human lymph nodes and are activated by T cell-derived IL-2: a potential new link be- tween adaptive and innate immunity. Blood 101, 3052-3057.

Fehniger, T.A., Wylie, T., Germino, E., Magrini, V.J., Koul S., Keppel1, C.R., Schneider, S.E. et al. (2010). Next-generation sequencing identifies the natural killer cell microRNA transcriptome. Genome Res. 20, 1590-1604.

Ferlazzo, G., Thomas, D., Lin, S.L., Goodman, K., Morandi, B., Muller, W. A., Moretta, A., Munz, C. (2004). The abundant NK cells in human secondary lymphoid tissues require activation to express killer cell Ig-like receptors and become cytolytic. J. Immunol. 172, 1455-1462.

Fernandez, N.C., Treiner, E., Vance, R.E., Jamieson, A.M., Lemieux, S., Raulet, D.H. (2005). A subset of natural killer cells achieves self-tolerance without expressing inhibitory receptors specific for self-MHC molecules. Blood 105, 4416–4423.

Fluidigm Corporation ©. (2014a). Using the C1TM Single cell Auto Prep System to Generate Single cell cDNA Libraries for mRNA Sequencing. PN 100-7168 E1

Fluidigm Corporation ©. (2014b). SINGuLARTM Analysis Toolset 3.5. User Guide. PN 100-5066 F1.

Freud, A.G., Yokohama, A., Becknell, B., Lee, M.T., Mao, H.C., et al. (2006). Evidence for discrete stages of human natural killer cell differentiation in vivo. J Exp. Med. 203, 1033–1043.

Fu, B., Tian, Z., Wei, H., (2014). Subsets of human natural killer cells and their regulatory effects. Immunology. 141, 483–489.

Garg, T.K., Szmania, S.M., Khan, J.A., Hoering, A., Malbrough, P.A., Moreno-Bost, A., Greenway, A.D. et al. (2012). Highly activated and expanded natural killer cells for multiple myeloma immunotherapy. Haematologica 97, 1348–1356.

Geremia, A., Arancibia-Carcamo, C.V., Fleming, M.P., Rust, N., Singh, B., et al. (2011). IL-23-responsive innate

Gilfillan, S., Ho, E.L., Cella, M., Yokoyama, W.M., Colonna, M. (2002). NKG2D recruits two distinct adapters to trigger NK cell activation and costimulation. Nat Immunol 3, 1150–1155.

Gillard-Bocquet, M., Caer, C., Cagnard, N., Crozet, L., Perez, M., Frid- man, W. H., Sautes-Fridman, C. et al. (2013). Lung tumor microenvironment induces specific gene expression signature in intratumoral NK cells. Front. Immunol. 4, 19.

Grégoire, C., Chasson, L., Luci, C., Tomasello, E., Geissmann, F., Vivier E., Walzer, T. (2007). The trafficking of natural killer cells. Immunol. Rev. 220, 169–182.

Guia, S., Jaeger, B.N., Piatek, S., Mailfert, S., Trombik, T., Fenis, A., Chevrier, N. et al. (2011). Confinement of activating receptors at the plasma membrane controls natural killer cell tolerance. Sci. Signal. 4, ra21.

Gurnett, C.A., Veile, R., Zempel, J., Blackburn, L., Lovett, M., Bowcock, A. (2008). Disruption of sodium bicarbonate transporter SLC4A10 in a patient with complex partial epilepsy and mental retardation. Arch. Neurol. 65, 550-553.

Guven, H., Gilljaml, M., Chambers, B.J., Ljunggren, H.G., Christensson, B., Kimby, E., Dilber, M.S. (2003). Expansion of natural killer (NK) and natural killer-like T (NKT)-cell populations derived from patients with B-chronic lymphocytic leukemia (B-CLL): a potential source for cellular immunotherapy. Leukemia 17, 1973–1980.

Hamann, I., Unterwalder, N., Cardona, A.E., Meisel, C., Zipp, F., Ransohoff, R.M., Infante-Duarte, C. (2011). Analyses of phenotypic and functional characteristics of CX3CR1-expressing natural killer cells. Immunology 133, 62–73.

Hanna, J., Bechtel, P., Zhai, Y., Youssef, F., McLachlan, K. and Man- delboim, O. (2004). Novel insights on human NK cells' immunological modalities revealed by gene expression profiling. J. Immunol. 173, 6547–6563.

Hayashi, T., Tsujii, S., Iburi, T., Tamanaha, T., Yamagami, K., Ishibashi, R., Hori, M. (2007). Laughter up-regulates the genes related to NK cell activity in diabetes. J-STAGE 28, 281-285.

Herberman, R.B., Nunn, M.E., Lavrin, D.H. (1975). Natural cytotoxic reactivity of mouse lymphoid cells against syngeneic and allogeneic tumors. I. Distribution of reactivity and specificity. Int. J. Cancer 16, 216–229.

Hoorweg, K., Peters, C.P., Cornelissen, F., Aparicio-Domingo, P., Papazian, N., Kazemier, G., Mjösberg, J.M. et al. (2012). Functional differences between human NKp44– and NKp44+ RORC+ innate lymphoid cells. Front. Immunol. 3, 72.

Horowitz, A., Strauss-Albee, D.M., Leipold, M., Kubo, J., Nemat-Gorgani, N., Dogan, O.C., Dekker, C.L. et al. (2013). Genetic and environmental determinants of human NK cell diversity revealed by mass cytometry. Sci. Transl. Med. 5, 208ra145.

Hoyler, T., Klose, C.S.N., Souabni, A., Turqueti-Neves, A., Pfeifer, D., Rawlins, E.L. et al. (2012). The transcription factor GATA-3 controls cell fate and maintenance of type 2 innate lymphoid cells. Immunity 37, 634–648.

Hromadnikova, I., Pirkova, P., Sedlackova, L. (2013). Influence of in vitro IL-2 or IL-15 alone or in combination with Hsp-70-derived 14-mer peptide (TKD) on the expression of NK cell activatory and inhibitory receptors. Mediators Inflam., 405295,

Hughes, T., Becknell, B., Freud, A.G., McClory, S., Briercheck, E., Yu, J., Mao, C. et al. (2010). Interleukin-1beta selectively expands and sustains interleukin-22+ immature human natural killer cells in secondary lymphoid tissue. Immunity 32, 803–814.

Illario, M., Giardino-Torchia, M.L., Sankar, U., Ribar, T.J., Galgani, M., Vitiello, L., Masci, A.M. et al. (2008). Calmodulin-dependent kinase IV links Toll-like receptor 4 signaling with survival pathway of activated dendritic cells. Blood 111, 723-731.

Jacobs, R., Hintzen, G., Kemper, A., Beul, K, Kempf, S., Behrens, G., Sykora, K-W., Schmidt, R.E. (2001). CD56bright cells differ in their KIR repertoire and cytotoxic features from CD56dim NK cells. Eur. J. Immunol. 31, 3121–3127.

Jiang, D., Liang, J., Hodge, J., Lu, B., Zhu, Z., Yu, S., Fan, J. et al. (2004). Regulation of pulmonary fibrosis by chemokine receptor CXCR3. J. Clin. Invest. 114, 291–299.

Johansson, S., Johansson, M., Rosmaraki, E., Vahlne, G., Mehr, R., Salmon-Divon, M., Lemonnier, F. et al. (2005). Natural killer cell education in mice with single or multiple major histocompatibility complex class I molecules. J. Exp. Med. 201, 1145–1155.

Joncker, N.T., Fernandez, N.C., Treiner, E., Vivier, E., and Raulet D.H. (2009). NK cell responsiveness is tuned commensurate with the number of inhibitory receptors for self-MHC Class I: the rheostat model. J. Immunol. 182, 4572–4580.

Juelke, K., Killig, M., Luetke-Eversloh, M., Parente, E., Gruen, J., Morandi, B., Ferlazzo, G., Thiel, A. et al. (2010). CD62L expression identifies a unique subset of polyfunctional CD56dim NK cells. Blood 116, 1299 – 1307.

Kaiser, B.K., Barahmand-Pour, F., Paulsene, W., Medley, S., Geraghty, D.E., Strong, R.K. (2005). Interactions between NKG2x immunoreceptors and HLA-E ligands display overlapping affinities and thermodynamics. J Immunol 174, 2878–2884.

Kao, I-T., Yao, C-L., Kong, Z-L., Wu, M-L., Chuang, T-L., Hwang, S-M. (2007). Generation of natural killer cells from serum-free, expanded human umbilical cord blood CD34+ cells. Stem Cel. Develop. 16, 1043-1052.

Karre, K., Ljunggren, H.G., Piontek, G., and Kiessling, R. (1986). Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defense strategy. Nature 319, 675–678

Kaur, J., Adya, R., Tan, B.K., Chen, J., Randeva, H.S. (2010). Identification of chemerin receptor (ChemR23) in human endothelial cells: chemerin-induced endothelial angiogenesis. Biochem. Biophys. Res. Commun. 391, 1762–1768.

Keller, C., Kulasegaran-Shylini, R., Shimada, Y., Hotz, H-R., Bühler, M. (2013). Noncoding RNAs prevent spreading of a repressive histone mark. Nat. Struct. Mol. Biolol. 20, 994-1000.

Kiessling, R., Klein, E., Wigzell, H. (1975). ‘‘Natural’’ killer cells in the mouse. I. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Specificity and distribution according to genotype. Eur. J. Immunol. 5, 112–117.

Killig, M., Glatzer, T., Romagnani, C. (2014). Recognition strategies of group 3 innate lymphoid cells. Frontiers Immunol. Volume 5, 142.

Kim, C. C. and Lanier, L. L., (2013). Beyond the transcriptome: completion of act one of the Immunological Genome Project. Curr. Opin. Immunol. 25, 593–597.

Kim, J.R., Horton, N.C., Mathew, S.O., Mathew, P.A. (2013). CS1 (SLAMF7) inhibits production of proinflammatory cytokines by activated monocytes. Inflamm Res. 62765-62772.

Kim, S., Poursine-Laurent, J., Truscott, S.M., Lybarger, L., Song, Y-J., Yang, L., French, A.R. et al. (2005). Licensing of natural killer cells by host major histocompatibility complex class I molecules. Nature 436, 709–713.

Kinkade, J.M., Pember, S.O., Barnes, K.C., Shapira, R., Spitznagel, J.K., Martin, L.E. (1983). Differential distribution of distinct forms of myeloperoxidase in different azurophilic granule subpopulations from human neutrophils. Biochem. Biophys. Res. Commun. 114, 296–303.

Koczan, D., Guthke, R., Thiesen, H-J., Ibrahim, S.M., Kundt, G., Krentz, H., Gross, G., Kunz, M. (2005). Gene expression profiling of peripheral blood mononuclear leukocytes from psoriasis patients identifies new immune regulatory molecules. Europ. J. Dermatol. 15, 80-84.

Koopman, L.A., Kopcow, H.D., Rybalov, B., Boyson, J.E., Orange, J.S., Schatz, F., Masch, R. et al. (2003). Human decidual natural killer cells are a unique NK cell subset with immunomodulatory potential. J. Exp. Med. 198, 1201–1212.

Koppelman, G.H., Meyers, D.A., Howard, T.D., Zheng, S.L., Hawkins, G.A., Ampleford, E.J., Xu, J. et al. (2009). Identification of PCDH1 as a novel susceptibility gene for bronchial hyperresponsiveness. Am. J. Respir. Crit. Care Med. 15, 929-935.

Kouno, T., de Hoon, M., Mar, J.M., Tomaru, Y., Kawano, M., Carninci, P., Suzuki, H. et al. (2013). Temporal dynamics and transcriptional control using single cell gene expression analysis. Genome Biol. 14, R118.

Koyasu, S. and Moro, K. (2012). Role of innate lymphocytes in infection and inflammation. Frontiers Immunol. 3, 101.

Krämer, B., Kebschull, M., Nowak, M., Demmer, R.T., Haupt, M., Körner, C., Papapanou, P. N. (2013). Role of the NK cell-activating eeceptor CRACC in periodontitis. Infection Immun. 81, 690–696.

Kuribayashi, K., Gillis, S., Kern, D.E., Henney, C.S. (1981). Murine NK cell cultures: effects of interleukin-2 and interferon on cell growth and cytotoxic reactivity. J Immunol 126, 2321-2327.

Kusakabe, M., Watanabe, K., Emoto, N., Aki, N., Kage, H., Nagase, T., Nakajima, J. (2009). Impact of DNA demethylation of the G0S2 gene on the transcription of G0S2 in squamous lung cancer cell lines with or without nuclear receptor agonists. Biochem. Biophys. Res. Com. 390, 1283-1287.

Lakschevitz, F.S., Visser, M.B., Sun, C. Glogauer, M. (2015). Neutrophil transcriptional profile changes during transit from bone marrow to sites of inflammation. Cell. Mol. Immunol. 12, 53–65.

Landau, D.A., Clement, K., Ziller, M.J., Boyle, P., Fan, J., Gu, H., Stevenson, K. et al. (2014). Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell 26, 813-825.

Lanier, L.L. (2000). Turning on natural killer cells. J. Exp. Med. 191, 1259–1262. receptors

Lanier, L.L., Le, A. M., Civin, C. I., Loken, M. R., Phillips, J. H. (1986). The relationship of CD16 (Leu-11) and Leu-19 (NKH-1) antigen expression on human peripheral blood NK cells and cytotoxic T lymphocytes. J. Immunol. 136, 4480–4486.

Lee, N., Llano, M., Carretero, M., Ishitani, A., Navarro, F., López-Botet, M., Geraghty, D. E. (1998). HLA-E is a major ligand for the natural killer inhibitory receptor CD94/ NKG2A. Proc. Natl. Acad. Sci. USA 95, 5199–5204.

Leibson, P.J. (1997) Signal transduction during natural-killer-cell activation: inside the mind of a killer. Immunity 6, 655–661.

Li, Y. Guo, X., Xue, Q., Zhu, M., Gao, L., & Wang, Y. (2013). Single cell gene profiling revealed heterogeneity of paracrine effects of bone marrow cells in mouse infarcted hearts. PloS One 8, e68270.

Liao, N., Bix, M., Zijlstra, M., Jaenisch, R., Raulet, D. (1991). MHC class I deficiency: susceptibility to natural killer (NK) cells and impaired NK activity. Science 253, 199–202.

Liu, A., Klein, G., Bandobashi, K., Klein, E., Nagy, N. SH2D1A expression reflects activation of T and NK cells in cord blood lymphocytes infected with EBV and treated with the immunomodulator PSK. Immunol. Letters 80, 181–188.

Liu, L., Huang, D., Matsui, M., He, T.T., Hu, T., DeMartino, J., Lu, B., Gerard, C., Ransohoff, R.M. (2006). Severe disease, unaltered leukocyte migration, and reduced IFN-γ production in CXCR3–/– mice with experimental autoimmune encephalomyelitis. J. Immunol. 176, 4399–4409.

Ljunggren, H-G., Karre, K. (1990). In search of the 'missing self': MHC molecules and NK cell recognition. Immunol. Today 11, 237–244.

Long, E.O., Sik Kim, H., Liu, D., Peterson, M.E., Rajagopalan, S. (2013). Controlling natural killer cell responses: integration of signals for activation and inhibition. Annu Rev Immunol. 21, 227–258.

Lopez-Verges, S., Milush, J.M., Pandey, S., York, V.A., Arakawa-Hoyt, J., Pircher, H., Norris, P.J. et al. (2010). CD57 defines a functionally distinct population of mature NK cells in the human CD56dimCD16+ NK-cell subset. Blood 116, 3865–3874.

Luci, C., Reynders, A., Ivanov, I.I., Cognet, C., Chiche, L., Chasson, L., Hardwigsen, J. et al. (2009). Influence of the transcription factor RORγt on the development of NKp46+ cell populations in gut and skin. Nat. Immunol. 10, 75–82.

Luetke-Eversloh, M., Killig, M., Romagnani, C. (2013). Signatures of human NK cell development and terminal differentiation. Frontiers Immunol. 4, 499.

Luther, C., Warner, K., Takei, F. (2011). Unique progenitors in mouse lymph node develop into CD127+ NK cells: thymus-dependent and thymus- independent pathways. Blood 117, 4012–4021.

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

Mandel, M., Gurevich, M., Pauzner, R., Kaminski, N., Achiron, A. (2004). Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus. Clin. Exp. Immunol. 138, 164–170.

Marquardt, N., Béziat, V., Nyström, S., Hengst, J., Ivarsson, M.A., Kekäläinen, E., Johansson, H. et al. (2015). Cutting edge: identification and characterization of human intrahepatic CD49a+ NK cells. J. Immunol. 194, 2467-2471.

Matsui, K., Nishizawa, M., Ozaki, T., Kimura, T., Hashimoto, I., Yamada, M., Kaibori, M. et al. (2008). Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes. Hepatology. 47, 686-97.

Matsumoto, Y., Maller, J.L. (2002). Calcium, calmodulin, and CaMKII requirement for initiation of centrosome duplication in Xenopus egg extracts. Science 295, 499-502.

Mebius, R.E,, Rennert, P., Weissman, I.L. (1997). 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.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. Nucl. Acids Res.

Miller, J.S. (2013). Therapeutic applications: natural killer cells in the clinic. Hematology Am. Soc. Hematol. Educ. Program, 247–253.

Min-Oo, G., Kamimura, Y., Hendricks, D.W., Nabekura, T. and Lanier, L.L. (2013). Natural killer cells: walking three paths down memory lane. Trends Immunol. 34, 251–258.

Miyake, N., Tonoki, H., Gallego, M., Harada, N., Shimokawa, O., Yoshiura, K., Ohta, T. et al. (2004). Phenotype–genotype correlation in two patients with 12q proximal deletion. J. Human Gen. 49, 282-284.

Mjosberg, J.M., Trifari, S., Crellin, N.K., Peters, C.P., van Drunen, C.M., Piet, B., Fokkens, W.J. et al. (2011). 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.

Montaldo, E., Juelke, K., Romagnani, C. (2015). Group 3 innate lymphoid cells (ILC3s): Origin, differentiation, and plasticity in humans and mice. Europ. J. Immunol. 45, 2171–2182.

Moro, K., Yamada, T., Tanabe, M., Takeuchi, T., Ikawa, T., Kawamoto, H., Furusawa, J.I. et al. (2010). Innate production of Th2 cytokines by adipose tissue-associated c-Kit+Sca-1+ lymphoid cells. Nature 463, 540–544.

Murphy, W.J., Kumar, V., Bennett, M. (1987). Rejection of bone marrow allografts by mice with severe combined immune deficiency (SCID): evidence that natural killer cells can mediate the specificity of marrow graft rejection. J. Exp. Med. 165, 1212-1217.

O'Leary, J.G., Goodarzi, M., Drayton, D.L., von Andrian, U.H. (2006). T cell ‐ and B cell ‐ independent adaptive immunity mediated by natural killer cells. Nat. Immunol. 7, 507–516.

Park, H.J., Hong, M., Bronson, R.T., Israel, M.A., Frankel. W.N., Yun, K. (2013). Elevated Id2 expression results in precocious neural stem cell depletion and abnormal brain development. Stem Cells. 31, 1010–21.

Park, K. U., Jin, P., Sabatino, M., Feng, J., Civini, S., Khuu, H., Berg, M. et al. (2010). Gene expression analysis of ex vivo expanded and freshly isolated NK cells from cancer patients. J. Immunother 33, 945–955.

Paust, S., Gill, H.S., Wang, B-Z., Flynn, M.P., Moseman, E.A., Senman, B., Szczepanik, M. et al. (2010). Critical role for the chemokine receptor CXCR6 in NK cell–mediated antigen-specific memory of haptens and viruses. Nat. Immunol. 11, 1127-1135.

Perera, N.C., Weismpller, K-H., Larsen, M.T., Schacher, B., Eickholy, P., Borregaard, N., Jenne D.E. (2013). NSP4 is stored in azurophil granules and released by activated neutrophils as active endoprotease with restricted specificity. J Immunol, 191, 2700-2707.

Perez, O.D., Mitchell, D., Jager, G.C., Nolan, G.P. (2004). LFA-1 signaling through p44/42 is coupled to perforin degranulation in CD56+CD8+ natural killer cells. Blood 104, 1083–1093.

Peterson, M.E., Long, E.O. (2008). Inhibitory receptor signaling via tyrosine phosphorylation of the adaptor Crk. Immunity 29, 578–588.

Racanelli, V. and Rehermann, B. (2006). The liver as an immunological organ. Hepathology 43, 54-62.

Rajagopalan, S., and Long, E.O. (1999). A human histocompatibility leukocyte antigen (HLA)-G-specific receptor expressed on all natural killer cells. J. Exp. Med. 189, 1093–1100.

Rascio, F., Pontrelli2, P., Accetturo, M., Oranger, A., Gigante, M., Castellano, G., Gigante, M. et al. (2015). A type I interferon signature characterizes chronic antibody-mediated rejection in kidney transplantation. J. Pathology 237, 72-84.

Raulet, D.H. (2006). Missing self recognition and self tolerance of natural killer (NK) cells. Semin. Immunol. 18, 145–150.

Ravetch, J.V., Lanier, L.L. (2000). Immune inhibitory receptors. Science 290, 84–89.

Reeves, R.K., Li, H., Jost, S., Blass, E., Li, H., Schafer, J.L.,Varner, V. et al. (2015). Antigen-specific NK cell memory in rhesus macaques. Nat. Immunol. 16, 927–932.

Renoux, V. M., Zriwil, A., Peitzsch, C., Michaëlsson, J., Friberg, D., Soneji, S., Sitnicka, E. (2015). Identification of a Human Natural Killer Cell Lineage-Restricted Progenitor in Fetal and Adult Tissues. Immunity 43, 394-407.

Ribeiro, V.S., Hasan, M., Wilson, A., Boucontet, L., Pereira, P., Lesjean-Pottier, S., Satoh-Takayama, N. et al. (2010). Cutting edge: Thymic NK cells develop independently from T cell precursors. J. Immunol. 185, 4993–4997.

Robertson, J.M., Caligiuri, M.A., Manley, T.J., Levine, H., Ritz, J. (1990). Human natural killer cell adhesion molecules. Differential expression after activation and participation in cytolysis. J. Immunol. 145, 3194–3201.

Robertson, M.J., Soiffer, R.J., Wolf, S.F., Manley, T.J., Donahue, C., Young, D., Herrmann, S.H., Ritz, J. (1992). Response of human natural killer (NK) cells to NK cell stimulatory factor (NKSF): cytolytic activity and proliferation of NK cells are differentially regulated by NKSF. J. Exp. Med. 175, 779–88.

Romagnani, C., Juelke, K., Falco, M., Morandi, B., D'Agostino, A., Costa, R., Ratto, G. et al. (2007). CD56brightCD16– killer Ig-like receptor– NK cells display longer telomeres and acquire features of CD56dim NK cells upon activation. J. Immunol. 178, 4947–4955.

Sánchez, L., Calvo, M., Brock, J.H. (1992). Biological role of lactoferrin. Arch. Dis. Child. 67, 657–661.

Sanos, S. L., Vonarbourg, C., Mortha, A., and Diefenbach, A. (2011). Control of epithelial cell function by interleukin-22-producing RORγt+ innate lymphoid cells. Immunology 132, 453–465.

Satoh-Takayama, N. et al. (2008). Microbial flora drives interleukin 22 production in intestinal NKp46+ cells that provide innate mucosal immune defense. Immunity 29, 958–970.

Satoh-Takayama, N., Lesjean-Pottier, S., Vieira, P., Sawa, S., Eberl, G., Vosshenrich, C.A., Di Santo, J.P. (2010). IL-7 and IL-15 independently program the differentiation of intestinal CD3–NKp46+ cell subsets from Id2-dependent precursors. J. Exp. Med. 207, 273–280.

Satoh-Takayama, N., Lesjean‐Pottier, S., Sawa, S., Vosshenrich, C.A., Eberl, G. and Di Santo, J.P. (2011) Lymphotoxin-beta receptor-independent development of intestinal IL-22-producing NKp46+ innate lymphoid cells. Eur. J. Immunol. 41, 780‐786.

Scheiter, M., Lau, U., van Ham, M., Bulitta, B., Gröbeß, L., Garritsen, H., Klawonn, F. (2013). Proteome analysis of distinct developmental stages of human natural killer (NK) cells. Mol. Cell. Proteomics 12, 1099–1114.

Seaman, W., Sleisenger, M., Eriksson, E., Koo, G. (1987). Depletion of natural killer cells in mice by monoclonal antibody to NK-1.1. Reduction in host defense against malignancy without loss of cellular or humoral immunity. J. Immunol. 138, 4539–4544.

Seillet, C., Huntington, N.D., Gangatirkar, P., Axelsson, E., Minnich, M., Brady, H.J., Busslinger, M. et al. (2014). Differential requirement for Nfil3 during NK cell development. J. Immunol. 192, 2667– 2676.

Sim, M.J.W., Stowell, J., Sergeant, R., Altmann, D.M., Long, E.O., Boyton, R.J. (2015). KIR2DL3 and KIR2DL1 show similar impact on licensing of human NK cells. Eur. J. Immunol. 46, 185-191.

Sivozhelezov, V., Giacomelli, L., Tripathi, S., Nicolini, C. (2006). Gene expression in the cell cycle of human T lymphocytes: I. Predicted gene and protein networks J. Cell. Biochem. 97, 1137–1150.

Specht, H.M., Ahrens, N., Blankenstein, C., Duell, T., Fietkau, R., Gaipl. U.S., Günther, C. et al. (2015). Heat shock protein 70 (Hsp70) peptide activated natural killer (NK) cells for the treatment of patients with non-small cell lung cancer (NSCLC) after radiochemotherapy (RCTx) - from preclinical studies to a clinical phase II trial. Front Immunol. 6, 162.

Spits, H., Artis, D., Di Santo, J.P., Eberl, G., Koyasu, S., Locksley, R.M. et al. (2013). Innate lymphoid cells — a proposal for uniform nomenclature. Nat. Rev. Immunol. 13, 145-149.

Spits, H., Di Santo, J.P. (2011). The expanding family of innate lymphoid cells: regulators and effectors of immunity and tissue remodeling. Nat. Immunol. 12, 21–27.

Stewart, C.A., Laugier-Anfossi, F., Vély, F., Saulquin, X., Riedmuller, J., Tisserant, A., Gauthier, L. et al. (2005). Recognition of peptide- MHC class I complexes by activating killer immunoglobulin-like receptors. Proc. Natl. Acad. Sci. USA 102, 13224– 13229.

Storkus, W.J., Salter, R.D., Alexander, J., Ward, F.E., Ruiz, R.E., Cresswell, P., Dawson, J.R. (1991). Class I-induced resistance to natural killing: identification of nonpermissive residues in HLA-A. PNAS 88, 14.

Su, A.I, Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block D. et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes PNAS 101, 6062–6067.

Suhre, K., Shin, S.Y, Petersen, A.K., Mohney, R.P., Meredith, D., Wägele, B., Altmaier, E., Deloukas, P. et al. (2011). Human metabolic individuality in biomedical and pharmaceutical research. Nature 477, 54–60.

Sun, H., Sun, C., Tian, Z., Xiao, W. (2013). NK cells in immunotolerant organs. Cell. Mol. Immunol. 10, 202–212.

Sun, J. C., Beilke, J. N., Lanier, L. L. (2009). Adaptive immune features of natural killer cells. Nature 457, 557–561.

Sun, J.C., Madera, S., Bezman, N.A., Beilke, J.N., Kaplan, M.H., Lanier, L.L. (2012). Proinflammatory cytokine signaling required for the generation of natural killer cell memory. JEM 209, 947-954.

Surmiak, M., Kaczor, M., Sanak, M. (2012). Expression profile of proinflammatory genes in neutrophil-enriched granulocytes stimulated with native anti-PR3 autoantibodies. J. Physiol. Pharmacol. 63, 249-256.

Takatori, H., Kanno, Y., Watford, W.T., Tato, C.M., Weiss, G., Ivanov, I.I., Littman, D.R. and O'Shea, J.J. (2009). Lymphoid tissue inducer-like cells are an innate source of IL-17 and IL-22. J. Exp. Med. 206, 35–41.

Talkowski, M.E., Maussion, G., Crapper, L., Rosenfeld, J.A., Blumenthal, I., Hanscom, C., Chiang, C. et al. (2012). Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities. Americ. J. Human Gen., 91, 1128-1134.

Tan, Z., Li, M., Wu, W., Zhang, L., Ding, Q., Wu, X., Mu, J., Liu, Y. (2012). NLK is a key regulator of proliferation and migration in gallbladder carcinoma cells. Mol. Cel. Bicohem. 369, 27-33.

Tang, Q., Ahn, Y.O., Southern, P., Blazar, B.R., Miller, J.S., Verneris, M.R. (2011). Development of IL-22- producing NK lineage cells from umbilical cord blood

hematopoietic stem cells in the absence of secondary lymphoid tissue. Blood 117, 4052–4055.

Thapa, M., Kuziel, W.A., Carr, D. J. (2007). Susceptibility of CCR5-deficient mice to genital herpes simplex virus type 2 is linked to NK cell mobilization. J. Virol. 81, 3704–3713.

Thomas, M.L., Peterson M.E., Long E.O. (2013). Cutting Edge: NK Cell Licensing Modulates Adhesion to Target Cells. J. Immunol. 191, 3981-3985.

Thomas, M.P., Whangbo, J., McCrossan, G., Deutsch, A.J., Martinod, K., Walch, M., Lieberman, J. (2014). Leukocyte protease binding to nucleic acids promotes nuclear localization and cleavage of nucleic acid binding proteins. J. Immunol. 192, 5390–5397.

Toncheva, A.A., Suttner, K., Michel, S., Klopp, N., Illig, T., Balschun, T., Vogelberg, C. et al. (2012). Genetic variants in Protocadherin-1, bronchial hyperresponsiveness, and asthma subphenotypes in German children. Pediatr. Allergy Immunol. 23, 636-641.

Trundley, A., Moffet, A. (2004). Human uterine leukocytes and pregnancy. Tis. Antigens 63, 1-12.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. (2015). Tissue-based map of the human proteome. Science 347, 1260419.

Uhrberg, M., Valiante, N.M., Young, N.T., Lanier, L.L., Phillips, J.H., Parham, P. (2001). The repertoire of killer cell Ig-like receptor and CD94:NKG2A receptors in T cells: clones sharing identical αβ TCR rearrangement express highly diverse killer cell Ig-like receptor patterns. J. Immunol. 166, 3923-3932.

Veillette, A., Dong, Z., Latour, S. (2007). Consequence of the SLAM-SAP signaling pathway in innate-like and conventional lymphocytes. Immunity 27, 698–710.

Vosshenrich, C.A., Garcı́a-Ojeda, M.E., Samson-Villéger, S.I., Pasqualetto, V., Enault, L., Richard-Le Goff, O., Corcuff, E. et al. (2006). A thymic pathway of mouse natural killer cell development characterized by expression of GATA-3 and CD127. Nat. Immunol. 7, 1217–1224.

Wang, F., Tian, Z., Wei, H. (2015). Genomic expression profiling of NK cells in health and disease Eur. J. Immunol. 45, 661–678.

Wang, F., Zhou, Y., Fu, B., Wu, Y., Zhang, R., Sun, R., Tian, Z., Wei, H. (2014). Molecular signatures and transcriptional regulatory networks of human immature decidual NK and mature peripheral NK cells. Eur. J. Immunol. 44, 2771–2784.

Warner, K., Luther, C., Takei, F., (2012). Lymphoid progenitors in normal mouse lymph nodes develop into NK cells and T cells in vitro and in vivo. Exp. Hematol. 40, 401–406.

Welch, C., Santra, M.K., El-Assaad, W., Zhu, X., Huber, W.E., Keys, R.A., Teodoro, J.G., Green, M.R. (2009). Identification of a protein, G0S2, that lacks Bcl-2 homology domains and interacts with and antagonizes Bcl-2. Cancer Res. 69, 6782.

Wendt, K., Wilk, E., Buyny, S., Buer, J., Schmidt, R.E., Jacobs, R. (2006). Gene and protein characteristics reflect functional diversity of CD56dim and CD56bright NK cells. J. Leukoc. Biol. 80, 1529–1541.

Whistler, T., Fletcher, M.A., Lonergan, W., Zeng, X-R., Lin, J-M., LaPerriere, A., Vernon, S.D., Klimas, N.G. (2009). Impaired immune function in Gulf War Illness. BMC Med. Genom. 2, 12.

White, A.K., VanInsberghe, M., Petriv, O.I., Hamidi, M., Sikorski, D., Marra, M.A. et al. (2011). High-throughput microfluidic single cell RT-qPCR. PNAS 108, 13999–14004.

Wilhelm, C., Hirota, K., Stieglitz, B., Van Snick, J., Tolaini, M., Lahl, K., Sparwasser, T. et al. (2011). An IL-9 fate reporter demonstrates the induction of an innate IL-9 response in lung inflammation. Nat. Immunol. 12, 1071–1077.

Witt, C. S., Whiteway, J.M., Warren, H.S., Barden, A., Rogers, M., Martin, A., Beilin, L. and Christiansen, F.T. (2002). Alleles of the KIR2DL4 receptor and their lack of association with preeclampsia. Eur. J. Immunol. 32, 18–29.

Wulfing, C., Purtic, B., Klem, J., Schatzle, J.D. (2003). Stepwise cytoskeletal polarization as a series of checkpoints in innate but not adaptive cytolytic killing. Proc. Natl. Acad. Sci. USA 100, 7767–7772.

Yamada, T., Park, C.S., Burns, A., Nakada, D., Lacorazza, H.D. (2012). The cytosolic protein G0S2 maintains quiescence in hematopoietic stem cells. PLoS ONE 7, e38280.

Yamagiwa, S., Kamimura, H., Ichida, T. (2009). Natural killer cell receptors and their ligands in liver diseases. Med. Mol. Morphol. 42, 1–8.

Yang, J.J., Preston, G.A., Alcorta, D.A., Waga, I., Munger, W.E., Hogan, S.L., Sekura, S.B. et al. (2002). Expression profile of leukocyte genes activated by anti-neutrophil cytoplasmic autoantibodies (ANCA). Kidney Internat. 62, 1638–1649.

Yang, Q., Monticelli, L.A., Saenz, S.A., Chi, A.W.S., Sonnenberg, G.F., Tang, J., Obaldia, M.E. et al. (2013). T cell factor 1 is required for group 2 innate lymphoid cell generation. Immunity 38, 694-704.

Yokota, Y., Mansouri, A., Mori, S., Sugawara, S., Adachi, S., Nishikawa, S-I., Gruss, P. (1999). Development of peripheral lymphoid organs and natural killer cells depends on the helix–loop–helix inhibitor Id2. Nature 397, 702-706.

Yokoyama, W.M., Kim, S. (2006). How do natural killer cells find self to achieve tolerance? Immunity 24, 249–257.

Young, N.T., Uhrberg, M., Phillips, J.H., Lanier, L.L., Parham, P. (2001). Differential expression of leukocyte receptor complex-encoded Ig-like receptors correlates with the transition from effector to memory CTL. J. Immunol. 166, 3933-3941.

Yu, J., Freud, A.G., Caligiuri, M.A. (2013). Location and cellular stages of natural killer cell development. Trends Immunol. 34, 573-582.

Yu, J., Mao, H.C., Wei, M., Hughes, T., Zhang, J., et al. (2010). CD94 surface density identifies a functional intermediary between the CD56bright and CD56dim human NK cell subsets. Blood 115, 274 – 281.

Zhang, Q.H., Ye, M., Wu, X.Y., Ren, S.X., Zhao, M., Zhao, C.J., Fu, G. et al. (2000). Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells. Genome Res. 10, 1546-60.

Zhang, Z., Wu, N., Lu, Y., Davidson, D., Colonna, M., Veillette, A. (2015). DNAM-1 controls NK cell activation via an ITT-like motif. JEM 212. 2165-2182.

APPENDIXES

Appendix 1 Ranked top 100 genes differentially expressed within all analyzed samples

Genes identified from the total number of 20858 genes using LoD1.

GeneID	Score (PC1-3)	Rank	GeneID	Score (PC1-3)	Rank
SH2D1B	0.046211855	1	PRTN3	0.030322462	51
GZMB	0.040545715	2	BCAT1	0.030315395	52
NMUR1	0.040207182	3	FAT4	0.030211952	53
FCRL6	0.040197263	4	TBX21	0.03012797	54
NCAM1	0.038953949	5	TNS3	0.030106641	55
GNLY	0.038288605	6	CD247	0.030077825	56
KLRD1	0.037813562	7	SLAMF7	0.030052784	57
NCR1	0.037653283	8	IGLL1	0.030017484	58
FCGR3A	0.037649381	9	FCRL3	0.029920427	59
MPO	0.037251151	10	IKZF3	0.029797926	60
AC017104.6	0.036800003	11	ITGAM	0.029513769	61
PTGDR	0.036258052	12	LMO2	0.029462625	62
TRDC	0.036126767	13	RASSF4	0.029388897	63
CLIC3	0.036004902	14	ANGPT1	0.029335414	64
SIGLEC7	0.035499688	15	CD300A	0.029213044	65
S1PR5	0.035409046	16	PDGFRB	0.02917724	66
CX3CR1	0.034972368	17	L3MBTL4	0.029160946	67
FGR	0.034625771	18	BNC2	0.02915969	68
BCL11B	0.034583362	19	CAMK4	0.029086805	69
XCL2	0.03455931	20	MYB	0.029067877	70
CD28	0.034289992	21	AC010468.2	0.029043678	71
FLT3	0.034002569	22	PLD4	0.028976966	72
DTHD1	0.033973869	23	PDZD4	0.028938765	73
CPA3	0.033922611	24	MSRB3	0.028857116	74
LINC00299	0.033912246	25	MAP7	0.028810967	75
CMKLR1	0.033562424	26	CD5	0.028809729	76
PRF1	0.033299088	27	PDGFD	0.028803034	77
XCL1	0.033258596	28	SH2D1A	0.028796607	78
NKG7	0.03300908	29	RGS9	0.028767589	79
FGFBP2	0.032872492	30	CCR4	0.028723691	80
GZMH	0.03281785	31	CTSG	0.028711259	81
MLC1	0.032537178	32	TRAT1	0.028709552	82
IFNG	0.032480375	33	COL13A1	0.028656065	83
GZMM	0.032363887	34	NELL2	0.028652299	84
SLC16A10	0.0322795	35	ALDH1A1	0.028585352	85
CACNA1I	0.032252599	36	ATP8B4	0.028576339	86
LEF1-AS1	0.032234284	37	KLRC1	0.028537801	87
TRDJ1	0.032179722	38	RNF165	0.028528544	88
RP11-	0.032059439	39	MAL	0.028444411	89

705C15.5					
TYROBP	0.031777736	40	EGFL7	0.028399162	90
PYHIN1	0.031688328	41	KRT86	0.028397972	91
B3GNT7	0.031598814	42	ITK	0.028382932	92
GZMA	0.031564505	43	LINGO2	0.02833046	93
TRGV9	0.031166321	44	PKIA-AS1	0.028317236	94
CCNJL	0.030822052	45	CTC-499J9.1	0.02823147	95
RP11-121A8.1	0.03081529	46	IL18RAP	0.028185366	96
SPINK2	0.030793462	47	LPCAT2	0.028134425	97
SIGLEC17P	0.030705875	48	RP11-104L21.3	0.028116524	98
NCR3	0.030591624	49	GRIK4	0.028057238	99
TENM1	0.030418915	50	HOPX	0.027980855	100

Appendix 2 Genes differentially expressed between PBMC populations identified using ANOVA analysis

Expression levels are indicated in Log_2 domain.

2.1 Genes differentially expressed between NK and T lymphocytes. Genes higher expressed in NK cells are marked in green.

		Expression levels				Expression levels	
GeneID	**p Value**	**NK**	**T**	**GeneID**	**p Value**	**NK**	**T**
GNLY	8.16E-13	14.27	6.47	LGR6	8.17E-13	8.68	2.33
PRF1	8.30E-13	14.18	7.78	FEZ1	8.25E-13	8.51	2.01
NKG7	8.16E-13	13.82	5.80	CMKLR1	7.95E-12	8.47	1.22
KLRD1	8.16E-13	12.74	4.40	L3MBTL4	3.63E-12	8.42	2.20
GZMB	8.16E-13	12.56	3.65	RNF165	8.16E-13	8.35	1.79
ITGAM	8.16E-13	12.17	5.32	GPR141	8.16E-13	8.12	0.83
FGFBP2	9.18E-13	12.13	4.87	KIR3DX1	8.16E-13	8.11	1.19
CST7	8.16E-13	12.09	5.46	DAB2	8.16E-13	7.91	0.94
S1PR5	8.16E-13	11.93	3.94	TRGV9	8.16E-13	7.90	1.43
GZMA	8.17E-13	11.81	4.92	PDGFRB	8.17E-13	7.88	0.50
FCGR3A	8.16E-13	11.60	2.88	B3GAT1	1.72E-12	7.79	1.37
KLRF1	8.19E-13	11.58	4.99	PCDH1	8.16E-13	7.68	0.68
TTC38	8.16E-13	11.47	4.83	NME8	8.16E-13	7.62	0.73
SH2D1B	8.16E-13	11.35	1.96	CXCR2	8.16E-13	7.45	0.85
GZMH	8.25E-13	11.32	3.70	PRSS30P	1.08E-12	7.44	1.58
FGR	8.16E-13	11.14	4.16	SIGLEC17P	8.16E-13	7.20	0.30
ADGRG1	8.18E-13	11.05	4.29	SLC1A7	1.36E-11	6.98	0.60
CD300A	8.16E-13	10.72	4.30	SIGLEC7	8.17E-13	6.79	0.24
HOPX	8.16E-13	10.70	3.43	LINGO2	8.21E-13	6.69	0.11
CX3CR1	1.19E-10	10.63	3.72	TRDJ1	8.16E-13	6.54	0.08
TRDC	8.16E-13	10.60	3.12	COL13A1	1.01E-11	6.32	0.22
PRSS23	8.16E-13	10.59	3.18	HCK	0.025926292	4.38	1.98
NCAM1	8.16E-13	10.42	2.26	IL7R	1.75E-11	5.85	13.59

SYK	8.16E-13	10.18	3.29	CAMK4	8.16E-13	5.30	12.80
ATP8B4	8.16E-13	10.17	2.43	NELL2	1.42E-12	4.11	11.97
TYROBP	8.16E-13	10.06	2.88	CCR7	8.28E-13	3.91	11.54
FCRL6	8.16E-13	10.06	1.32	CD5	8.26E-13	3.73	11.65
PTGDR	8.16E-13	10.02	2.83	TRAT1	7.61E-12	3.55	10.53
LINC00299	8.16E-13	9.93	3.06	CD28	8.16E-13	2.96	12.34
BNC2	8.16E-13	9.89	1.50	SLAMF1	8.58E-13	2.69	9.02
NCR1	8.16E-13	9.87	2.47	AXIN2	9.18E-13	2.03	8.56
B3GNT7	8.17E-13	9.83	1.97	SLC16A10	8.20E-13	1.43	8.58
MLC1	8.16E-13	9.74	1.58	ALS2CL	8.19E-13	1.40	7.80
PDGFD	8.16E-13	9.51	2.16	SIT1	8.20E-13	1.36	8.12
RASSF4	8.16E-13	9.37	2.71	CCR4	8.16E-13	0.81	9.08
NMUR1	8.16E-13	9.35	0.88	LEF1-AS1	8.16E-13	0.80	8.48
FAT4	8.16E-13	9.15	0.66	FAAH2	8.16E-13	0.71	7.40
SLCO4C1	8.16E-13	9.07	2.01	MAL	8.16E-13	0.69	8.72
CLIC3	8.19E-13	8.97	2.59	CACNA1I	8.16E-13	0.37	8.05
AC017104.6	8.16E-13	8.88	1.07	CTC-499J9.1	8.16E-13	0.23	6.94

2.2 Genes differentially expressed between NK and B lymphocytes. Genes higher expressed in NK cells are marked in green.

GeneID	p Value	NK	B	GeneID	p Value	NK	B
GNLY	2.38E-12	14.27	3.87	CLIC3	3.91E-06	8.97	3.29
PRF1	3.41E-08	14.18	6.45	LGR6	2.64E-10	8.68	0.34
NKG7	1.82E-09	13.82	4.88	FEZ1	5.24E-08	8.51	0.97
KLRD1	8.14E-10	12.74	3.18	CMKLR1	0.000204057	8.47	1.56
GZMB	3.76E-09	12.56	4.10	RNF165	1.15E-10	8.35	1.34
ITGAM	0.014038922	12.17	9.87	GPR141	2.08E-09	8.12	0.34
FGFBP2	2.28E-08	12.13	2.49	KIR3DX1	5.89E-08	8.11	1.50
CST7	1.51E-12	12.09	3.29	DAB2	3.82E-05	7.91	3.06
S1PR5	5.77E-10	11.93	2.83	TRGV9	1.13E-11	7.90	1.00
GZMA	5.61E-11	11.81	2.49	PDGFRB	0.007194957	7.88	4.33
FCGR3A	2.75E-07	11.60	3.56	B3GAT1	2.36E-05	7.79	1.24
KLRF1	3.81E-09	11.58	3.56	PCDH1	3.59E-11	7.68	0.00
TTC38	1.70E-09	11.47	6.44	NME8	8.75E-08	7.62	1.57
SH2D1B	4.81E-11	11.35	2.16	CXCR2	3.20E-08	7.45	0.97
GZMH	4.08E-08	11.32	2.35	PRSS30P	1.40E-07	7.44	0.00
ADGRG1	1.35E-08	11.05	3.37	SIGLEC17P	2.44E-08	7.20	1.00
CD300A	2.95E-05	10.72	6.03	SLC1A7	0.000146564	6.98	0.63
HOPX	0.000263012	10.70	6.52	SIGLEC7	3.36E-07	6.79	0.34
CX3CR1	7.81E-06	10.63	1.76	LINGO2	1.82E-06	6.69	0.34
TRDC	8.28E-13	10.60	1.63	TRDJ1	8.20E-13	6.54	0.00
PRSS23	1.86E-10	10.59	1.69	COL13A1	0.000155986	6.32	0.34
NCAM1	1.57E-12	10.42	2.02	CYBB	0.007079707	6.20	12.16
ATP8B4	1.16E-07	10.17	2.75	MPEG1	0.031777488	5.05	10.72

TYROBP	1.72E-07	10.06	4.23	HCK	0.009445927	4.38	10.03
FCRL6	6.09E-10	10.06	1.57	SPI1	0.049981251	4.05	9.01
PTGDR	5.40E-10	10.02	2.00	CD86	0.046443612	3.96	8.91
LINC00299	1.20E-10	9.93	1.84	CCR7	0.000103201	3.91	9.99
BNC2	1.71E-10	9.89	2.07	TNS3	0.000213329	3.58	10.32
NCR1	1.33E-12	9.87	2.10	CD1D	0.046105377	3.05	8.17
MLC1	9.86E-08	9.74	2.78	SLAMF1	0.000472038	2.69	7.46
PDGFD	2.74E-08	9.51	3.83	AXIN2	0.039476765	2.03	5.26
RASSF4	1.08E-05	9.37	4.50	DGAT2	0.013581184	1.99	6.83
NMUR1	8.41E-13	9.35	1.13	SIT1	1.71E-09	1.36	10.13
FAT4	8.47E-13	9.15	1.26	FAAH2	0.007946423	0.71	3.23
SLCO4C1	6.11E-11	9.07	1.82				

2.3 Genes differentially expressed between NK and monocytes. Genes higher expressed in NK cells are marked in green.

		Expression levels				Expression levels	
GeneID	**p Value**	**NK**	**MC**	**GeneID**	**p Value**	**NK**	**MC**
GNLY	2.31E-10	14.27	5.53	LGR6	3.67E-10	8.68	0.44
PRF1	4.51E-11	14.18	4.21	FEZ1	5.72E-06	8.51	2.48
NKG7	4.02E-06	13.82	7.47	L3MBTL4	0.000628966	8.42	3.13
KLRD1	4.68E-10	12.74	2.98	RNF165	2.37E-07	8.35	3.19
GZMB	9.21E-12	12.56	1.98	KIR3DX1	3.48E-10	8.11	0.00
FGFBP2	4.57E-09	12.13	1.84	TRGV9	2.88E-11	7.90	1.23
CST7	9.79E-09	12.09	5.86	PDGFRB	0.00960803	7.88	4.43
S1PR5	4.38E-10	11.93	2.73	B3GAT1	8.92E-07	7.79	0.00
GZMA	1.60E-11	11.81	2.06	PCDH1	3.59E-11	7.68	0.00
KLRF1	5.43E-12	11.58	1.34	PRSS30P	1.27E-06	7.44	0.73
TTC38	9.44E-06	11.47	8.07	SIGLEC17P	6.73E-05	7.20	3.04
SH2D1B	9.07E-11	11.35	2.37	S100A9	0.005111889	7.14	15.46
GZMH	3.02E-09	11.32	1.34	SLC1A7	3.27E-05	6.98	0.00
ADGRG1	1.64E-07	11.05	4.15	LINGO2	6.31E-07	6.69	0.00
HOPX	3.13E-12	10.70	1.10	TRDJ1	8.20E-13	6.54	0.00
TRDC	8.17E-13	10.60	0.44	COL13A1	6.70E-05	6.32	0.00
PRSS23	1.42E-10	10.59	1.60	CYBB	0.000289019	6.20	14.04
NCAM1	1.39E-12	10.42	1.94	FCN1	0.001716661	5.68	14.99
SYK	0.036299499	10.18	12.57	TNFAIP2	0.001055934	5.36	14.37
TYROBP	0.02847195	10.06	12.55	MPEG1	0.002874598	5.05	12.48
FCRL6	4.16E-08	10.06	2.88	SERPINA1	0.000392649	4.76	13.47
PTGDR	2.74E-09	10.02	2.48	HCK	9.97E-05	4.38	12.62
LINC00299	1.37E-09	9.93	2.54	SPI1	0.000333274	4.05	12.34
BNC2	1.27E-10	9.89	1.98	CD86	0.00034272	3.96	12.13
NCR1	2.20E-10	9.87	3.57	TNS3	9.86E-05	3.58	10.68
B3GNT7	5.01E-07	9.83	2.48	RBM47	0.000203469	3.56	11.32
MLC1	0.007189711	9.74	6.32	CFD	0.000147124	3.44	11.59
PDGFD	8.35E-13	9.51	0.00	ANPEP	0.000166772	3.30	11.87

NMUR1	8.25E-13	9.35	0.73	JDP2	0.000237226	3.06	10.80
FAT4	8.46E-13	9.15	1.25	CD1D	0.000554689	3.05	11.18
SLCO4C1	0.000225536	9.07	5.50	THBD	0.002397956	2.43	10.54
CLIC3	1.62E-07	8.97	2.35	DGAT2	5.58E-05	1.99	9.63
AC017104.6	5.16E-08	8.88	1.98	ALS2CL	0.000105678	1.40	6.16

2.3 Genes differentially expressed between T lymphocytes and monocytes. Genes with higher expression in T lymphocytes are marked in green.

		Expression levels				Expression levels	
GeneID	**p Value**	**T**	**MC**	**GeneID**	**p Value**	**T**	**MC**
AXIN2	2.83E-06	8.56	1.25	CYBB	1.23E-05	3.98	14.04
CACNA1I	8.33E-10	8.05	0.00	DAB2	1.03E-05	0.94	6.45
CAMK4	3.05E-11	12.80	3.38	DGAT2	3.97E-05	1.42	9.63
CCR4	1.19E-06	9.08	1.54	FCGR3A	0.000193193	2.88	8.78
CCR7	4.90E-05	11.54	4.83	FCN1	0.000182095	3.36	14.99
CD28	2.22E-09	12.34	1.91	FGR	2.12E-11	4.16	13.52
CD5	0.000298924	11.65	5.54	GPR141	5.65E-07	0.83	7.29
CTC-499J9.1	8.93E-13	6.94	0.00	HCK	2.78E-06	1.98	12.62
FAAH2	4.19E-11	7.40	0.00	ITGAM	6.55E-12	5.32	13.02
GZMA	0.043966888	4.92	2.06	JDP2	0.0001538	2.41	10.80
IL7R	1.52E-07	13.59	1.75	MLC1	0.000282552	1.58	6.32
KLRF1	0.006778523	4.99	1.34	MPEG1	0.00031325	3.13	12.48
LEF1-AS1	9.97E-13	8.48	0.00	NME8	2.93E-05	0.73	5.48
MAL	6.20E-12	8.72	0.00	PDGFRB	0.004413248	0.50	4.43
NELL2	1.52E-06	11.97	2.34	RASSF4	4.79E-10	2.71	10.71
PRF1	0.012864422	7.78	4.21	RBM47	3.96E-05	2.25	11.32
SIT1	1.49E-07	8.12	0.44	S100A9	0.000440521	4.66	15.46
SLAMF1	0.006804697	9.02	5.05	SERPINA 1	1.63E-05	2.22	13.47
SLC16A10	0.000303481	8.58	3.27	SIGLEC17 P	0.016930394	0.30	3.04
TRAT1	1.04E-07	10.53	0.00	SIGLEC7	2.99E-07	0.24	7.07
TRDC	0.004786652	3.12	0.44	SLCO4C1	0.000607553	2.01	5.50
ANPEP	5.47E-05	2.18	11.87	SPI1	8.11E-05	2.74	12.34
ATP8B4	4.73E-06	2.43	8.96	SYK	3.88E-12	3.29	12.57
CD1D	0.000104203	1.56	11.18	THBD	0.001983222	1.86	10.54
CD300A	2.96E-06	4.30	9.91	TNFAIP2	0.000217891	3.69	14.37
CD86	0.000149182	3.02	12.13	TNS3	5.44E-06	1.82	10.68
CFD	9.88E-06	1.49	11.59	TTC38	5.35E-05	4.83	8.07
CXCR2	1.54E-05	0.85	5.90	TYROBP	1.84E-12	2.88	12.55

2.4 Genes differentially expressed between T and B lymphocytes. Genes with higher expression in T lymphocytes are marked in green.

GeneID	p Value	T	B	GeneID	p Value	T	B
ALS2CL	2.17E-06	7.80	1.56	B3GNT7	0.000107051	1.97	7.73

AXIN2	0.048211056	8.56	5.26	CD1D	0.009631394	1.56	8.17
CACNA1I	5.78E-08	8.05	1.26	CD86	0.019394662	3.02	8.91
CAMK4	8.17E-08	12.80	5.87	CYBB	0.000325067	3.98	12.16
CCR4	3.32E-06	9.08	1.92	DGAT2	0.007955644	1.42	6.83
CD28	3.82E-09	12.34	2.13	FGR	3.73E-07	4.16	10.51
CD5	0.005121458	11.65	6.83	HCK	0.000291421	1.98	10.03
CTC-499J9.1	1.15E-12	6.94	0.34	HOPX	0.013661026	3.43	6.52
FAAH2	2.18E-05	7.40	3.23	ITGAM	2.53E-06	5.32	9.87
IL7R	0.000173187	13.59	5.50	JDP2	0.040610676	2.41	7.32
LEF1-AS1	8.29E-12	8.48	1.00	MPEG1	0.003950965	3.13	10.72
MAL	2.67E-08	8.72	2.38	PDGFRB	0.00585475	0.50	4.33
NELL2	3.08E-07	11.97	1.57	RBM47	0.036521636	2.25	7.19
SLC16A10	0.003422509	8.58	4.23	SPI1	0.012959867	2.74	9.01
TRAT1	2.03E-06	10.53	1.38	SYK	7.50E-12	3.29	12.33
AC017104.6	1.06E-07	1.07	8.11	TNS3	1.16E-05	1.82	10.32

2.5 Genes differentially expressed between B lymphocytes and monocytes. Genes with higher expression in B lymphocytes are marked in green.

		Expression levels				Expression levels	
GeneID	**p Value**	**B**	**MC**	**GeneID**	**p Value**	**B**	**MC**
AC017104.6	0.000183343	8.11	1.98	CXCR2	0.001110193	0.97	5.90
B3GNT7	0.008095343	7.73	2.48	FCGR3A	0.014748615	3.56	8.78
CCR7	0.022474728	9.99	4.83	GPR141	2.12E-05	0.34	7.29
FAAH2	0.014169341	3.23	0.00	NME8	0.009836482	1.57	5.48
HOPX	0.000550519	6.52	1.10	RASSF4	3.67E-05	4.50	10.71
ITGAM	0.014035438	9.87	13.02	SERPINA 1	0.016392762	5.04	13.47
PDGFD	0.004142158	3.83	0.00	SIGLEC7	5.60E-05	0.34	7.07
SIT1	2.89E-07	10.13	0.44	SLCO4C1	0.006585699	1.82	5.50
ALS2CL	0.007455837	1.56	6.16	THBD	0.038578435	2.44	10.54
ATP8B4	0.000674948	2.75	8.96	TNFAIP2	0.026581098	5.52	14.37
CD300A	0.014802003	6.03	9.91	TYROBP	6.82E-08	4.23	12.55
CX3CR1	0.029098815	1.76	7.99				

Appendix 3 Genes differentially expressed between ex vivo NK developmental stages identified using ANOVA analysis

Expression levels are indicated in Log_2 domain.

3.1 Genes differentially expressed between stage 1 and stage 2 *ex vivo* NK samples. All identified genes are higher expressed by stage 1 samples.

GeneID	**p Value**	**Expression levels**		**GeneID**	**p Value**	**Expression levels**	
		Stage 1	**Stage 2**			**Stage 1**	**Stage 2**
CA2	0.008581077	11.98	7.36	PDZD4	0.01072162	3.95	0.95
CD8A	0.003595695	5.65	2.19	TGFBR3	0.007657388	6	3.64

3.2. Genes differentially expressed between stage 2 and stage 3 *ex vivo* NK samples. Genes with higher expression in stage 2 samples are marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Stage 2	Stage 3			Stage 2	Stage 3
ADGRG6	0.014087346	9.35	6.30	IL7R	0.000242399	6.10	12.98
CPA3	0.040783765	10.98	6.90	KIAA1211L	3.87E-05	1.30	7.48
HBD	0.04314571	7.75	2.73	KIAA1324	1.35E-08	1.53	9.46
MAP7	0.04423324	9.77	7.83	KRT81	0.005183573	1.53	6.43
MPO	0.024955782	12.96	5.48	KRT86	0.001308175	1.29	7.29
PRSS23	0.011445942	3.81	1.40	LIF	0.000665343	4.25	9.78
PRSS57	0.001368987	10.58	5.01	LINC00892	4.51E-05	1.20	6.74
SPINK2	0.009062321	10.28	7.79	LINGO4	7.56E-05	2.65	8.15
TAL1	0.016399197	9.23	4.68	PDZD4	4.42E-05	0.95	3.99
AL450992.2	0.005320224	2.05	6.45	PKDCC	3.46E-05	2.21	7.89
B3GALT5	1.35E-06	0.81	6.95	RORC	0.000155328	3.71	10.72
CD2	0.000343415	3.79	8.97	RP11-330A16.1	0.003430403	2.23	5.80
CSPG4	0.003126537	1.96	5.46	RP5-1028K7.2	0.011937266	0.75	5.20
DLL1	0.005091447	4.53	9.16	SLC4A10	0.000151548	0.68	7.75
ENPP1	3.01E-06	2.47	9.37	SMIM10L2A	0.001883157	1.51	5.96
EOMES	0.011641871	0.81	3.02	TGFBR3	0.000278677	3.64	5.71
GNLY	0.001305565	5.94	9.04	THEM5	0.006462324	1.80	5.89
GPR55	0.00898273	2.17	6.72	TNFRSF18	4.45E-05	3.89	9.13
GPR68	0.001166204	5.05	9.27	TNFSF11	4.60E-05	1.97	8.80
IGFBP4	0.045589771	3.77	7.02	TOX2	0.002858866	3.60	8.19
IL1R1	0.00270958	5.22	10.39	XCL1	0.000106199	1.61	8.08
IL23R	0.000206492	2.01	8.95	XCL2	0.000116576	1.70	7.03
IL4I1	0.001311971	3.73	9.17				

3.3 Genes differentially expressed between stage 3 and stage 4 *ex vivo* NK samples. Genes with higher expression in stage 3 samples are marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Stage 3	Stage 4			Stage 3	Stage 4
ADGRG6	6.41E-06	6.30	2.27	BNC2	2.22E-05	1.81	4.93
AKR1C3	0.029933829	5.93	3.16	CD8A	0	3.14	8.20
AL450992.2	0.044095866	6.45	3.78	DTHD1	0	1.05	8.28
B3GALT5	1.05E-09	6.95	0.69	EOMES	0	3.02	9.34
CA2	5.45E-07	9.23	4.87	FCGR3A	0	1.95	8.53
CDC42BPA	1.18E-06	8.00	3.99	FCRL3	0.00435096	4.44	7.92
CDH1	2.06E-07	8.24	2.81	FCRL6	1.49E-08	3.25	8.04
CPA3	0.001761078	6.90	2.57	FEZ1	2.94E-06	0.74	4.15
EREG	0.021786606	9.71	5.82	FGFBP2	0	1.05	7.36
FLT3	7.82E-09	9.16	3.33	GNLY	0	9.04	14.99

GRB10	1.55E-07	8.07	3.84	GZMA	0	3.03	10.29
IL1R1	0.000149211	10.39	5.49	GZMB	0	4.85	10.30
IL23R	0.00354844	8.95	4.62	GZMH	0	0.98	9.01
IL4I1	0.002282303	9.17	5.16	GZMM	0	1.79	9.47
KIAA1211L	0.001281245	7.48	3.69	IFNG	0	1.17	8.11
KIAA1324	1.80E-05	9.46	5.01	KIR3DX1	0	1.07	7.19
KIT	1.66E-06	12.49	8.88	KLRD1	0	4.31	12.12
LIF	0.002490478	9.78	5.93	KLRF1	0	2.87	11.15
LINGO4	4.17E-10	8.15	1.04	PDZD4	0	3.99	8.47
MAP7	5.16E-08	7.83	4.19	PRF1	1.36E-11	6.73	12.49
MRC2	1.16E-08	9.04	4.18	PRSS23	4.04E-09	1.40	5.63
MYB	3.22E-08	9.17	4.74	PYHIN1	0	2.92	8.65
PKDCC	4.40E-08	7.89	1.96	RP11-222K16.2	0	0.37	7.07
RORC	1.04E-05	10.72	4.44	S1PR5	0	0.79	8.40
STAC	0	7.91	2.73	TBX21	0	3.88	10.52
TGM2	8.13E-07	9.13	3.41	TENM1	0	2.58	8.89
VWDE	4.43E-08	7.69	3.29	TGFBR3	0	5.71	11.00
ACAN	0.002163214	2.82	6.18	XCL2	0.017352484	7.03	9.77

3.4 Genes differentially expressed between stage 4 and stage 5 *ex vivo* NK samples. Genes with higher expression in stage 4 samples are marked in green.

		Expression levels				Expression levels	
GeneID	p Value	Stage 4	Stage 5	GeneID	p Value	Stage 4	Stage 5
ACAN	2.58E-06	6.18	1.92	STAC	1.78E-05	2.73	0.40
CA2	0.019618121	4.87	3.02	TNFSF11	0.000349946	7.48	3.56
CDC42BPA	0.035914272	3.99	2.36	TOX2	0.000167189	6.82	3.16
CSPG4	1.31E-07	4.83	0.81	VWDE	0.000384224	3.29	0.92
DLL1	0.000257419	9.23	5.44	XCL1	9.93E-05	10.71	6.49
ENPP1	1.12E-10	8.00	1.35	XCL2	0.018317172	9.77	7.47
GNAI1	0	6.19	1.05	AKR1C3	6.02E-08	3.16	8.54
GPR55	0.000555762	4.91	1.18	B3GAT1	0	1.78	7.79
GRB10	0.007817987	3.84	1.96	BNC2	0	4.93	9.89
IGFBP4	1.65E-10	8.81	2.47	CD8A	0.00046797	8.20	9.97
IL1R1	0.016830891	5.49	2.67	DTHD1	5.74E-07	8.28	10.67
IL23R	0.024566706	4.62	1.62	FCGR3A	1.31E-06	8.53	11.60
IL7R	2.11E-06	11.65	5.85	FCRL6	0.005243711	8.04	10.06
KIT	7.81E-12	8.88	4.17	FEZ1	5.05E-11	4.15	8.51
KRT81	0.016064946	5.09	2.25	FGFBP2	0	7.36	12.13
LIF	5.20E-05	5.93	1.70	GZMA	0.04406799	10.29	11.81
LINC00892	4.30E-05	5.42	1.81	GZMB	0.000106211	10.30	12.56
LPAR1	5.30E-08	6.80	1.48	GZMH	0.000228603	9.01	11.32
MAP7	4.68E-07	4.19	1.38	GZMM	0.004292417	9.47	10.84
MRC2	0.012626141	4.18	2.32	IFNG	0.001285718	8.11	9.86
MYB	0.000233402	4.74	2.31	LGR6	0	2.37	8.68

PRSS23	0	5.63	10.59	PDZD4	5.73E-05	8.47	10.42
RORC	0.026651162	4.44	1.49	PRF1	0.019104451	12.49	14.18
RP11-330A16.1	6.77E-11	5.95	0.54	PYHIN1	7.26E-05	8.65	10.65
SLC4A10	0.013550666	7.23	4.02	S1PR5	3.16E-12	8.40	11.93
SMIM10L2A	9.28E-05	4.53	0.97	SLC1A7	8.04E-11	1.11	6.98
SPINK2	0	6.66	0.72	TGFBR3	2.67E-05	11.00	12.55

Appendix 4 Ranked top 100 genes differentially expressed within ex vivo stage 4 NK samples

Genes identified from the total number of 20160 genes using LoD1.

GeneID	Score (PC1-3)	Rank	GeneID	Score (PC1-3)	Rank
HLA-U	0.044872866	1	CPA1	0.028576064	51
PPBP	0.042793137	2	MMP8	0.02856053	52
HBG1	0.041395926	3	RN7SKP291	0.028503449	53
HBG2	0.041259942	4	S100A12	0.028286044	54
CTSG	0.040896323	5	ABCD2	0.028237051	55
RPSAP53	0.040235825	6	TCL1A	0.028144963	56
IGHG1	0.040112401	7	RP5-857K21.11	0.028035752	57
ELANE	0.039878047	8	CENPQ	0.027890947	58
RP11-380G5.2	0.038727222	9	PGLYRP1	0.027885592	59
DEFA4	0.037676366	10	KIF21A	0.027829581	60
IGHG2	0.037580756	11	CXCL8	0.027425628	61
LTF	0.036850387	12	IGLC3	0.027378225	62
PRTN3	0.036692085	13	CPB1	0.027368866	63
COLGALT2	0.036618122	14	AZU1	0.027278489	64
MTRNR2L1	0.035658654	15	TLR10	0.027253323	65
CELA1	0.035132499	16	MTRNR2L12	0.027102442	66
PF4	0.034670961	17	RP11-263C24.1	0.027003067	67
S100A8	0.03419542	18	TAS2R64P	0.026991588	68
FCRL5	0.033676034	19	CRISP3	0.026955824	69
BIRC6-AS1	0.033640429	20	LYZ	0.026891252	70
DEFA3	0.033612078	21	FCRL2	0.026886044	71
CAMP	0.033589194	22	CHI3L1	0.026870583	72
IGHA1	0.033359998	23	SLC22A20	0.02676414	73
MPO	0.033142477	24	GP9	0.02674691	74
RNASE3	0.032762949	25	RP11-726G1.1	0.026689902	75
IGHG3	0.032494735	26	CYTL1	0.026634354	76
BPI	0.032218286	27	MTRNR2L10	0.026615115	77
MIR4742	0.032112952	28	MIR1255A	0.026588787	78
ITGA1	0.032095342	29	SLPI	0.026578227	79
RNU2-64P	0.032058646	30	TBC1D9	0.026558456	80
MDGA1	0.031874738	31	IFIT1B	0.026538442	81
IGHG4	0.031872443	32	GAPT	0.026520743	82

IGKV3-20	0.031693363	33	TNFAIP8L2	0.026509954	83
IGLV3-1	0.031495208	34	FCRLA	0.026485506	84
ITGB3	0.031253428	35	IGKV4-1	0.026483515	85
IGHV1-69	0.030969613	36	RP3-465N24.5	0.026481947	86
OPA1-AS1	0.030330073	37	MS4A3	0.026453607	87
CEACAM8	0.030115459	38	IGHV3-15	0.026358545	88
AC079949.1	0.030039327	39	IGLV3-19	0.026255917	89
LCN2	0.029988944	40	RP11-319G9.3	0.02623054	90
IGLV2-23	0.029916983	41	RN7SKP137	0.0262097	91
ABCA13	0.029850262	42	RP11-286O18.1	0.026191232	92
RNASE2	0.029780793	43	CPA2	0.026183614	93
RP4-591B8.2	0.029394141	44	ADAM19	0.026148029	94
CR2	0.029237168	45	RP11-225N10.3	0.026123355	95
JCHAIN	0.028958795	46	RP11-513M16.7	0.026107061	96
IGKV1-5	0.02877502	47	SLC2A5	0.026048112	97
S100A9	0.028721833	48	RN7SL753P	0.025948049	98
IGHGP	0.028692624	49	RP11-394O4.5	0.025946907	99
PNLIPRP1	0.028650991	50	FCN1	0.025789268	100

Appendix 5 Genes differentially expressed between ex vivo stage 4 NK samples of different tissue origin identified using ANOVA analysis

Expression levels are indicated in Log_2 domain.

5.1 Genes differentially expressed between PB- and BM-derived stage 4 NK samples. All identified genes are higher expressed by BM-derived samples.

		Expression levels				Expression levels	
GeneID	**p Value**	**BM**	**PB**	**GeneID**	**p Value**	**BM**	**PB**
ABCA13	0.040775111	10.00	0.64	IGLC3	0.008147998	8.67	0.54
AZU1	0.03521059	12.13	3.58	IGLV2-23	0.007418155	9.31	0.00
BPI	0.002510486	10.46	0.00	JCHAIN	0.036939929	10.06	1.61
CAMP	0.004808829	11.12	0.40	LCN2	0.030612573	11.04	1.77
CEACAM8	0.006281595	9.77	0.00	LTF	0.009246428	13.25	1.62
CHI3L1	0.016231377	8.18	0.54	LYZ	0.030318717	13.16	5.37
CRISP3	0.004488564	9.37	0.00	MMP8	0.019786435	9.98	0.20
CTSG	0.009180755	12.44	0.54	MPO	0.030080094	13.61	2.44
DEFA3	0.022124486	11.30	0.70	MS4A3	0.028057495	9.74	2.01
DEFA4	0.002021867	11.55	0.73	PGLYRP1	0.030652781	8.73	0.64
ELANE	0.012103025	12.19	0.00	PLBD1	0.01528074	9.35	1.77
IGHA1	0.007347146	10.59	1.40	PRTN3	0.027559544	12.13	0.67
IGHA2	0.000142033	8.11	0.20	RETN	0.015686867	8.04	0.54
IGHG1	0.007268767	13.01	1.02	RNASE2	0.005186584	9.01	0.20
IGHG3	0.005615043	10.87	0.00	RNASE3	0.002182448	9.66	0.20
IGHGP	0.00213222	9.39	0.00	S100A12	0.028312931	9.92	1.57
IGHV3-23	0.011206482	8.13	0.20	S100A9	0.03551124	13.93	5.70

IGKV1-5	0.041200739	9.17	1.02	SLC2A5	0.007586502	7.91	0.54
IGLC2	0.027201772	8.59	0.80	SLPI	0.00935506	7.87	0.00

5.2 Genes differentially expressed between BM- and tonsil-derived stage 4 NK samples. Genes with higher expression in BM-derived samples are marked in green

		Expression levels				Expression levels	
GeneID	**p Value**	**BM**	**Tonsils**	**GeneID**	**p Value**	**BM**	**Tonsils**
BPI	0.003804809	10.46	0.00	RETN	0.017197016	8.04	0.00
CAMP	0.016920463	11.12	2.35	RNASE2	0.022826538	9.01	2.20
CEACAM8	0.013070427	9.77	0.74	RNASE3	0.008358835	9.66	1.98
CRISP3	0.006732058	9.37	0.00	S100A12	0.032073791	9.92	1.06
DEFA3	0.024648072	11.30	0.00	S100A9	0.021420888	13.93	3.68
DEFA4	0.003370746	11.55	0.95	SLPI	0.041410376	7.87	1.90
FCN1	0.043941761	10.12	0.00	CR2	0.01843573	0.00	8.36
LTF	0.023714767	13.25	3.09	ITGA1	0.021789257	5.27	11.42
MMP8	0.026281099	9.98	0.00				

5.3 Genes differentially expressed between BM- and CB-derived stage 4 NK samples. All identified genes are higher expressed in BM-derived samples.

		Expression levels				Expression levels	
GeneID	**p Value**	**BM**	**CB**	**GeneID**	**p Value**	**BM**	**CB**
AZU1	0.045469492	12.13	3.38	IGHV3-23	0.014803247	8.13	0.00
BPI	0.010797969	10.46	2.20	IGKV1-5	0.038088192	9.17	0.06
CAMP	0.038327985	11.12	3.98	IGLC2	0.026430575	8.59	0.00
CEACAM8	0.029997395	9.77	2.41	IGLC3	0.0094361	8.67	0.06
CTSG	0.013021856	12.44	0.42	IGLV2-23	0.01100207	9.31	0.00
DEFA4	0.004014637	11.55	1.35	MS4A3	0.041255488	9.74	2.09
ELANE	0.030275333	12.19	1.53	RNASE2	0.009809633	9.01	0.67
IGHA1	0.007040331	10.59	0.42	RNASE3	0.003090902	9.66	0.06
IGHA2	0.000256057	8.11	0.42	RPSAP53	0.028849284	8.23	0.00
IGHG1	0.011762502	13.01	1.26	SLC2A5	0.015613447	7.91	1.09
IGHG3	0.008381973	10.87	0.00	SLPI	0.021009958	7.87	0.76
IGHGP	0.003239692	9.39	0.00				

5.4 Genes differentially expressed between CB- and tonsil-derived stage 4 NK samples. All identified genes are higher expressed in tonsil-derived samples.

		Expression levels				Expression levels	
GeneID	**p Value**	**CB**	**Tonsils**	**GeneID**	**p Value**	**CB**	**Tonsils**
CR2	0.036908257	1.35	8.36	IGHGP	0.030686376	0.00	5.54
IGHA1	0.023729829	0.42	8.04	IGHV3-23	0.022368078	0.00	7.36
IGHA2	0.001257025	0.42	5.94	IGLC3	0.032341008	0.06	6.44
IGHG1	0.030808707	1.26	10.54	ITGA1	0.007386238	3.47	11.42
IGHG3	0.018994393	0.00	8.95				

5.5 Genes differentially expressed between PB- and tonsil-derived stage 4 NK samples. All identified genes are higher expressed in tonsil-derived samples.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		PB	Tonsils			PB	Tonsils
CR2	0.023055852	1.14	8.36	IGHV3-23	0.017286332	0.20	7.36
IGHA1	0.028465654	1.40	8.04	IGLC3	0.03074734	0.54	6.44
IGHA2	0.000679536	0.20	5.94	ITGA1	0.002921869	2.49	11.42
IGHG1	0.019333371	1.02	10.54	SLC2A5	0.039414738	0.54	5.47
IGHG3	0.01300668	0.00	8.95	TLR10	0.041480249	1.21	7.26
IGHGP	0.021370757	0.00	5.54				

5.6 Genes differentially expressed between PB- and CB-derived stage 4 NK samples. The only identified gene is higher expressed in PB-derived samples.

GeneID	p Value	Expression levels	
		CB	PB
RPSAP53	0.014617274	0.00	8.88

Appendix 6 Genes differentially expressed between BM-derived and non-BM derived stage 4 NK samples identified among top 400 genes using ANOVA analysis

Expression levels are indicated in Log_2 domain.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		CB	PB			CB	PB
ABCA13	0.006997875	10.00	2.18	LPCAT2	0.012541798	7.57	2.69
ANXA3	0.003571723	8.33	2.02	LTF	0.003082246	13.25	3.42
AZU1	0.000980616	12.13	3.60	LYZ	0.006791651	13.16	6.73
BEX1	0.000858393	5.99	0.78	MCEMP1	0.00097922	6.27	0.79
BPI	0.000133552	10.46	0.63	MGST1	0.00012882	6.11	0.64
CAMP	0.001083936	11.12	1.98	MMP8	0.002840424	9.98	1.09
CD177	0.009211566	7.15	1.11	MNDA	0.00343195	8.41	2.33
CEACAM6	4.82E-05	8.95	1.33	MPO	0.003409805	13.61	3.91
CEACAM8	0.000440334	9.77	0.90	MS4A3	0.000812184	9.74	2.19
CHI3L1	0.003755841	8.18	1.78	OLFM4	0.003013909	8.15	0.87
CHIT1	0.002575087	7.63	1.79	PADI2	0.00555747	7.30	1.95
CLC	0.002836451	6.66	0.87	PGLYRP1	0.002404518	8.73	1.45
CLEC4D	0.000468154	6.69	0.70	PLBD1	0.004602579	9.35	3.13
CLEC5A	0.000225758	6.51	0.83	PRRT4	5.81E-07	6.57	0.19
CRISP3	0.002570828	9.37	1.15	PRTN3	0.002778924	12.13	1.98
CSF3R	0.022993957	9.24	4.19	RAB44	5.36E-05	6.91	0.62
CSTA	0.003757188	7.51	2.21	RETN	0.000863251	8.04	0.82

CTSG	0.001571086	12.44	1.66	RNASE2	0.000295559	9.01	0.91
CYP4F3	0.000587681	8.15	1.19	RNASE3	9.08E-05	9.66	0.67
DEFA3	0.001465622	11.30	1.19	S100A12	0.013417495	9.92	2.67
DEFA4	1.90E-05	11.55	0.97	S100A8	0.011079726	13.23	4.45
ELANE	0.0013738	12.19	1.47	S100A9	0.007013196	13.93	5.95
FCAR	0.009551089	8.22	2.27	S100P	0.001243559	9.16	2.96
FCN1	0.02213523	10.12	2.96	SERPINB 10	0.000448408	7.45	0.84
GLT1D1	0.002676796	6.71	1.74	SIRPB1	0.009921931	5.97	1.48
H1F0	0.001628369	8.81	3.70	SLC2A5	0.020687636	7.91	2.10
HK3	0.010539678	8.12	1.96	SLPI	0.000638316	7.87	0.76
IGHA1	0.031039151	10.59	3.02	SPI1	0.041546652	7.77	2.95
IGHA2	0.020810183	8.11	1.90	SPP1	0.004740411	5.13	0.21
IGHG1	0.042965989	13.01	3.81	TCN1	0.002764109	7.89	0.89
IGHG3	0.048660702	10.87	2.56	VSTM1	6.74E-06	7.25	0.16
IGHGP	0.00931305	9.39	1.58	CTD-2228K2.7	0.007793918	2.07	6.44
IGHV3-21	0.027836493	7.43	1.76	CYCSP40	0.008757902	0.00	4.70
IGKV1-5	0.029075362	9.17	2.21	FAM129C	0.010387984	2.57	7.24
IGKV1-9	0.001005592	7.28	0.82	KCND3	0.031720814	0.85	4.11
IGKV3-11	0.018213792	8.03	1.46	LRRC48	0.017006933	2.33	6.19
IGLC2	0.027538621	8.59	2.01	MFAP1	0.03774401	3.63	7.72
IGLC3	0.031006099	8.67	2.09	PHKG1	0.046158696	2.71	6.48
IGLV2-11	0.015224378	5.84	1.23	RP11-533E19.3	0.0157923	1.53	5.47
IGLV2-23	0.00964024	9.31	1.49	RP11-567M16.5	0.041374679	1.45	5.46
JDP2	0.008172979	7.08	2.50	RP11-701H24.4	0.038097903	3.84	8.22
KCNH2	0.006150404	6.84	1.53	RP11-702F3.1	0.00184513	0.00	4.61
LCN2	0.001579508	11.04	2.49	RP11-802E16.3	0.02451216	2.77	6.77
LIN7A	0.003364205	6.82	1.75	UBAP1L	0.016958663	3.05	7.46

Appendix 7 Ranked top 100 genes differentially expressed within ex vivo stage 5 NK samples

Genes identified from the total number of 21903 genes using LoD1.

GeneID	Score (PC1-3)	Rank	GeneID	Score (PC1-3)	Rank
CXCL8	0.048725036	1	SPI1	0.029533229	51
G0S2	0.043684424	2	MAFB	0.029496721	52
HBA1	0.042841837	3	ANPEP	0.029383244	53
HBA2	0.041642666	4	ALAS2	0.029269775	54
HBB	0.040229567	5	LTF	0.029251677	55
THBD	0.040130548	6	RNU1-120P	0.029196621	56
AQP9	0.039567551	7	SLC4A1	0.029129114	57
S100A9	0.039030113	8	DGAT2	0.029064446	58

FCAR	0.038887971	9	LILRA5	0.028954611	59
IL1B	0.038283023	10	FAM20C	0.028944054	60
FCN1	0.037708582	11	IL31RA	0.02882478	61
TSIX	0.037573964	12	CD1D	0.028817054	62
C5AR1	0.037082444	13	CYP27A1	0.028736263	63
ARSI	0.036744692	14	MNDA	0.028558618	64
EREG	0.036421824	15	RNU1-59P	0.02851339	65
MPEG1	0.036171307	16	TBC1D9	0.028397372	66
SEMA6B	0.035340882	17	CXCL3	0.028318437	67
HBEGF	0.03524662	18	RAB20	0.028281443	68
SERPINB2	0.034846779	19	SERPINA1	0.028269283	69
TNFAIP2	0.034834576	20	CTSL	0.028144706	70
ADM	0.034558569	21	CTD-2303H24.2	0.028024837	71
IL1RN	0.034497212	22	VCAN	0.027997839	72
FPR1	0.034276824	23	JDP2	0.027793213	73
CXCL2	0.034252954	24	CD14	0.027770132	74
TREM1	0.034020456	25	ZNF683	0.027607461	75
CTB-61M7.2	0.033906973	26	EGR3	0.02756121	76
CLEC7A	0.03379109	27	MS4A6A	0.027448848	77
FFAR2	0.033624425	28	CD36	0.027372058	78
KIR3DL1	0.03342716	29	ALPL	0.027348023	79
LRRK2	0.032874463	30	HCAR3	0.027243408	80
GPR84	0.032622855	31	WDFY3	0.027230761	81
LYZ	0.032501323	32	CD300E	0.027209143	82
CSF3R	0.032489089	33	RP11-443B7.3	0.027123718	83
TRIB1	0.032471461	34	PLAU	0.026949121	84
APOBEC3A	0.032241561	35	BASP1	0.026769882	85
GNA15	0.032155158	36	SGK1	0.026750555	86
RNU1-34P	0.032102292	37	MGLL	0.026717902	87
DUSP4	0.031772339	38	VDR	0.026706643	88
S100A12	0.031687864	39	RBM47	0.026696244	89
TGFBI	0.031597074	40	PID1	0.026660303	90
S100A8	0.031259906	41	ICOSLG	0.026625361	91
PTGS2	0.031232339	42	OLR1	0.026599181	92
CAMK2A	0.031168223	43	RNU5F-1	0.026594636	93
CXCL16	0.030985889	44	ZMAT4	0.026584874	94
SDC2	0.030640419	45	CTB-114C7.4	0.026560423	95
SIRPA	0.030564235	46	SELENBP1	0.026524697	96
PTX3	0.030388862	47	MME	0.026432205	97
HCAR2	0.030237623	48	DEFA3	0.026354172	98
LRG1	0.030138753	49	PTGES	0.026346759	99
SLC8A1	0.029844379	50	BOK	0.026224765	100

Appendix 8 Genes differentially expressed between nonlicensed, licensed and memory-like ex vivo stage 5 NK populations identified from top 100 genes using ANOVA analysis

Expression levels are indicated in Log_2 domain.

8.1 Genes differentially expressed between nonlicensed and licensed samples. All identified genes are higher expressed in nonlicensed samples.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Non-licensed	Licensed			Non-licensed	Licensed
AMICA1	0.017310103	7.89	4.50	MNDA	0.036935536	5.99	3.01
CD4	0.018416018	6.41	3.11	MS4A6A	0.012419481	6.34	3.01
FAM20C	0.012196882	5.32	1.37	NAPSB	0.015336566	6.91	3.33
GNA15	0.046179566	6.16	2.83	PLD4	0.010970554	6.01	2.67
HCK	0.026177292	6.64	3.42	PTGS2	0.029512842	5.84	2.32
IL1B	0.028955093	7.68	2.94	RP11-443B7.3	0.040515378	4.22	1.18
IL1RN	0.042781891	5.02	1.51	SLC8A1	0.036845802	6.79	3.64
JDP2	0.027446496	4.83	1.73	SPI1	0.037400647	6.38	2.76
KCTD12	0.018728145	7.22	3.72	TGFBI	0.006044786	8.15	4.14
LILRA5	0.041437349	6.41	3.13	TNFAIP2	0.040965892	8.20	4.46
LYZ	0.031494793	10.64	6.54	WDFY4	0.016001568	7.60	4.54

8.2 Genes differentially expressed between licensed and memory-like samples. Genes with higher expression in licensed samples are marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Licensed	Memory			Licensed	Memory
ADGRA3	0.000362118	6.78	1.85	HCAR3	0.007034287	0.56	4.00
ZMAT4	0.010841639	5.71	1.50	HCK	0.00852025	3.42	6.72
ADM	0.006639844	1.67	5.86	IL1B	0.00402717	2.94	8.37
AIF1	0.012708486	3.67	6.94	IL1RN	0.001224724	1.51	6.44
ANPEP	0.037673756	2.29	5.62	JDP2	0.003534086	1.73	5.31
APOBEC3A	0.011131988	1.94	6.01	KCTD12	0.037786526	3.72	6.42
AQP9	0.03566371	1.61	5.66	LILRA5	0.00668812	3.13	6.89
BANK1	0.037663273	4.20	6.37	LRG1	0.00499452	0.83	4.46
C5AR1	0.007449463	2.72	7.33	LTF	0.000848569	1.93	7.95
CD14	0.002721083	3.19	7.30	LYZ	0.00691872	6.54	10.99
CD1D	0.024198792	2.10	5.77	MAFB	0.009714003	2.71	6.76
CD300E	0.013735272	2.86	6.60	MNDA	0.009040035	3.01	6.22

CD4	0.00508166	3.11	6.52	MPEG1	0.013888814	4.31	7.96
CLEC7A	0.018381993	3.27	7.41	MS4A1	0.029387246	4.95	7.44
CSF3R	0.002881093	4.00	8.50	MS4A6A	0.011036874	3.01	5.94
CTB-114C7.4	0.00238312	0.96	4.88	PTGS2	0.002151916	2.32	6.67
CTB-61M7.2	0.04675631	0.79	4.20	PTX3	0.02015828	1.02	4.37
CTSL	0.022884927	1.14	4.42	RAB20	0.002948134	1.24	5.25
CXCL16	0.02174077	2.08	5.76	RP11-443B7.3	0.002839788	1.18	5.03
CXCL8	0.00488742	5.12	11.68	S100A12	0.001824637	1.41	5.75
CYBB	0.014355388	5.45	8.48	S100A8	0.001184547	4.17	9.27
CYP27A1	0.01126043	1.32	4.88	S100A9	0.002490332	5.77	10.62
DGAT2	0.016979779	1.14	4.35	SDC2	0.04257517	0.85	3.98
EREG	0.008680707	2.94	7.54	SERPINA1	0.002142821	3.52	8.21
FAM20C	0.009816509	1.37	4.90	SIRPA	0.013335425	1.71	5.41
FCAR	0.001280308	1.33	6.82	SLC8A1	0.044459555	3.64	6.27
FCN1	0.004944217	4.96	9.32	SPI1	0.004790846	2.76	7.01
FFAR2	0.008995255	1.33	5.33	TGFBI	0.006922601	4.14	7.55
FPR1	0.006778137	2.45	6.99	TNFAIP2	0.007174018	4.46	8.69
G0S2	0.003957594	3.12	8.95	TREM1	0.00425362	2.89	7.62
GNA15	0.029189105	2.83	5.96	TRIB1	0.007907009	3.29	7.47
GPR84	0.009054005	1.16	5.10	VCAN	0.021883606	5.78	9.27
HBEGF	0.007256206	1.83	6.24	WDFY4	0.024770358	4.54	7.01
HCAR2	0.007329884	0.96	4.79				

8.3 Genes differentially expressed between nonlicensed and memory-like samples. The gene with higher expression in nonlicensed samples is marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Non-licensed	Memory			Non-licensed	Memory
ADGRA3	0.017992673	5.88	1.85	LTF	0.017889801	2.60	7.95

Appendix 9 Genes differentially expressed between nonlicensed, licensed and memory-like *ex vivo* stage 5 NK populations (excluding samples from the donor 6) identified from top 400 genes using ANOVA analysis

Expression levels are indicated in Log_2 domain.

9.1 Genes differentially expressed between nonlicensed and licensed samples. All identified genes are higher expressed in nonlicensed samples.

GeneID	p Value	Expression levels	GeneID	p Value	Expression levels

		Non-licensed	Licensed			Non-licensed	Licensed
ALDH2	0.019756039	5.96	2.72	MEFV	0.042740551	4.71	2.17
CD4	0.048825127	5.80	2.76	MS4A6A	0.039622601	5.93	2.70
CFP	0.044028532	6.47	3.21	NAPSB	0.043232221	6.86	3.10
CLEC10A	0.041427478	5.56	1.85	PLD4	0.042811678	6.19	2.62
FLT3	0.005461092	5.41	1.39	SEMA4A	0.036546325	5.86	2.37
IFI30	0.046531008	6.11	2.65	TGFBI	0.016728485	7.69	3.78
IL1RN	0.032653486	3.64	0.95	TNS3	0.033987674	4.88	2.05
KCTD12	0.046912382	6.98	3.39				

9.2 Genes differentially expressed between nonlicensed and memory-like samples. Genes with higher expression in nonlicensed samples are marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Non-licensed	Memory			Non-licensed	Memory
AC009505.2	0.027433182	3.56	0.68	CAMP	0.000879347	1.05	6.75
AC010886.2	0.024800011	5.22	0.99	CEACAM8	0.000901707	0.48	5.65
ADGRA3	0.031749081	6.83	2.48	DEFA3	0.000270012	0.05	8.17
DUTP1	0.012906403	3.78	1.27	DEFA4	7.90E-05	0.00	5.90
MGAM2	0.004072577	5.17	0.10	FCAR	0.041827861	2.67	5.31
MPPED2	0.003327628	4.40	0.00	FPR2	0.032052052	0.37	2.84
MYO1H	0.021001958	4.59	1.46	IL10	0.03493464	1.72	5.15
RASGRF2	0.009361244	8.62	4.75	LAG3	0.041672348	2.44	6.91
RP11-588K22.2	0.001940757	5.84	0.58	LCN2	0.000428392	0.00	6.35
RP11-706O15.3	0.045795177	4.47	0.00	LTF	0.00063943	0.70	9.20
SARDH	0.009948945	6.16	1.33	MPO	0.030842798	1.71	6.32
TRAT1	0.049303543	4.06	0.86	SELENBP1	0.012812314	1.42	5.88
WSCD1	0.025893635	4.44	0.86	TCN1	0.018736252	1.57	4.85
BPI	0.000294287	0.26	6.49	ZNF683	0.03906203	4.90	9.12

9.3 Genes differentially expressed between licensed and memory-like samples. Genes with higher expression in licensed samples are marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Licensed	Memory			Licensed	Memory

AC009505.2	0.029071684	2.95	0.68	HK3	0.028067068	2.76	6.01
AC010886.2	0.047819505	3.97	0.99	HLA-DQA2	0.017708629	2.97	8.65
ADGRA3	0.005578575	7.05	2.48	HLA-DRB6	0.014870348	4.40	8.70
C1QTNF7	0.04512044	4.54	1.94	IFI30	0.034336294	2.65	6.32
DIO1	0.013493943	3.20	0.84	IGHV1-3	0.038941597	1.89	5.10
GJB7	0.01172082	4.30	1.00	IL10	0.003486882	1.25	5.15
MPPED2	0.027335908	2.56	0.00	IL1B	0.03133587	2.36	7.02
RASGRF2	0.036438926	7.21	4.75	IL1RN	0.001469331	0.95	5.17
RP11-588K22.2	0.005871095	4.22	0.58	KB-1615E4.3	0.02003347	1.58	4.50
RP11-706O15.3	0.044894366	3.58	0.00	LAG3	0.005447808	1.96	6.91
RP11-873E20.1	0.030648154	3.85	0.58	LCN2	0.000130518	0.62	6.35
SARDH	0.037394314	4.41	1.33	LILRA5	0.032802459	2.69	5.78
TAS2R50	0.027184442	4.85	1.72	LILRB2	0.014852505	2.13	5.40
BPI	7.62E-05	0.80	6.49	LINC00384	0.045907086	2.45	5.26
BTBD19	0.029351751	2.12	5.16	LINC00877	0.012985086	2.53	6.23
C6orf99	0.048250825	5.57	8.35	LTF	0.000190775	1.49	9.20
CAMP	3.94E-05	0.51	6.75	LYZ	0.047472132	6.07	10.24
CCDC144A	0.033959705	2.03	5.93	MEFV	0.03741949	2.17	4.77
CD4	0.040208376	2.76	5.92	MPO	0.01144651	1.98	6.32
CDA	0.015253614	0.73	3.30	PAX5	0.025461976	2.73	5.61
CEACAM8	0.000260916	0.93	5.65	PTMS	0.009293334	2.83	6.86
CFD	0.021990616	2.40	5.42	RP11-159J3.1	0.028213187	4.81	7.71
CFP	0.023872987	3.21	6.87	RP11-443B7.3	0.011337436	0.77	4.13
CSF3R	0.010244388	3.48	7.33	RP11-689B22.2	0.044481994	2.35	5.00
CTD-2562J15.6	0.041796946	3.29	6.31	RXRG	0.030573467	1.24	4.60
CXCL8	0.020211892	4.33	9.44	S100A8	0.003601321	3.75	9.39
DEFA3	8.18E-05	0.87	8.17	S100A9	0.010120386	5.31	10.33
DEFA4	2.46E-05	0.65	5.90	SELENBP1	0.002943572	1.51	5.88
FCAR	9.64E-05	0.68	5.31	SERPINA1	0.006362658	3.04	7.53
FCN1	0.020629501	4.49	8.54	SLC34A1	0.012251396	2.80	6.33
FMR1-AS1	0.02281783	1.63	5.04	SLC4A1	0.008788832	2.89	6.89
FPR1	0.02826293	1.89	5.30	SPATA22	0.028400685	1.36	4.72
FPR2	0.019208331	0.68	2.84	SPI1	0.027304489	2.29	5.87
G0S2	0.003194579	2.42	7.04	TAS2R63P	0.009431207	5.14	2.05
HBA1	0.021566674	6.53	11.09	TCN1	0.001331981	1.08	4.85
HBA2	0.032073681	8.54	12.71	TNFAIP2	0.049648127	3.96	7.43
HCK	0.04979665	3.10	6.45	TRIB1	0.035694725	2.76	5.83
HEMGN	0.035318619	2.90	5.62	ZNF683	0.033706893	5.66	9.12

Appendix 10 Genes differentially expressed between *ex vivo* and *in vitro* stage 3 NK populations identified from top 400 genes using ANOVA analysis

Expression levels are indicated in Log_2 domain. Genes with higher expression in *in vitro* samples are marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Ex vivo	In vitro			Ex vivo	In vitro
ASPM	0.001757709	7.51	12.67	KIR2DL4	0.008265714	2.34	6.75
SLFN12L	0.007565211	7.56	11.69	KLRC4	0.000979751	1.84	6.74
MYO18B	0.005352704	6.82	11.61	KLRC4-KLRK1	0.004846741	2.35	6.74
NCAM1	0.027386606	6.02	11.53	NOSTRIN	0.005882118	2.28	6.72
CENPE	0.00589387	7.01	11.29	HLA-V	0.006307461	1.89	6.69
CDK1	0.003167144	6.13	10.96	FBXO43	0.010586229	2.38	6.31
CFH	0.001164924	6.21	10.81	AMPH	0.000674115	1.87	6.23
ARHGAP11A	0.002894541	6.09	10.79	RNA5SP390	0.000125648	1.29	6.22
KIF14	0.00106013	5.75	10.72	LDB2	0.028897802	2.56	6.18
RASSF8	0.005942996	6.84	10.70	SGCE	0.000862963	1.76	6.13
GZMA	5.22E-05	3.03	10.59	AC092580.1	0.005863426	1.57	5.70
PTGER3	0.00376404	6.02	10.29	RP11-799B12.2	0.034997144	2.50	5.49
TMEM200A	0.007448746	5.14	10.23	VSTM4	0.041437639	2.22	5.48
HJURP	0.004484735	5.11	9.94	SIK1	0.003435981	9.19	6.23
BUB1B	0.024450973	5.91	9.92	CD300LF	0.007326864	9.32	4.33
ADGRL2	0.000403131	3.52	9.92	CDH1	0.006806712	8.24	3.64
AFAP1L2	0.000124148	4.21	9.88	RP11-307C12.13	0.037303753	6.54	2.92
KIF18A	0.023548809	6.06	9.81	RP4-765C7.2	0.014912757	6.19	2.89
RP11-1094M14.5	0.001449569	3.31	9.54	SEMA3C	0.042358593	5.48	2.57
LZTS1	4.95E-06	2.17	9.51	EREG	9.00E-07	9.71	2.49
RP11-1094M14.8	0.002847264	4.08	9.47	RRAD	0.037927553	6.00	2.49
B4GALNT1	0.049607825	5.10	9.38	NUAK2	0.017550491	5.02	2.44
PEG10	0.000166902	2.28	9.31	AREG	0.014439112	6.05	2.05
IGF2BP3	0.001553652	2.77	9.29	A2M-AS1	0.001517961	5.46	1.96
NTRK2	8.18E-07	1.12	9.27	AKR1C3	0.021980398	5.93	1.85
DRAXIN	3.01E-08	1.70	9.24	AF213884.2	0.003927121	6.31	1.72
NDFIP2	0.011519835	4.66	9.02	C17orf107	0.00376429	5.61	1.66
RP11-277P12.20	0.012629887	5.04	8.99	MSRB3	0.000226098	6.47	1.43
RNF165	0.003193698	3.85	8.94	C1QTNF1	0.031765431	4.81	1.23
DEPDC1	0.008097417	4.34	8.88	RP11-44N11.2	0.001444691	6.50	1.22
E2F3P1	0.007612725	3.84	8.87	NELL2	0.018406665	7.03	1.19
CENPI	0.009730398	4.61	8.79	LINC00612	0.003868064	4.98	1.12
CLECL1	0.001301643	3.26	8.76	ADGRE1	0.017518438	4.48	0.85
TOMM20P2	0.007000877	4.27	8.76	NCF1	0.012215451	4.10	0.83
RP11-	0.003829609	4.34	8.61	CTD-	0.016912858	4.49	0.82

686D22.10				2526A2.2			
NEK2	0.008592297	4.26	8.61	NFIX	2.12E-05	6.08	0.70
E2F8	0.01643722	4.26	8.58	BEX5	0.012240932	5.12	0.70
THRA1/BTR	0.035823687	5.07	8.54	UPB1	0.013121344	3.37	0.58
PTGIS	1.26E-07	0.63	8.45	RP11-252C15.1	0.019035142	5.03	0.53
PVRL3	0.019278926	4.58	8.40	A2MP1	0.026483283	3.39	0.53
AK4	0.000332975	2.09	8.39	IL5RA	0.005842729	4.15	0.52
PBK	0.01290095	3.03	8.33	HLA-U	0.001938466	6.10	0.49
KLRC2	0.001137854	2.56	8.32	CACNA1D	0.031633645	3.22	0.49
ERCC6L	0.038711474	4.71	8.23	BTG1P1	0.039814816	3.19	0.49
ITGA1	0.013994977	3.89	8.12	KRT17P8	0.042201187	3.27	0.49
MYO6	0.026833354	4.87	7.80	ALS2CL	0.004079629	3.89	0.46
RP11-70L8.4	0.03038884	4.97	7.65	CYP2F1	0.028210804	3.78	0.16
OMD	0.007422255	3.23	7.51	LINC00309	0.010292542	4.50	0.16
ASPN	0.011855546	3.17	7.45	AC012307.2	0.023731202	3.88	0.16
SPC25	0.004589728	2.99	7.43	EVPL	0.033548827	3.41	0.16
ALDH1A3	3.06E-05	1.77	7.41	DTX1	0.003035575	6.60	0.10
LINC01146	0.005959369	3.59	7.37	AC022182.1	0.002732063	5.51	0.00
IL26	0.006537708	2.70	7.32	EDARADD	0.01232804	3.11	0.00
CTA-125H2.2	1.42E-05	1.58	7.25	IGHG1	0.016847725	7.59	0.00
PRKCA-AS1	0.013649211	3.51	7.17	CH25H	0.019515034	4.64	0.00
ABLIM3	5.69E-06	1.22	7.08	PCDH20	0.023501069	3.72	0.00
KLRC3	0.000435018	2.23	6.93	IGHG4	0.036326526	5.29	0.00
CYP1A1	9.82E-05	1.44	6.76				

Appendix 11 Genes differentially expressed between *ex vivo* and *in vitro* stage 4 NK populations identified from top 400 genes using ANOVA analysis

Expression levels are indicated in Log_2 domain. Genes with higher expression in *in vitro* samples are marked in green.

GeneID	p Value	Expression levels		GeneID	p Value	Expression levels	
		Ex vivo	In vitro			Ex vivo	In vitro
AC020571.3	0.010554827	1.85	6.01	SLC7A11	0.000122661	1.43	6.54
AFAP1L2	0.010273428	5.07	10.27	SOS1-IT1	0.00271236	3.69	7.81
AMPH	0.006154799	2.91	6.94	TAS2R43	0.016817182	2.97	6.46
ASPM	0.001172334	7.00	12.28	TAS2R64P	0.025780473	3.51	7.62
BCAT1	0.004659351	4.26	9.15	TDRD15	0.00080337	1.87	6.69
C10orf128	0.00175295	1.96	7.25	THEMIS	0.000183037	3.62	10.19
C18orf54	0.003541747	3.58	7.48	TMEM200A	0.019881808	5.98	9.88
C3AR1	0.018126229	3.68	7.57	TPTEP1	0.001377652	4.30	8.16

CCDC150	0.007310545	3.37	7.45	TTLL7	0.000807138	2.41	7.11
CDK1	0.00454538	4.56	10.31	BEX5	0.008738484	4.18	0.78
CDKN2B	0.001295273	2.23	7.69	CD2	0.000342337	10.46	5.80
CENPA	0.004390287	3.27	7.63	COL16A1	0.001670428	4.48	0.00
DEPDC1	0.000398382	2.47	8.54	CYGB	0.000178752	5.29	0.28
DIAPH3	0.007251608	4.42	8.26	EDARADD	0.000103799	4.91	0.70
DLGAP5	0.005140536	4.43	9.88	ETNK2	0.005998641	3.53	0.39
DSCC1	0.007564444	3.21	7.04	FCRL1	0.029697826	4.36	0.18
E2F7	0.000753758	3.37	7.93	FCRL3	0.008851847	7.92	1.28
E2F8	0.005633857	3.43	7.72	HBA1	0.008359766	6.14	0.04
EGLN3	0.001380868	4.07	9.32	HBA2	0.001886636	7.43	0.45
ESCO2	0.009446238	3.58	8.04	HBB	0.002754031	10.16	2.27
GPR15	0.000980635	2.85	8.78	HID1	0.003884196	4.03	0.00
GSG2	0.016159942	3.08	6.91	HOXA-AS3	0.006004183	4.31	0.21
HLA-V	0.003350906	1.97	7.34	IFNG-AS1	0.000114308	7.88	2.52
HMMR	0.000605365	4.73	9.42	IGHM	0.009239485	7.30	2.63
IGF2BP3	0.01181309	5.69	9.24	IGLV3-13	0.002942657	3.10	0.00
IKBIP	0.020018107	4.83	8.46	IL21R-AS1	0.000133906	7.12	2.15
IL7R	0.000111365	11.65	7.07	KCNK17	0.005811878	5.21	0.77
KCNK5	0.02305587	2.73	6.18	KIAA1211L	0.009890233	3.69	0.08
KDELC1	0.004516924	3.31	7.80	L3MBTL4	0.000493433	7.14	0.92
KIF14	0.001796609	4.89	10.19	LARGE	0.011181527	5.04	0.41
KIF20A	0.002954548	3.77	9.07	LINC00925	0.001923865	4.70	0.00
KIZ-AS1	0.010516244	2.32	6.16	LINGO2	0.00335457	4.23	0.03
KLRF2	0.005624246	4.35	9.25	LMNA	0.000152129	10.97	6.12
KNSTRN	0.014351063	4.80	8.37	MS4A1	0.001711099	7.22	1.67
LILRP2	0.007144067	2.00	6.17	MUC16	0.01167359	3.78	0.41
LINC01355	0.007216156	3.78	8.03	NKX3-1	0.018567559	3.25	0.51
MYO5B	0.000537664	2.96	8.00	PMEPA1	0.007668535	4.45	1.02
NCR2	0.001834598	1.16	6.00	RP11-44N11.2	0.000152996	7.10	1.89
NDFIP2	0.000994266	4.06	10.93	RP11-545P6.2	0.017442561	3.13	0.03
NUF2	0.007174789	3.96	9.03	SDPR	0.0103272	4.52	0.21
PBK	0.001053652	1.93	7.72	SH3BP4	0.000196451	6.93	1.95
PRKCA-AS1	0.001148395	3.22	7.72	SLC4A1	0.021005154	4.53	0.00
RP11-21J7.1	0.002488062	2.39	7.45	SNORD3B-1	0.010274656	8.54	5.20
RP11-28F1.2	0.000186709	1.78	7.19	T	0.012970318	3.22	0.03
RP11-328K2.1	0.005608019	2.28	6.31	TMEM204	0.017371895	5.99	3.35
RP11-563D10.1	0.002076368	1.91	6.37	TNFSF9	0.004536794	8.01	3.59
RP11-932O9.10	0.001318369	1.54	5.99	WWC2-AS2	0.001685186	4.87	0.00
RP11-98G13.1	0.001003641	2.25	8.23	ZDHHC20P1	0.00771788	3.34	0.00
RPL23AP22	0.00045306	1.63	6.89	ZMAT4	0.000635	8.06	2.95
SGOL1	0.011877292	4.24	8.05	ZNF365	0.005313754	7.25	3.34

Appendix 12 Ranked top 100 protein-genes differentially expressed within *ex vivo* NK samples

Genes identified from the total number of 14177 genes using LoD1.

GeneID	Score (PC1-3)	Rank	GeneID	Score (PC1-3)	Rank
ENPP1	0.054752916	1	PRSS23	0.038967922	51
IL23R	0.054545918	2	SPINK2	0.038809767	52
RORC	0.054213487	3	CTSG	0.038738028	53
IL7R	0.052599853	4	B3GALT5	0.038733592	54
S1PR5	0.051887348	5	KLRF1	0.038693149	55
MPO	0.050715683	6	MS4A3	0.038543886	56
FGFBP2	0.049178562	7	DEFA3	0.038202797	57
XCL1	0.048513928	8	S100A12	0.03793209	58
LIF	0.047318836	9	FEZ1	0.037894379	59
ARSI	0.046889232	10	CLC	0.037885662	60
TNFSF11	0.046831308	11	TOX2	0.037800911	61
DUSP4	0.046767582	12	CXCL8	0.037683811	62
GPR55	0.04652845	13	FFAR2	0.037664656	63
SLC4A10	0.046490401	14	THEM5	0.037655324	64
IL4I1	0.04568056	15	CDH1	0.037536095	65
GZMH	0.045637846	16	KLRF2	0.037498171	66
KRT81	0.045627665	17	BCAS1	0.037481396	67
PRTN3	0.045165573	18	DTX1	0.03746515	68
KRT86	0.045087726	19	HBG2	0.037388629	69
DTHD1	0.044790453	20	RHAG	0.037264379	70
AZU1	0.04424315	21	CPNE7	0.037233014	71
LINGO4	0.043954412	22	FCAR	0.037176649	72
KLRD1	0.043631485	23	CSPG4	0.037128125	73
ELANE	0.043423083	24	CAMK2A	0.037056427	74
LTF	0.043375786	25	TGFBR3	0.036987477	75
FCGR3A	0.04337522	26	PKDCC	0.036978542	76
HBD	0.043358476	27	N4BP3	0.036915877	77
GZMA	0.042300228	28	PRSS57	0.036787298	78
LPAR1	0.042006769	29	TNFSF9	0.036640388	79
GZMM	0.04180026	30	VWDE	0.03661077	80
KIAA1211L	0.041663323	31	HBB	0.036438067	81
IFNG	0.041566923	32	LTB	0.036420826	82
IL1R1	0.041191723	33	HBG1	0.036358759	83
DLL1	0.04110398	34	MAP7	0.036351706	84
MYO7A	0.040790579	35	ITGB3	0.036302531	85
GZMB	0.040763344	36	B4GALNT1	0.036220613	86
SMIM10L2A	0.040701178	37	IL22	0.036099059	87
KIAA1324	0.040475128	38	CSF2	0.036079662	88
FLT3	0.040460333	39	NCR2	0.036033816	89

MYB	0.040440553	40	EOMES	0.035964139	90
ACAN	0.04033615	41	SORCS1	0.035856048	91
PYHIN1	0.040300451	42	TNFRSF18	0.035815999	92
BNC2	0.039937801	43	CPA3	0.035729712	93
SPTA1	0.03964422	44	TNFAIP8L2	0.035678744	94
TBX21	0.03953964	45	PDZD4	0.03567442	95
IGFBP4	0.039194002	46	B3GAT1	0.03566897	96
C5AR1	0.039185552	47	CD200R1	0.035653068	97
XCL2	0.039132871	48	RNASE2	0.035603947	98
AHSP	0.039095362	49	KIR3DX1	0.035532949	99
PRF1	0.03908271	50	NELL2	0.035521544	100

Appendix 13 Ranked top 100 non-protein genes differentially expressed within *ex vivo* NK samples

Genes identified from the total number of 6613 genes using LoD1.

GeneID	Score (PC1-3)	Rank	GeneID	Score (PC1-3)	Rank
LINC00892	0.058799913	1	AP001171.1	0.038234855	51
RP11-98G13.1	0.055845921	2	AC004791.2	0.038167016	52
RP11-21J7.1	0.050096205	3	RP11-463J10.2	0.037918677	53
AL450992.2	0.04939482	4	RP3-467K16.4	0.037852336	54
RP11-563D10.1	0.048783287	5	LPAL2	0.037764726	55
RP6-91H8.3	0.048270487	6	RP1-206D15.6	0.037718527	56
RP11-330A16.1	0.04733271	7	AMZ2P2	0.037639982	57
LINC01132	0.04728287	8	RP11-624C23.1	0.037454564	58
RP11-354E11.2	0.04629156	9	XXbac-BPG181B23.6	0.037442059	59
RPL23AP22	0.046228632	10	RP5-1028K7.2	0.037408791	60
AC092580.4	0.04553675	11	RP11-705C15.5	0.037379497	61
EDDM3CP	0.045018206	12	AF131217.1	0.037330811	62
RP11-973H7.1	0.043704866	13	HSPE1P18	0.037163487	63
HLA-DQB1-AS1	0.043562304	14	LINC00309	0.037138922	64
AC017104.6	0.043350433	15	AC022182.1	0.037026399	65
RP1-131F15.2	0.043310571	16	RP11-305O6.3	0.036936453	66
GATA2-AS1	0.04330196	17	RP11-693J15.5	0.036901389	67
RP11-15B24.5	0.043140309	18	RP4-738P11.3	0.036822959	68
LINC00299	0.042441564	19	FCGR2C	0.036811323	69
RP11-104L21.3	0.042305008	20	KIAA0125	0.036791233	70
AC007278.2	0.042039656	21	RP11-475O6.1	0.036600145	71
RP11-158I9.5	0.04201662	22	RP11-456H18.2	0.036486872	72
MAGI2-AS3	0.041594372	23	RP11-277P12.20	0.036462396	73
AC092667.2	0.04158054	24	TVP23CP1	0.036452761	74
FTH1P22	0.041559638	25	RP11-121A8.1	0.036337256	75
RP11-326C3.2	0.041541675	26	RP11-47L3.1	0.036250539	76
RP11-212I21.4	0.041498893	27	RP11-757F18.5	0.036215634	77

LINC01268	0.041233947	28	RP11-686D22.10	0.036193839	78
RP11-104L21.2	0.04113833	29	HCP5	0.036026386	79
AOX2P	0.041121305	30	RPS20P2	0.035945755	80
AC002454.1	0.04109729	31	CASC15	0.035938704	81
RP11-252C15.1	0.041032289	32	RP11-567M16.1	0.035876267	82
RP11-520A21.1	0.040856994	33	RP11-347C18.3	0.035807167	83
AC020571.3	0.040596058	34	AC007319.1	0.03577111	84
RP11-389C8.2	0.040435381	35	RP11-794P6.1	0.03572974	85
RP11-366L20.2	0.040123848	36	TOMM20P2	0.035698244	86
RP11-222K16.2	0.040042764	37	NAPSB	0.035674891	87
IGF2BP2-AS1	0.039933586	38	RP11-796E2.4	0.035637531	88
RP1-18D14.7	0.039901388	39	RP11-1094M14.5	0.035556444	89
LINC00880	0.039861751	40	RP11-424C20.2	0.03553821	90
LINC00877	0.039747578	41	AL356585.1	0.035528516	91
AC005083.1	0.039632494	42	RP11-693N9.2	0.035441543	92
RP11-20I20.4	0.039513625	43	RP11-678G14.3	0.035390647	93
RP13-786C16.1	0.039269059	44	KIAA0087	0.035240211	94
AC010761.14	0.039101463	45	RP11-44N11.3	0.035125387	95
SRGAP3-AS2	0.039013378	46	C10orf91	0.03510002	96
AC010886.2	0.038877675	47	B3GALT5-AS1	0.035026045	97
SCART1	0.038576601	48	PCED1B-AS1	0.03485257	98
RP11-342I1.2	0.038341606	49	PLA2G4E-AS1	0.034813687	99
RP11-6F2.5	0.038235569	50	EGFEM1P	0.03480097	100

Appendix 14 Correlation coefficients between expression of protein-coding and non-protein coding genes

Only correlation coefficients >0.8 or <-0.8 ($p<0.05$) and only genes involved in at least one correlation are shown

Gene name	AC006369.2	AC007620.3	AC092580.4	AF131217.1	AL450992.2	AP001372.2	CASC15	CYP4F29P	FTH1P22	GS1-115G20.1	KIAA0087	KIAA0125	KIR2DP1	LINC00892	PCED1B.AS1	PRSS30P
ARSI						-0.81										
B3GAT1	0.83															
BNC2										0.84			0.84			
CD2			0.84													
CDH1										-0.82						
DTHD1										0.85			0.84			
ENPP1														0.84		

Gene name	THEM5	TGFBR3	TBX21	STAC	SPINK2	SLC1A7	SH2D1B	S1PR5	PYHIN1	PRSS23	PRF1	MRC2	MAP7	LGR6	KLRF1	KLRD1	GZMM	GZMH	GZMB	GZMA	GNAI1	FGFBP2	FEZ1	FCGR3A	BNC2	B3GAT1	ARSI
RN7SL653P						0.81				0.82	0.86						0.89	0.90	0.83	0.85		0.84		0.86	0.86		
RP1-244F24.1																			-0.80								
RP11-1094M14.5																											
RP11-20I20.4															0.88					0.81					0.82		
RP11-222K16.2	0.90																										
RP11-229P13.19																										0.84	
RP11-372K14.2				0.81								0.82															0.86
RP11-438B23.2									0.82																0.80		
RP11-463J10.2															0.82										0.84		
RP11-532M24.1									0.80	0.80		-0.81													0.83		
RP11-686D22.10				0.80																							
RP11-693N9.2					0.83			-0.80					0.80								0.82						
RP11-973H7.1		0.81	0.81		-0.84	0.82		0.84	0.80	0.83	0.80			0.87			0.84		0.86			0.85	0.82	0.80		0.82	
SNX18P13																											
ZNF667-AS1		0.83	0.86				0.82	0.82	0.89	0.82	0.93				0.81	0.82	0.83	0.85	0.91	0.91		0.84	0.84	0.89			
			0.83								0.84							0.84									

DTHD1	0.89	-0.80	0.84	0.81				0.82		0.85							
FCGR3A			0.82	0.85				0.83					0.87				
FEZ1			0.80	0.89							0.81		0.89				
FGFBP2		-0.83		0.84									0.87				
FLT3											-0.86						
GNAI1															0.81		
GZMA			0.85	0.81	0.86							0.80	0.85				
GZMB				0.92									0.92				
GZMH													0.81				
GZMM				0.82									0.83				
IFNG													0.85				
KIT															0.86		
KLRD1				0.81									0.83				
KLRF1			0.82		0.84												
LGR6		-0.80								0.81							
LIF		0.82													0.83		
MAP7				-0.83													
MRC2	-0.81	0.80													0.80		
MYB	-0.85			-0.83							-0.81						
PRF1				0.87									0.88				
PRSS23	0.84	-0.83		0.81				0.80		0.85			0.81		-0.84		
PYHIN1	0.87	0.00	0.89	0.87				0.86			0.86		0.88	0.80			
S1PR5		-0.82		0.91				0.81					0.85				
SLC1A7															-0.83		
SPINK2		0.89		-0.82						-0.87			-0.80		0.85		
STAC		0.82													0.82		
TBX21	0.82			0.83									0.83				
TGFBR3			0.81	0.85							0.82		0.85				
VWDE				-0.81													

Appendix 15 Correlation coefficients between expression of non-protein coding and those of closely localized protein-coding genes

Only correlation coefficients >0.5 or <-0.5 ($p<0.05$) and only genes involved in at least one correlation are shown

15.1. Correlation coefficients between expression of non-protein coding and those of closely localized protein-coding genes within all analyzed samples

Gene name	AC002454.1	AC007278.2	AC092580.4	AL450992.2	AOX2P	B3GALT5-AS1	FTH1P22	GATA2-AS1	HCP5	HSPE1P18	LINC00299	LINC00892	LINC01132
AIF1							-0.54		-0.58				
AOX1	0.50				0.53								
B3GALT5						0.61							
C2CD4D				0.81								0.66	
CD101										0.57			
CD2				0.57			0.55		0.50				
CD40LG										0.58			0.59
CD58			0.53										
CLEC2D	-0.53												
EOMES		0.74	0.63				0.74						
GATA2	0.61							0.69					
HMGN1								0.51	-0.74				
HTATSF1									-0.53				
ID2			0.75	0.51							0.68	0.58	
IFITM2								-0.55	0.54		0.50		

IGFBP4				0.59								0.65	
IL10RA	-0.52						0.51		0.54				
IL18RAP		0.91	0.54				0.82		0.52		0.61		
KLRB1		0.71	0.62				0.64	-0.51			0.75		
KLRC1		0.63	0.71				0.65				0.60	0.50	
KLRC2		0.63	0.69				0.63				0.52		
KLRC3		0.65	0.69				0.68						
KLRC4		0.68	0.60				0.57						
KLRD1		0.78	0.65				0.83		0.61		0.56		
KLRF1		0.82					0.82			-0.51			
KLRF2												0.60	
KLRK1		0.70	0.53				0.53						
LTA				0.55									
LTB										0.65		0.52	
NCR3		0.54	0.64				0.68	-0.57	0.55		0.70		
PPIL3									-0.64				
PTGFRN										0.55			
SLFN12L	-0.54	0.70	0.62				0.62	-0.54			0.56		
THEM5			0.51	0.90								0.70	
TRGC2		0.63	0.74				0.76				0.68		

Gene name	RP11-1094M14.5	RP11-121A8.1	RP11-222K16.2	RP11-277P12.20	RP11-326C3.2	RP11-330A16.1	RP11-567M16.1	RP11-686D22.10	RP11-705C15.5	RP11-796E2.4	RP4-738P11.3	RP5-1028K7.2	Xxbac-BPG181B23.6
ACAD9			-0.53	-0.51									
AIF1	-0.57		-0.51					-0.60					
ATHL1					0.86					0.70			
BAG6	-0.51												
BTG1					0.56					0.57			0.51
C2CD4D					0.52					0.52			
CCR7					0.75					0.59			
CD101													
CD2					0.52					0.52			
CD3E					0.54								
CD58		0.51											
CD69													0.69
CLEC2D	0.80							0.69					
TRGV8		0.57	0.61				0.70		0.58		0.61		
TRGV9		0.56	0.79				0.73		0.52		0.73		
TRGVA		0.54					0.55						
XCL1			0.67				0.56				0.69	0.61	
XCL2		0.57	0.75				0.63				0.66	0.58	

EOMES	0.78	0.75	0.88	0.79				0.81	0.68		0.59		
GATA2	-0.51							-0.59					
HMGN1													-0.55
HTATSF1													-0.53
ID2		0.63							0.54		0.67	0.56	
IFITM2		0.60	0.52						0.70		0.65	0.50	
IFITM3									0.52				
IFITM5					0.70					0.57			
IGFBP4						0.76				0.56		0.53	
IL18RAP	0.61	0.77	0.80	0.63				0.71	0.77		0.59		0.52
KLRB1	0.70	0.69	0.67	0.64				0.77	0.72		0.64	0.59	
KLRC1	0.64	0.63	0.70	0.75				0.68	0.73		0.74	0.60	
KLRC2	0.70	0.61	0.74	0.85				0.72	0.68		0.69		
KLRC3	0.68	0.68	0.79	0.89				0.74	0.71		0.67		
KLRC4	0.78	0.62	0.71	0.93				0.76	0.63		0.59		
KLRD1	0.68	0.83	0.86	0.70				0.73	0.72		0.59		0.60
KLRF1	0.69	0.73	0.88	0.67				0.68	0.71				0.53
KLRF2						0.65					0.53		
KLRK1	0.69	0.60	0.69	0.93				0.74	0.55		0.59		
LTA					0.67					0.57			

LTB					0.71	0.58				0.61			
NCR3	0.55	0.72	0.66	0.57				0.64	0.81		0.62	0.70	0.52
NFATC1					0.51		0.51						
PPIL3													-0.59
SLFN12L	0.95	0.71	0.77	0.72				0.95	0.66				
THEM5					0.52	0.52						0.53	
TNF					0.54								
TRGC2	0.60	0.75	0.67	0.59				0.62	0.68		0.52	0.56	
TRGV8	0.54	0.88	0.68	0.54				0.62	0.68		0.56		0.52
TRGV9	0.51	0.80	0.67	0.55				0.59	0.76		0.76	0.61	
TRGVA	0.55	0.74	0.60	0.51				0.61	0.58				
XCL1		0.60		0.56		0.52		0.51	0.69		0.84	0.63	
XCL2		0.67	0.57	0.59				0.55	0.71		0.87	0.61	

15.2. Correlation coefficients between expression of non-protein coding and those of closely localized protein-coding genes within *ex vivo* NK samples

Gene name	**AC002454.1**	**AC007278.2**	**AC092580.4**	**AL450992.2**	**AOX2P**	**B3GALT5-AS1**	**FTH1P22**	**GATA2-AS1**	**HCP5**	**HSPE1P18**	**LINC00299**	**LINC00892**	**LINC01132**	**RP1-206D15.6**
ACAD9	0.58	-0.59											0.60	
AIF1	0.55						-0.61		-0.58					
AOX1					0.62			0.51						

ATHL1				0.61								0.68		
B3GALT5				0.54		0.64				0.75		0.53	0.66	0.58
BTG1														
C2CD4D				0.76		0.52				0.57		0.68	0.58	0.69
CCR7				0.52								0.57		0.63
CD2			0.84	0.67			0.61	-0.53				0.58		
CD3D							0.54							
CD3E			0.55						0.53					
CD40LG										0.56			0.59	
CD69									0.67					
CLEC2D		0.70					0.72							
CLK1		0.58												
EOMES		0.57					0.71		0.55					
FAM133B		0.50												-0.55
GATA2	0.61							0.64						
HMGN1					0.51			0.50	-0.73					
HTATSF1		-0.58					-0.50		-0.59					
ID2			0.59	0.54								0.53		
IFITM2								-0.53	0.60					
IFITM5												0.56		

IGFBP4				0.60						0.52		0.75		0.68
IGSF5										0.52			0.61	0.51
IL10RA							0.65		0.52					
IL18RAP	-0.53	0.80					0.71							
KLRB1		0.60						-0.54			0.55			
KLRC1							0.58							
KLRC2							0.57							
KLRC3							0.66							
KLRC4		0.53					0.55							
KLRD1		0.59					0.77		0.72					
KLRF1		0.68				-0.55	0.82							-0.51
KLRF2				0.70						0.63		0.69	0.57	0.68
KLRK1		0.70												
LST1										0.63			0.70	0.73
LTA				0.55								0.54		
LTB				0.68						0.66		0.76	0.60	0.74
NCR3	-0.58						0.64	-0.64	0.67		0.51			
NFATC1						0.55								
PPIL3									-0.53		-0.53			
PTGFRN						0.53				0.54			0.55	

Gene name	RP11-1094M14.5	RP11-121A8.1	RP11-222K16.2	RP11-277P12.20	RP11-326C3.2	RP11-330A16.1	RP11-567M16.1	RP11-686D22.10	RP11-693J15.5	RP11-705C15.5	RP11-796E2.4	RP4-738P11.3	RP5-1028K7.2	Xxbac-BPG181B23.6
ACAD9	-0.51		-0.57	-0.61				-0.58						
AIF1	-0.65	-0.62	-0.59					-0.73						
AOX1								-0.51						
ATHL1					0.80	0.58					0.80	0.64	0.51	
B3GALT5						0.57	0.58		0.62					
BTG1					0.58						0.57	0.52		
C2CD4D					0.70	0.64					0.72	0.56		
CCR7					0.82	0.71					0.76	0.64		
CD2											0.53	0.52	0.75	
SELL														
SLFN12L	-0.52	0.68					0.72		0.50					-0.64
TAF5L													-0.52	
THEM5				0.90								0.71	0.56	0.53
TRGC2			0.64				0.74	-0.56	0.52					
TRGV8							0.60	-0.50	0.64					
TRGV9			0.69				0.72	-0.62	0.75					
XCL1				0.57								0.71		
XCL2			0.54						0.54			0.58		

CD3D	0.63		0.61											
CD3E											0.58	0.59		0.60
CD40LG														-0.55
CD69												0.59		0.74
CD83					0.60						0.55			
CLEC2D	0.89	0.62	0.68	0.72				0.78		0.70				
CLK1				0.51										
EOMES	0.72	0.69	0.80	0.70				0.70		0.56				
FAM133B						-0.56								
HMGN1		-0.66						-0.55		-0.52				-0.57
HTATSF1			-0.53	-0.57										-0.63
ID2									0.54				0.57	
IFITM2										0.60	0.56	0.60		0.64
IFITM3												0.51		
IFITM5					0.69	0.53					0.65	0.52		
IGFBP4					0.77	0.77					0.73	0.54		
IGSF3	-0.53							-0.52		-0.51				
IGSF5									0.57				0.55	
IL10RA			0.60	0.60						0.59				
IL18RAP	0.69	0.61	0.70	0.52				0.80		0.69				

KLRB1	0.56							0.59		0.53			0.53	
KLRC1			0.54		0.50					0.52	0.56			
KLRC2	0.52		0.64	0.77										
KLRC3	0.58		0.72	0.84			-0.52	0.56		0.53				
KLRC4	0.73		0.61	0.90				0.60		0.52				
KLRD1	0.73	0.78	0.73	0.53				0.69		0.60				0.61
KLRF1	0.83	0.74	0.84	0.62				0.70		0.72				
KLRF2					0.64	0.73			0.61		0.59			
KLRK1	0.70		0.58	0.87				0.71						
LST1	-0.69	-0.53	-0.62		0.52	0.61		-0.73						
LTA					0.72						0.74	0.67		
LTB					0.82	0.82			0.57		0.81	0.56	0.54	
NCR3		0.56	0.59					0.54		0.66			0.52	0.70
NFATC1							0.54		0.51					
PPIL3										-0.57		-0.52		-0.60
PTGFRN									0.52					
SELL						0.53								
SLFN12L	0.94	0.80	0.75	0.63		-0.51		0.88		0.64				
TAF5L														
THEM5					0.71	0.65					0.67		0.54	

TNF					0.60						0.60	0.66		
TRGC2	0.58	0.72	0.60							0.58			0.54	
TRGV8	0.53	0.84	0.53					0.53		0.51				0.54
TRGV9		0.65	0.54							0.56		0.54	0.68	0.69
TRGVA	0.63	0.61						0.62						
XCL1					0.80	0.57					0.85	0.71	0.56	
XCL2					0.62						0.67	0.79	0.54	

15.3. Correlation coefficients between expression of non-protein coding and those of closely localized protein-coding genes within *ex vivo* stage 3 NK samples

Gene name	**AC002454.1**	**AC007278.2**	**AC092580.4**	**AL450992.2**	**AOX2P**	**B3GALT5-AS1**	**FTH1P22**	**GATA2-AS1**	**HCP5**	**HSPE1P18**	**LINC00299**	**LINC00892**	**LINC01132**	**RP1-206D15.6**
ACAD9		-0.89										-0.69		
AOX1					0.62									
ATHL1	-0.75		0.61	0.65	-0.91	0.74		-0.63	0.92	0.65			0.90	
C2CD4D	-0.79		0.73	0.70	-0.87	0.81		-0.72	0.94	0.65			0.95	0.65
CCR7					-0.64									
CD2	-0.88		0.95	0.95		0.95	0.68	-0.78	0.91	0.80	0.77	0.78	0.91	0.96
CD40LG	-0.69		0.80	0.75		0.71	0.75	-0.62			0.83	0.88		0.80
CLK1		0.85												
FAM133B			-0.82	-0.64		-0.72			-0.65	-0.73	-0.70		-0.63	-0.72

GATA2	0.87		-0.79	-0.83		-0.88	-0.80	0.91	-0.74	-0.71	-0.71		-0.80	-0.76
GATAD1							0.61							
HTATSF1							0.70							
ID2	-0.70			0.67		0.62		-0.74		0.65				0.63
IGFBP4	-0.75		0.79	0.66		0.77		-0.72	0.78		0.82	0.68	0.85	0.69
IGSF3			-0.65								-0.71		-0.62	-0.65
IL18RAP	-0.63	0.61		0.61										
IL1RL1						-0.66		0.69						
KLRC1		0.79									0.65	0.82		0.65
KLRC2		0.86												
KLRC3											0.61	0.66		
KLRC4		0.94										0.71		
KLRK1		0.83												
PPIL3					0.88				-0.63					
SLFN12L	-0.69		0.71	0.65		0.66		-0.62			0.91	0.69	0.64	0.70
THEM5	-0.86		0.89	0.89	-0.69	0.92	0.64	-0.74	0.96	0.82	0.65	0.65	0.95	0.89
TRGV8					-0.79									
B3GALT5	-0.85		0.94	0.90		0.91	0.63	-0.82	0.81	0.75	0.90	0.81	0.85	0.91
BAG6					-0.62									
BTG1	-0.62	0.75		0.62										0.64

TRGV9	TRGC2	TNF	NFKBIL1	NFATC1	NCR3	LTB	LTA	LST1	KLRF2	KLRF1	KLRB1	IGSF5	IFITM2	HMGN1	CLEC2D	CD83	CD69	CD3E	CD3D
-0.93	-0.76			-0.69	-0.80	-0.77	-0.91	-0.62	-0.67			-0.84					-0.80	-0.90	
										0.63	0.64				0.69				0.78
0.91	0.61			0.66	0.79	0.61	0.93		0.73			0.94					0.66	0.91	
0.91	0.70			0.62	0.79	0.61	0.95		0.67			0.86					0.87	0.89	
		-0.90	-0.72	-0.92	-0.83	-0.83	-0.61	-0.75					-0.72	0.78		-0.61			
0.96	0.61	0.62	0.64	0.75	0.85	0.71	0.95		0.63	-0.67		0.91					0.70	0.94	
0.70	0.80					0.62	0.72					0.74					0.73	0.62	
-0.88	-0.61			-0.69	-0.66	-0.69	-0.76			0.61		-0.74						-0.87	
0.87		0.75	0.75	0.87	0.97	0.89	0.95	0.65	0.61	-0.62		0.80					0.70	0.82	
0.76		0.65		0.70	0.74		0.79					0.61					0.66	0.80	
0.82	0.75						0.72		0.79		0.72	0.78						0.87	
0.71	0.86						0.71		0.84		0.81	0.72					0.75	0.70	0.62
0.92		0.71	0.75	0.86	0.95	0.90	0.95	0.68	0.65	-0.64		0.85					0.67	0.87	
0.90	0.73				0.75		0.93		0.74			0.85					0.85	0.89	

IGSF3	IGFBP4	ID2	GATA2	FAM133B	EOMES	CLK1	CD58	CD40LG	CD2	CCR7	C2CD4D	ATHL1	AOX1	ACAD9	Genen name	KLRD1	XCL2	XCL1
						0.68								-0.68	**RP11-1094M14.5**	-0.72	-0.82	-0.71
	0.63			-0.70						0.73	0.64				**RP11-121A8.1**			
						0.69								-0.62	**RP11-222K16.2**		0.95	
						0.78		0.66						-0.82	**RP11-277P12.20**		0.92	
	0.72	0.72	-0.82						0.86		0.88	0.90			**RP11-326C3.2**			-0.64
	0.78	0.69	-0.92					0.62	0.75		0.75	0.66			**RP11-330A16.1**	0.63	0.91	0.61
	0.74	0.65	-0.71	-0.64					0.74		0.95	0.94			**RP11-567M16.1**			
-0.71	0.76		-0.72						0.64						**RP11-686D22.10**	-0.76	-0.77	
		0.71			0.73		0.63	0.66		0.62				-0.66	**RP11-693J15.5**		0.81	0.69
-0.63	0.79										0.65	0.65	-0.71		**RP11-705C15.5**		0.81	
-0.77	0.69						-0.61				0.76	0.77	-0.84		**RP11-796E2.4**	0.77	0.89	
-0.61	0.85	0.65	-0.74	-0.61					0.64	0.63	0.83	0.72	-0.69		**RP4-738P11.3**		0.81	
-0.76	0.79		-0.75	-0.82				0.82	0.95		0.66				**RP5-1028K7.2**	0.71	0.84	0.76
	0.82		-0.81						0.77		0.85	0.81			**Xxbac-BPG181B23.6**		0.94	

IL1RL1					-0.65									
KLRC1	0.78			0.94					0.71				0.63	
KLRC2	0.90			0.81										
KLRC3				0.72				0.61	0.65					
KLRC4	0.90			0.97										
KLRK1	0.86			0.72										
PPIL3					-0.66		-0.85			-0.63	-0.88	-0.73		
PTGFRN		0.66							0.68					
SLFN12L						0.78		0.92		0.75	0.67	0.74	0.84	0.67
TAF5L		-0.61	0.63									-0.64		
THEM5					0.87	0.73	0.81	0.61				0.62	0.86	0.86
TRGV8		0.84	-0.85				0.78				0.62	0.80		
TRGVA		0.81	-0.61											
B3GALT5					0.82	0.84	0.72	0.76			0.64	0.77	0.95	0.76
BAG6			-0.65	-0.71										
BTG1				0.70										
CD3D	0.69			0.83						0.73				
CD3E					0.89	0.93	0.77	0.86		0.63	0.68	0.84	0.94	0.81
CD69					0.66				0.74				0.73	
CD83														

CLEC2D	0.67			0.73				0.72		0.63				
HMGN1		-0.83	0.78				-0.83					-0.84		
IFITM2					0.68	0.65	0.76	0.65		0.74	0.92	0.86		
IFITM5	0.61													
IGSF5					0.75	0.77	0.61	0.67				0.62	0.88	0.83
KLRB1				0.76					0.79					
KLRF1	0.62				-0.74	-0.64	-0.75					-0.69		
KLRF2	0.65			0.67				0.64	0.69				0.79	0.63
LST1		0.70			0.70	0.68	0.82				0.71	0.86		
LTA					0.89	0.78	0.79	0.65				0.65	0.90	0.84
LTB					0.82	0.67	0.87			0.67	0.64	0.75		0.75
NCR3		0.62			0.87	0.69	0.91				0.65	0.68	0.73	0.82
NFATC1		0.73	-0.61		0.85	0.71	0.98			0.63	0.79	0.85	0.62	0.73
NFKBIL1					0.67		0.78							0.70
TNF		0.81			0.71		0.92				0.73	0.78		
TRGC2				0.68		0.67		0.63	0.85				0.69	
TRGV9					0.94	0.90	0.80	0.79		0.62	0.68	0.82	0.91	0.81
XCL1					0.74	0.73	0.70	0.83		0.87	0.89	0.75	0.62	0.85
XCL2					0.78	0.80	0.69	0.75				0.70	0.98	0.76
KLRD1					0.64	0.83		0.85		0.88		0.79	0.65	0.81

15.4. Correlation coefficients between expression of non-protein coding and those of closely localized protein-coding genes within *ex vivo* stage 4 NK samples

Gene name	AC002454.1	AC007278.2	AC092580.4	AL450992.2	AOX2P	FTH1P22	GATA2-AS1	HCP5	HSPE1P18	LINC00299	LINC00892	LINC01132	RP1-206D15.6	RP11-1094M14.5
ACAD9								0.53						
B3GALT5									0.55					
BAG6								0.67						
BTG1						-0.61				-0.56				
C2CD4D				0.59									0.56	
CCR7														-0.65
CD2			0.63	0.65		0.63								
CD3E		-0.75												
CD40LG		0.54									0.57			0.52
CD69								0.64						
CD83						-0.81	0.57			-0.52	-0.55			-0.61
CLEC2D						0.69					0.58			0.87
CLK1			-0.56											
GATAD1						0.56								0.68
HMGN1								-0.68						
HTATSF1													-0.51	
IFITM3												0.50	0.58	

IGSF3					0.57									
IL18RAP		0.67												
KLRB1		0.62	-0.53											
KLRC2		-0.53		0.59										
KLRC3				0.53										
KLRF1						0.79				0.61	0.59			0.71
KLRK1							0.55							
LST1												0.61	0.56	
LTA						-0.56								-0.62
LTB												0.58	0.60	
NCR3							-0.60			0.58				
NFKBIL1						0.59						0.55		
PPIL3								-0.53				-0.58	-0.54	
RAB7A													0.55	-0.58
S100A10		-0.61												-0.58
SELL						0.73					0.66			
SLFN12L						0.65				0.53	0.53			0.88
TAF5L			-0.53			-0.53						-0.57		
THEM5				0.64										
TNF						-0.56	0.62							-0.63

CD69	CD58	CD40LG	CD3E	CD2	CD101	CCR7	C2CD4D	BTG1	BAG6	B3GALT5	ATHL1	AIF1	Genen name	XCL2	TSGA10	TRGV9	TRGV8	TRGC2
	0.54							-0.60					**RP11-121A8.1**	0.52				
				0.53									**RP11-222K16.2**					
									-0.60				**RP11-277P12.20**					
						0.65	0.53	0.52					**RP11-326C3.2**					
													RP11-330A16.1					
										0.50			**RP11-567M16.1**		0.77	0.56		0.59
		0.59	-0.56		-0.56							-0.52	**RP11-686D22.10**					-0.52
										0.52			**RP11-693J15.5**			0.66	0.53	
								-0.80					**RP11-705C15.5**					
0.56							0.53		0.66		0.53		**RP11-796E2.4**					
0.66						0.69	0.52		0.58				**RP4-738P11.3**		0.63			
								-0.58					**RP5-1028K7.2**					
0.69		-0.57							0.77				**Xxbac-BPG181B23.6**					
															0.65			0.52

CD83	-0.67	-0.65							-0.82			-0.65	
CLEC2D							0.65		0.57				
FAM133B			0.64										
GATA2													
GATAD1		0.54					0.60						
HMGN1	-0.63		0.63										
HTATSF1											-0.63		
IFITM2			-0.55							0.59			0.54
IFITM3							-0.54			0.50	0.51		
IGSF3											0.66		
IL18RAP	0.51						0.65						
IL1RL1								0.56				0.54	
KLRB1	0.52												
KLRC3			0.60										
KLRC4			0.83										
KLRD1	0.70			-0.51									
KLRF1	0.75	0.56							0.85			0.67	
KLRK1			0.66										
LST1							-0.62						
LTA	-0.62								-0.63				

LTB				0.51	0.57					0.66	0.56		
NCR3									0.60			0.67	0.50
NFATC1			-0.68										
NFKBIL1									0.53	0.58		0.61	
PPIL3										-0.58		-0.59	-0.62
PTGFRN				0.54									
RAB7A							-0.62				0.53		
S100A10							-0.61						
SELL	0.54	0.75			0.52				0.70			0.66	
SH3BGR								0.57				0.55	
SLFN12L	0.73	0.62					0.58		0.76			0.59	
TAF5L					-0.57								
TNF	-0.62	-0.56					-0.62		-0.57		0.70		0.57
TRGC2	0.58	0.59							0.52				
TRGV8	0.67		-0.54										
TRGV9	0.56								0.50			0.54	0.54
TSGA10		0.73							0.82			0.72	

15.5. Correlation coefficients between expression of non-protein coding and those of closely localized protein-coding genes within *ex vivo* stage 5 NK samples

ID2	HTATSF1	GATAD1	GATA2	EOMES	CLK1	CLEC2D	CD58	CD40LG	CD3E	CD3D	CD2	CCR7	C2CD4D	BAG6	B3GALT5	AIF1	ACAD9	Gene name
																		AC002454.1
	-0.67					0.70												**AC007278.2**
0.87							0.82	0.63			0.76		0.66		0.67			**AC092580.4**
			-0.80					0.79								-0.75	0.85	**AL450992.2**
									-0.68									**AOX2P**
																	0.68	**B3GALT5-AS1**
				0.78								0.67						**FTH1P22**
																		GATA2-AS1
											0.66							**HCP5**
0.86			-0.66				0.69	0.89			0.84				0.88		0.77	**HSPE1P18**
		-0.81								-0.72				0.78				**LINC00299**
																		LINC00892
0.79			-0.66					0.81			0.83				0.88		0.85	**LINC01132**
0.84							0.73				0.83		0.70					**RP1-206D15.6**

IFITM2						-0.70								
IFITM5											0.81			
IGSF3									0.65					
IGSF5			0.75	0.70						0.90			0.84	0.85
IL10RA														
IL18RAP														
IL1RL1			-0.70					0.75				-0.67		-0.87
KLRC2		0.78									-0.64			
KLRC3							0.66				-0.94			
KLRD1							0.83							
KLRF1	0.68				0.70		0.77							
KLRF2				0.76										
LTA					-0.64							0.75		
LTB			0.78						0.68	0.66		0.68	0.66	0.89
NCR3					-0.67									
NFKBIL1									0.65					
PPIL3			0.66								-0.75			
PTGFRN			0.64							0.70				
S100A10				-0.76						-0.71			-0.63	
SELL						-0.70	0.79							

CD3D	CD2	CCR7	C2CD4D	B3GALT5	ATHL1	AIF1	Gene name	XCL2	XCL1	TRGVA	TRGV9	TRGV8	TRGC2	TNF	THEM5	TAF5L	SLFN12L	SH3BGR
							RP11-1094M14.5											
	0.84			0.68			**RP11-121A8.1**											
0.65					-0.85		**RP11-222K16.2**					0.76				-0.70		0.66
			-0.75				**RP11-277P12.20**			0.83					0.70			
					0.79		**RP11-326C3.2**						0.65				0.69	
-0.65	0.71					-0.64	**RP11-330A16.1**							0.70				
							RP11-567M16.1						0.72					
	-0.75		-0.73	-0.74			**RP11-686D22.10**											
						-0.90	**RP11-693J15.5**	0.89								-0.84		
		0.66					**RP11-705C15.5**			0.80		0.71	0.67			-0.77		
							RP11-796E2.4											
							RP4-738P11.3		0.89									
	0.84		0.77	0.77			**RP5-1028K7.2**			0.81			0.63		0.66	-0.75		
							Xxbac-BPG181B23.6			0.77						-0.81		

CD40LG									0.73				0.68	
CD58			0.68		-0.67			-0.81					0.82	
CD69														0.67
CD83		-0.87				-0.67								
CLEC2D	0.75			0.77										
CLK1		-0.67		0.66										
EOMES	0.67	0.65												
GATA2									-0.65					
HMGN1						-0.75								
HTATSF1											-0.77			
ID2								-0.98					0.95	
IFITM5														0.81
IGFBP4				-0.68										
IGSF3			0.71					-0.76					0.68	0.64
IGSF5								-0.79	0.68				0.85	
IL10RA										0.78				
IL18RAP											-0.70			
IL1RL1								0.68					-0.78	
KLRB1		-0.75												
KLRC1			0.68		-0.77		-0.63							

KLRC2						-0.67								
KLRC3			0.73											
KLRC4	0.70			0.71										
KLRD1			0.77		-0.81		-0.75							
KLRF1			0.67		-0.82									
KLRF2						0.71			0.69					
KLRK1				0.90										
LST1				-0.65										
LTB								-0.83					0.90	0.73
NFKBIL1				-0.72				-0.64					0.68	0.67
PTGFRN		0.64												
RAB7A				-0.64										
S100A10									-0.74					
SELL							-0.66							
SH3BGR				-0.64				-0.66					0.70	
SLFN12L	0.92	0.70												
TAF5L								0.87				-0.64	-0.89	-0.67
THEM5						0.66								
TRGC2		0.65												
TRGV8		0.73						-0.68					0.71	

TRGV9								-0.79				0.71	0.67
TRGVA		0.63									0.74	0.63	
XCL1										0.70			0.68
XCL2										0.70			0.73

Appendix 16 NK-specific genes used for single cell analysis

Protein-coding			Non-protein-coding	
ADAMTS1	FGR	PDGFD	A2MP1	LUCAT1
ATP8B4	FOSL2	PDGFRB	AC017104.6	PLCB1-IT1
B3GAT1	GNLY	PRF1	AC092580.4	RP11-1094M14.5
BNC2	GZMA	PRSS23	AF131217.1	RP11-183I6.2
C1orf21	GZMB	PTGDR	ANKRD20A11P	RP11-222K16.2
CCNJL	GZMH	RASSF4	FCGR2C	RP11-277P12.20
CD300A	HOPX	RNF165	GS1-115G20.1	RP11-305L7.1
CLIC3	IL18RAP	S1PR5	HLA-DRB6	RP11-452D12.1
CMKLR1	ITGAM	SH2D1A	LINC00299	RP11-456D7.1
COL13A1	ITK	SH2D1B	LINC00861	RP11-81H14.2
CST7	KIR3DX1	SIGLEC7	LINC00877	RP13-516M14.2
CX3CR1	KLRD1	SLAMF7	LINC01108	SIGLEC17P
CXCR2	KLRF1	SLC1A7	LPAL2	U91324.1
DAB2	L3MBTL4	SLC4A4		
DTHD1	LINGO2	SLCO4C1		
FAT4	NCAM1	TRDC		
FCGR3A	NCR1	TRDJ1		
FCRL6	NKG7	TRGV9		
FEZ1	NME8	TTC38		
FGFBP2	PCDH1	TYROBP		

Appendix 17 Ranked top 89 genes differentially expressed within single cell samples

Genes identified from the total number of 172 genes using LoD1.

GeneID	Score (PC1-3)	Rank	GeneID	Score (PC1-3)	Rank
RP11-277P12.20	0.299472148	1	CX3CR1	0.137893714	46
LINC00299	0.261794142	2	IGLL5	0.13705548	47
CST7	0.253418982	3	SLAMF7	0.136352997	48
AC092580.4	0.236154646	4	IGHM	0.134048472	49
KLRD1	0.233888096	5	RP11-452D12.1	0.133816264	50
IGF1R	0.222037847	6	FOSL2	0.132753339	51

CD52	0.219360474	7	IGHA2	0.125868851	52
CLIC3	0.218038909	8	IGHG1	0.120504667	53
IL6ST	0.216909275	9	GZMH	0.119713812	54
NKG7	0.214374081	10	NELL2	0.118108594	55
FCGR3A	0.212705178	11	IGLV6-57	0.11563056	56
CD37	0.211009942	12	CD19	0.112480522	57
PRF1	0.207685164	13	IGLC2	0.110738037	58
GZMB	0.207636059	14	L3MBTL4	0.108641719	59
TRDJ1	0.206808793	15	IGHMBP2	0.106941898	60
IL32	0.204479471	16	IGHG3	0.105174127	61
SH2D1A	0.197002444	17	IGLC3	0.104732575	62
ITK	0.196889663	18	FEZ1	0.10295379	63
TYROBP	0.196845947	19	LINC00861	0.099697808	64
CD59	0.196802894	20	ANKRD20A11P	0.098024219	65
CD3D	0.194678508	21	NCR1	0.097355715	66
TRDC	0.193499757	22	CCR7	0.085949742	67
IGKC	0.192615816	23	FGR	0.085158447	68
GNLY	0.187869334	24	IGHV1-3	0.078918547	69
RP11-81H14.2	0.187684034	25	IGF2R	0.073016456	70
IL7R	0.186110755	26	IGBP1	0.072450968	71
CD3E	0.182619866	27	FCGR2C	0.072379785	72
GZMA	0.182606671	28	DTHD1	0.069269068	73
AF131217.1	0.175067402	29	TRGV9	0.068354073	74
IGJ	0.173851	30	IGHG4	0.06353194	75
IGHV5-51	0.173795501	31	RP11-1094M14.5	0.051424615	76
IGHA1	0.172992872	32	TTC38	0.050187768	77
IGLV3-19	0.169290032	33	ADAMTS1	0.045409684	78
HOPX	0.168971947	34	GS1-115G20.1	0.044262823	79
IL27RA	0.165765091	35	DAB2	0.04248583	80
C1orf21	0.160622497	36	NCAM1	0.041222876	81
CD82	0.159271777	37	ITGAM	0.03912345	82
FGFBP2	0.158105365	38	RP11-456D7.1	0.036664183	83
SH2D1B	0.157642569	39	ATP8B4	0.036534158	84
IGLC1	0.15281992	40	LEF1	0.033533177	85
IL18RAP	0.151770948	41	IGIP	0.032284662	86
KLRF1	0.150570604	42	IGHV1OR15-1	0.029346265	87
RP11-222K16.2	0.150133842	43	CMKLR1	0.028405811	88
SLCO4C1	0.14532052	44	CD300A	0.024991681	89
AHR	0.141135408	45			

Appendix 18 Correlation between expression of non-protein coding genes and those of protein-coding genes within single NK cells

Only significant correlation coefficients ($r>0.5$ or $r<-0.5$, $p<0.05$) and only genes involved in at least one significant correlation are shown.

Gene name	AC079949.1	AC092580.4	AF131217.1	ANKRD20A11P	CASC15	EDDM3CP	FCGR2C	FCGR2C	GS1.115G20.1	HCP5	LINC00299	LINC00861
ATP8B5									0.66			
BNC3	0.50									0.54		
C1orf22		0.59								0.68		
CCNJL			0.57									
CLIC4	0.58			0.51		0.52						
CMKLR2										0.53		
CST8											0.51	
DAB3									0.68			
DTX2			0.62									
DUSP5			0.55									
FCGR3A				0.66								
FEZ2												
GNLY		0.62		0.50								
GZMA		0.53										
GZMB		0.66										
IL18RAP						0.53		-0.69				
IL23R					0.69							
ITGAM										0.55		
KLRD2				0.65		0.52						
L3MBTL5			0.82									
MYO7A							0.54					
N4BP4									0.51			
NCAM2		0.57										
NCR2												0.52
PCDH2					0.69							
PDGFD												
PDGFRB					0.69							
PRF2											0.52	
PYHIN2		0.58										
SH2D1A	0.57	0.58										
SLC4A5					0.69							
SLCO4C2											0.70	
SPINK3												
TBX22												
TGFBR4											0.65	
TTC39										0.53		
XCL2									0.75	0.51	0.63	
XCL3											0.52	
TRGV10								-0.53				

Gene name	LINC01108	MAGI2.AS3	RP11.1094M14.5	RP11.222K16.2	RP11.277P12.20	RP11.366L20.2	RP11.452D12.1	RP11.456D7.1	RP11.686D22.10	RP11.693N9.2	RP11.81H14.2	RP5.857K21.11
ADAMTS2						0.53						
BNC3								0.51				
C1orf22				0.52								
CAMK2A	0.72	0.84				0.58				0.84		
CLIC4					0.65		0.51					
CMKLR2		0.72	0.52			0.58				0.72		
CX3CR2				0.57	0.52				0.51			
DTX2							0.56					
DUSP5	0.64		0.59									
FCGR3A					0.69							
FEZ2					0.51	0.50						
GNLY					0.71							
GZMA					0.59			0.54				
GZMB								0.50				
GZMH						0.60						
IL18RAP					0.58							
IL23R		0.50										
KLRD2					0.52							
NCAM2		0.69		0.54		0.65				0.70		
NELL3		0.72								0.72		
NKG8			0.53					0.51				
PCDH2		0.50										
PDGFD		1.00				0.59				0.99		
PDGFRB		0.50										
SH2D1A					0.65							
SLC4A11											0.53	
SLC4A5		0.50										
SPINK3									0.53			
TBX22		0.61				0.59				0.61		0.57
TTC39		0.63								0.63		
TYROBP												
VWDE		1.00				0.59				0.99		
XCL2												0.61
FOSL2								-0.56				
NFIL4				0.59								

Appendix 19 Correlation between expression of non-protein coding genes and those of protein-coding genes within cell populations

Only significant correlation coefficients (r>0.5 or r<-0.5, p<0.05) and only genes involved in at least one significant correlation are shown.

19.1 Correlation coefficients between expression of protein-coding and non-protein genes within all analyzed samples

Gene name	AC022182.1	AC092580.4	AF131217.1	ANKRD20A11P	CASC15	EDDM3CP	FCGR2C	GS1-115G20.1	HCP5	LINC00299	LINC00861
ADAMTS1			0.72	0.51	-0.56		0.51	0.75			0.55
BNC2			0.70	0.58	-0.57		0.69	0.75			0.63
C1orf21			0.74	0.53	-0.57		0.61	0.83			0.56
CCNJL		0.61	0.73					0.51	0.50	0.60	
CLIC3		0.60	0.78		-0.60		0.66	0.56	0.69	0.67	0.68
CMKLR1			0.59	0.56			0.63	0.59		0.56	0.57
CST7			0.68	0.55	-0.63		0.59	0.64	0.54		0.61
CXCR1			0.53	0.60	-0.52		0.56	0.53			0.57
DAB2							0.57	0.56			
FCGR3A			0.75	0.51	-0.63		0.77	0.74			0.62
FEZ1			0.74	0.63	-0.60		0.60	0.63	0.58		0.66
FOSL2			0.53			-0.50	0.58	0.59			
GNLY		0.69	0.77		-0.64		0.53	0.60	0.52	0.60	0.62
GZMA		0.57	0.79		-0.67		0.58	0.75			0.66
GZMB		0.61	0.72	0.53	-0.61		0.64	0.72	0.53	0.64	0.62
GZMH			0.78	0.55	-0.66		0.63	0.75			0.71
IL18RAP		0.55	0.81		-0.57		0.56	0.68	0.54	0.61	0.61
IL23R		0.54								0.51	
ITGAM			0.58				0.72	0.54			
KLRD1		0.65	0.83	0.53	-0.61		0.56	0.68	0.61	0.57	0.69
L3MBTL4	0.58		0.62	0.65					0.66		0.51
NCAM1		0.68	0.72		-0.54			0.58		0.64	0.53
NCR1		0.66	0.79		-0.57		0.54	0.50	0.65	0.68	0.65
NKG7			0.76		-0.61		0.66	0.71		0.50	0.64
PCDH1			0.63	0.71			0.53	0.63	0.52		0.59
PDGFD			0.80		-0.55		0.58	0.79			0.57
PDGFRB				0.57			0.58	0.54		0.50	
PRF1		0.59	0.81		-0.68		0.61	0.69	0.59	0.57	0.72
PYHIN1			0.84	0.54	-0.74		0.56	0.74	0.51		0.78
SH2D1A			0.59					0.54			
SLC4A10		0.57								0.56	
SLC4A4			0.80	0.51	-0.57		0.57	0.78			0.62
SLCO4C1			0.66	0.54	-0.55		0.72	0.79		0.54	0.58
SPINK2					0.54						-0.55
TBX21		0.52	0.81	0.59	-0.76		0.61	0.69	0.64	0.50	0.74

TGFBR3			0.79	0.63	-0.72		0.60	0.74	0.55		0.78
TRGV9		0.79	0.65						0.52	0.73	
TTC38		0.53	0.71	0.57	-0.58		0.65	0.67	0.57	0.62	0.56
TYROBP			0.60				0.66		0.57	0.58	
VWDE			-0.51		0.58		-0.63		-0.62		-0.51
XCL1		0.67								0.69	
XCL2		0.75	0.55							0.67	

Gene name	LINC01108	RP11-1094M14.5	RP11-222K16.2	RP11-277P12.20	RP11-366L20.2	RP11-452D12.1	RP11-456D7.1	RP11-686D22.10	RP11-693N9.2	RP11-81H14.2
ADAMTS1		0.65	0.75	0.61		0.68	0.50	0.66	0.51	0.55
ATP8B4						0.58				
BNC2		0.59	0.72	0.60		0.57		0.62	0.55	0.70
C1orf21		0.54	0.75	0.63		0.59		0.60	0.56	0.69
CAMK2A	0.66									
CCNJL			0.73	0.60		0.75		0.63		0.57
CLIC3		0.50	0.78	0.56		0.65		0.59	0.65	0.78
CMKLR1			0.64			0.55			0.53	0.73
CST7			0.74	0.50				0.55	0.56	0.68
CXCR1			0.61						0.51	0.66
DAB2										0.55
FCGR3A		0.58	0.82	0.61		0.56		0.63	0.62	0.70
FEZ1		0.54	0.73	0.51				0.55	0.69	0.67
FOSL2			0.58	0.52						
GNLY		0.61	0.82	0.73		0.75		0.69	0.53	0.61
GZMA		0.80	0.86	0.72		0.72	0.58	0.81	0.61	0.59
GZMB		0.65	0.78	0.65		0.67		0.71	0.63	0.76
GZMH		0.61	0.81	0.64		0.52		0.62	0.67	0.69
IL18RAP		0.60	0.80	0.63		0.70		0.71	0.63	0.70
ITGAM			0.64							0.62
KLRD1		0.68	0.86	0.71		0.70		0.73	0.67	0.72
L3MBTL4			0.55						0.56	0.76
NCAM1		0.67	0.76	0.78		0.84		0.74		0.53
NCR1		0.55	0.81	0.59		0.70		0.67	0.59	0.68
NKG7		0.54	0.81	0.64		0.58		0.61	0.56	0.71
PCDH1			0.60						0.60	0.70
PDGFD		0.55	0.81	0.64		0.66		0.58	0.59	0.67
PDGFRB										0.68
PRF1		0.72	0.85	0.65		0.68		0.77	0.67	0.70
PYHIN1		0.83	0.80	0.63		0.59	0.70	0.82	0.77	0.60
SH2D1A		0.58	0.53	0.55			0.65	0.58	0.57	
SLC4A10						0.61				
SLC4A4		0.59	0.82	0.68		0.66		0.63	0.65	0.67

SLCO4C1		0.60	0.73	0.65		0.64		0.63		0.70
SPINK2					0.57				-0.55	
TBX21		0.64	0.82	0.60		0.57		0.70	0.76	0.76
TGFBR3		0.72	0.80	0.61		0.54	0.56	0.76	0.74	0.72
TRGV9		0.51	0.67	0.56		0.77		0.60		0.62
TTC38			0.73	0.53		0.63		0.56	0.58	0.83
TYROBP			0.64			0.56				0.68
VWDE			-0.52						-0.58	-0.52
XCL1				0.56		0.73		0.51		
XCL2			0.57	0.59		0.77		0.55		

19.2 Correlation coefficients between expression of protein-coding and non-protein genes within *ex vivo* NK samples

Gene name	AC092580.4	AF131217.1	ANKRD20A11P	CASC15	EDDM3CP	FCGR2C	GS1-115G20.1	HCP5	LINC00861	LINC01108
ADAMTS1		0.79	0.75	-0.59		0.56	0.67		0.71	
BNC2		0.73	0.71	-0.65		0.75	0.84		0.77	
C1orf21		0.74	0.67	-0.70		0.69	0.77		0.70	
CAMK2A										0.69
CCNJL		0.62						0.55		
CLIC3		0.82	0.58	-0.67		0.68	0.52	0.60	0.85	
CMKLR1		0.69	0.74	-0.58		0.65	0.64		0.79	
CST7		0.55	0.70	-0.67		0.59	0.52	0.54	0.69	
CXCR1		0.56	0.77	-0.52		0.66	0.58		0.76	
DAB2			0.66			0.66	0.65			
DTX1			-0.53							
DUSP4			-0.66							
FCGR3A		0.76	0.72	-0.70		0.73	0.79		0.80	
FEZ1		0.70	0.77	-0.68		0.66	0.70	0.57	0.71	
FOSL2				-0.52	-0.55		0.52			
GNLY		0.73		-0.64				0.59	0.70	
GPR55							-0.53			
GZMA		0.82	0.72	-0.69		0.66	0.71		0.81	
GZMB		0.66	0.74	-0.69		0.67	0.64	0.66	0.78	
GZMH		0.67	0.65	-0.71		0.68	0.70		0.83	
IL18RAP		0.79		-0.57		0.53	0.63		0.55	
IL23R					0.51		-0.53			
ITGAM		0.66				0.55	0.54		0.67	
KLRD1		0.73	0.74	-0.63		0.53	0.54	0.72	0.73	
L3MBTL4		0.72	0.75					0.53	0.62	
MYO7A			-0.59	0.51			-0.65			

NCAM1		0.77		-0.56					0.68	
NCR1		0.71		-0.56				0.68	0.73	
NKG7		0.74	0.65	-0.70		0.69	0.59	0.58	0.88	
PCDH1		0.57	0.73			0.69	0.69		0.65	
PDGFD		0.84	0.60	-0.61		0.63	0.79		0.74	
PDGFRB			0.69			0.61			0.60	
PRF1		0.78	0.75	-0.66		0.69	0.64	0.63	0.83	
PYHIN1		0.81	0.79	-0.68		0.73	0.81		0.80	
SH2D1A		0.63	0.52	-0.61			0.74		0.50	
SLC4A10					0.55					
SLC4A4		0.81	0.54	-0.62		0.64	0.77		0.73	
SLCO4C1		0.74	0.75	-0.65		0.79	0.81		0.84	
SPINK2		-0.53	-0.71	0.72		-0.70	-0.76		-0.67	
TBX21		0.72	0.74	-0.73		0.69	0.72	0.53	0.76	
TGFBR3		0.64	0.69	-0.71		0.71	0.78		0.76	
TRGV9	0.69	0.55						0.75	0.56	
TTC38		0.70	0.76	-0.70		0.68	0.71	0.55	0.76	
TYROBP		0.64		-0.59		0.51		0.50	0.71	
VWDE		-0.57		0.78		-0.65	-0.62	-0.58	-0.65	
XCL2	0.54							0.54		
Gene name	**MAGI2-AS3**	**RP11-1094M14.5**	**RP11-222K16.2**	**RP11-277P12.20**	**RP11-366L20.2**	**RP11-452D12.1**	**RP11-456D7.1**	**RP11-686D22.10**	**RP11-693N9.2**	**RP11-81H14.2**
ADAMTS1		0.69	0.69				0.58	0.64	0.62	0.68
BNC2		0.80	0.67	0.56	-0.54		0.69	0.77	0.64	0.79
C1orf21	-0.55	0.79	0.68	0.51	-0.50		0.58	0.79	0.76	0.71
CCNJL	-0.51		0.54					0.56		
CLIC3	-0.54	0.68	0.78		-0.58		0.57	0.67	0.73	0.63
CMKLR1		0.77	0.68		-0.53		0.56	0.69	0.76	0.80
CST7		0.60	0.60		-0.52			0.59	0.65	0.65
CXCR1		0.64	0.55				0.51	0.53	0.59	0.74
DAB2		0.58						0.57		0.67
DUSP4										-0.53
FCGR3A	-0.58	0.82	0.77	0.55	-0.52		0.61	0.76	0.78	0.76
FEZ1	-0.51	0.80	0.69		-0.54		0.51	0.82	0.77	0.77
FOSL2								0.54		
GNLY			0.72	0.61			0.55		0.53	
GZMA	-0.53	0.85	0.86	0.57	-0.51		0.63	0.77	0.80	0.73
GZMB	-0.60	0.75	0.68		-0.65			0.75	0.76	0.80
GZMH		0.71	0.71	0.55	-0.56		0.59	0.64	0.73	0.72
IL18RAP	-0.57	0.69	0.70	0.52				0.80	0.57	
ITGAM		0.52	0.56					0.59		
KLRD1	-0.60	0.73	0.73	0.53				0.69	0.73	0.68

L3MBTL4		0.53	0.59				0.50		0.57	0.63
MYO7A		-0.57						-0.59	-0.54	-0.58
NCAM1	-0.51	0.57	0.69	0.70		0.58	0.59	0.56		
NCR1	-0.53	0.51	0.68					0.56	0.59	
NKG7		0.66	0.74		-0.51		0.55	0.61	0.70	0.70
PCDH1		0.67	0.53					0.60	0.59	0.73
PDGFD	-0.52	0.84	0.85	0.61			0.63	0.74	0.75	0.58
PDGFRB		0.57							0.54	0.71
PRF1	-0.59	0.78	0.77		-0.56		0.55	0.73	0.77	0.74
PYHIN1	-0.62	0.90	0.73	0.61	-0.56	0.56	0.71	0.86	0.79	0.81
SH2D1A		0.67	0.63	0.75			0.59	0.73	0.66	0.58
SLC4A4		0.83	0.81	0.63			0.63	0.78	0.76	0.59
SLCO4C1		0.82	0.68	0.57	-0.51		0.69	0.75	0.67	0.77
SPINK2	0.52	-0.69	-0.53		0.64		-0.53	-0.73	-0.64	-0.78
TBX21	-0.62	0.76	0.69	0.54	-0.60		0.60	0.75	0.77	0.80
TGFBR3	-0.54	0.80	0.66	0.57	-0.61		0.59	0.81	0.74	0.78
TRGV9			0.54						0.52	
TTC38	-0.58	0.75	0.66		-0.64		0.56	0.76	0.77	0.83
TYROBP			0.54				0.52			
VWDE	0.51	-0.56	-0.57		0.68			-0.67	-0.61	-0.72

19.3 Correlation coefficients between expression of protein-coding and non-protein genes within PB-derived NK samples

Gene name	AC022182.1	AC079949.1	AC092580.4	AF131217.1	ANKRD20A11P	EDDM3CP	FCGR2C	GS1-115G20.1	HCP5	LINC00299	LINC00861
ADAMTS1				0.76	0.75	0.71	0.51			0.64	0.57
ATP8B4				0.79		0.54					
BNC2				0.55	0.64	0.55		0.64			
C1orf21				0.56	0.64	0.55	0.51			0.56	
CAMK2A				-0.54		-0.53				-0.73	-0.57
CCNJL	-0.54										
CLIC3				0.53	0.56		0.58			0.71	0.77
CMKLR1				0.51	0.76	0.52				0.76	0.68
CST7			0.59		0.58				0.60		
CXCR1					0.70		0.52			0.62	0.71
DAB2				0.70	0.75	0.62	0.64	0.57		0.61	
DTX1					-0.57						
DUSP4				-0.58	-0.62	-0.72	-0.57			-0.63	-0.62
FCGR3A					0.70	0.58	0.50			0.66	0.57
FEZ1					0.67						
GZMA				0.59	0.65	0.65				0.54	

GZMB					0.58				0.56		0.52
GZMH			0.51							0.51	0.76
IL18RAP	-0.55			0.59							
ITGAM				0.52						0.60	
KLRD1					0.53				0.52		
L3MBTL4				0.76	0.66	0.72	0.63			0.59	0.65
MYO7A					-0.67						
N4BP3				-0.60	-0.63	-0.73	-0.54			-0.63	-0.67
NCAM1				0.61						0.66	
NCR1										0.53	0.51
NELL2					-0.53		-0.56			-0.61	
NKG7		-0.53	0.55		0.58					0.64	0.84
PCDH1					0.56		0.59			0.67	0.60
PDGFD				0.67	0.61	0.58	0.62	0.58		0.68	0.50
PDGFRB					0.71					0.74	0.80
PRF1				0.53	0.74		0.52			0.76	0.77
PYHIN1				0.66	0.77	0.64	0.56	0.53		0.52	
SH2D1A										-0.58	
SLCO4C1				0.62	0.73	0.64	0.63	0.52		0.73	0.65
SLC4A4				0.61	0.54	0.55	0.55			0.60	
TBX21					0.61		0.52			0.57	
TGFBR3					0.60						
TTC38					0.60					0.64	0.62
FOSL2				-0.59	-0.55	-0.74				-0.59	-0.66
TRGV9									0.74	0.52	0.63

Gene name	LINC01108	RP11-1094M14.5	RP11-222K16.2	RP11-277P12.20	RP11-452D12.1	RP11-456D7.1	RP11-686D22.10	RP11-693N9.2	RP11-81H14.2	RP5-857K21.11
ADAMTS1		0.67	0.53		0.63	0.51			0.52	-0.62
ATP8B4					0.81	0.54				
BNC2		0.71	0.58		0.51	0.54				-0.58
C1orf21	-0.53	0.86	0.62				0.57	0.67	0.65	
CAMK2A	0.78	-0.59	-0.75							
CLIC3	-0.74		0.56							
CMKLR1	-0.68	0.73	0.68					0.59	0.71	
CST7								0.55		
CXCR1	-0.58									
DAB2		0.82	0.56		0.54					-0.62
DTX1		-0.51		-0.57		-0.53			-0.61	
DUSP4	0.53	-0.72	-0.68		-0.59	-0.62		-0.58	-0.72	0.55
FCGR3A	-0.62	0.78	0.74					0.68	0.76	
FEZ1		0.69					0.64	0.57		
GZMA	-0.57	0.81	0.85					0.60	0.63	

GZMB				-0.56						
GZMH	-0.52									
IL18RAP							0.64			
ITGAM	-0.54									
L3MBTL4		0.58	0.58		0.60					-0.58
MYO7A		-0.78					-0.57	-0.68	-0.71	
N4BP3	0.54	-0.71	-0.73		-0.66	-0.68		-0.61	-0.70	0.54
NCAM1	-0.60				0.59					
NCR1	-0.57									
NKG7	-0.57		0.58							
PCDH1	-0.68	0.52								
PDGFD	-0.69	0.81	0.75		0.66	0.56				-0.57
PDGFRB	-0.59	0.54							0.62	
PRF1	-0.67	0.59	0.59							
PYHIN1	-0.55	0.96	0.71		0.62	0.55	0.59	0.60	0.56	-0.60
SLCO4C1	-0.65	0.81	0.68		0.58	0.55			0.60	-0.60
SPINK2									-0.58	
SLC4A4	-0.75	0.80	0.79		0.57			0.55		
TBX21		0.64						0.57	0.56	
TGFBR3		0.57							0.62	
TTC38	-0.58	0.64	0.62					0.63	0.67	
XCL1									-0.58	
FOSL2	0.55	-0.67	-0.75		-0.64	-0.60		-0.58	-0.64	
TRGV9	-0.55			-0.57						

19.4 Correlation coefficients between expression of protein-coding and non-protein genes within *ex vivo* stage 5 NK samples

Gene name	AF131217.1	ANKRD20A11P	EDDM3CP	FCGR2C	GS1-115G20.1	HCP5	LINC00299	LINC00861	LINC01108	MAGI2-AS3
ADAMTS1	0.66	0.77	0.54				0.52			
ATP8B4	0.84		0.73		0.51					
BNC2	0.56	0.56			0.60		0.60			
C1orf21	0.57	0.53								
CAMK2A									0.77	
CLIC3	0.53			0.50			0.60	0.78	-0.57	
CMKLR1	0.61	0.62					0.78	0.72	-0.58	
CST7		0.60				0.61				
CXCR1		0.68					0.67	0.70		
DAB2	0.74	0.66	0.52	0.57			0.74			
DTX1		-0.60								
DUSP4	-0.56	-0.62	-0.59				-0.67	-0.64		

FCGR3A	0.54	0.54					0.69	0.57		
FEZ1		0.66								
FOSL2	-0.52		-0.68							
GZMA	0.62	0.56	0.51				0.52			
GZMB		0.53				0.58	0.56			
IL18RAP	0.58									
ITGAM	0.54						0.73	0.62		
KLRD1		0.66								
L3MBTL4	0.71	0.68	0.61	0.52				0.54		
MYO7A										0.50
N4BP3	-0.54	-0.66	-0.63				-0.59	-0.58		
NCAM1	0.68		0.56	0.58			0.68	0.70	-0.58	
NCR1								0.63	-0.63	
NELL2		-0.57		-0.51			-0.64			
NKG7							0.65	0.83		
PCDH1							0.70	0.58	-0.53	
PDGFD	0.78		0.63				0.50		-0.70	
PDGFRB		0.66					0.76	0.81		
PRF1	0.51	0.70					0.71	0.69	-0.56	
PYHIN1	0.73	0.73	0.60				0.63			
SH2D1A							-0.57	-0.56	0.54	
SLC4A10									-0.51	
SLC4A4	0.71		0.64						-0.65	
SLCO4C1	0.69	0.62	0.58	0.63			0.85	0.73		
SPINK2										
TBX21		0.58					0.54			
TRGV9						0.66		0.55		
TTC38		0.59					0.70	0.58		
TYROBP							0.60	0.65		
Gene name	**RP11-1094M14.5**	**RP11-222K16.2**	**RP11-277P12.20**	**RP11-366L20.2**	**RP11-452D12.1**	**RP11-456D7.1**	**RP11-686D22.10**	**RP11-693N9.2**	**RP11-81H14.2**	**RP5-857K21.11**
ADAMTS1					0.61				0.50	
ATP8B4	0.61	0.60			0.78	0.60				-0.65
BNC2					0.56	0.52				
C1orf21	0.75				0.51		0.50	0.50	0.57	
CAMK2A	-0.50	-0.77								
CMKLR1	0.59	0.56			0.52				0.52	
CST7			-0.54							
DAB2	0.64				0.58					
DTX1									-0.60	
DUSP4					-0.67	-0.64			-0.67	
FCGR3A	0.64	0.58							0.60	

FEZ1							0.57			
FOSL2		-0.54			-0.64					
GZMA	0.69	0.71								
GZMB			-0.65							
IL18RAP							0.65			
L3MBTL4					0.61					
MYO7A	-0.57						-0.56			
N4BP3					-0.67	-0.62			-0.65	
NCAM1	0.60	0.52			0.57	0.52				
NELL2									-0.59	
NKG7										
PCDH1										
PDGFD	0.70	0.78			0.68					-0.53
PDGFRB									0.57	
PYHIN1	0.77				0.70		0.52			
SH2D1A			0.50							
SLC4A4	0.76	0.80			0.59					
SLCO4C1	0.65				0.67	0.58				
SPINK2				0.54					-0.50	
TTC38									0.56	
TYROBP										
XCL1									-0.50	

19.5 Correlation coefficients between expression of protein-coding and non-protein genes within nonlicensed *ex vivo* stage 5 NK samples

Gene name	AC022182.1	AC079949.1	AC092580.4	AF131217.1	ANKRD20A11P	CASC15	EDDM3CP	FCGR2C	GS1-115G20.1	HCP5	LINC00299	LINC00861
ADAMTS1				0.83	0.83		0.83					0.83
ATP8B4												
BNC2						0.89					0.89	
CAMK2A	0.81								-0.93			
CLIC3												0.94
CMKLR1												0.94

CST7									0.89		
DAB2								0.89			
DTX1		0.89		-0.94			-0.94	-0.89			
DUSP4		0.83		-0.89			-0.89	-0.94			
FCGR3A				0.83			0.83				0.83
FEZ1											
FOSL2		0.83		-0.94			-0.94	-0.89			
GNLY			0.83		1.00						0.89
GZMH										0.94	
ITGAM						0.94				0.94	
KLRD1					0.83						
L3MBTL4				0.83	0.83		0.83				0.83
MYO7A			-0.83								
N4BP3		0.94		-1.00			-1.00	-0.94			
NCAM1											0.89
NCR1						0.83				0.89	0.83
NELL2											
NKG7	0.83				0.83						
PCDH1						0.89				1.00	
PRF1					0.89						1.00

Gene name	LINC01108	MAGI2-AS3	RP11-1094M14.5	RP11-222K16.2	RP11-277P12.20	RP11-366L20.2	RP11-452D12.1	RP11-456D7.1	RP11-686D22.10	RP11-693N9.2	RP11-81H14.2	RP5-857K21.11
ADAMTS1		-0.83						0.83				
ATP8B4			0.89	0.94								
C1orf21			0.94	0.83	0.89		0.89			0.83		
CAMK2A				-0.90								
CCNJL						-0.82						
DTX1					-0.89		-0.89	-0.94				
DUSP4								-0.89				
FCGR3A			1.00		0.94		0.94	0.83				-0.89
FEZ1									0.89	0.89		
FOSL2					-0.89		-0.89	-0.94				
GNLY		-1.00										
PYHIN1				0.83			0.83					0.83
SH2D1A											-0.94	
SLC4A4									0.94			
TBX21		-0.83		0.89			0.89	0.94				
TTC38									0.83			
VWDE				0.88			0.88	0.88				
XCL2						-0.94					-0.94	

GPR55									0.82	0.88		
GZMA									0.89	1.00		
GZMB						-0.82						
IL23R			0.81		0.81		0.81			0.93		
ITGAM	-0.83											
KLRD1		-0.83										
L3MBTL4		-0.83						0.83				
MYO7A									-0.83			
N4BP3			-0.83		-0.94		-0.94	-1.00				0.83
NELL2						-0.94						
NKG7		-0.83										
PDGFD			0.89	0.94								
PRF1		-0.89	0.83									
PYHIN1			1.00		0.94		0.94	0.83				-0.89
SH2D1A											0.83	
SLC4A10				0.83						0.89		
SLC4A4				0.94								
SLCO4C1				0.89								
SPINK2				0.85								
TBX21								0.89				

TRGV9						-0.82						
VWDE			0.88	0.82	0.88		0.88	0.88				
XCL2	0.89										0.89	

19.6 Correlation coefficients between expression of protein-coding and non-protein genes within licensed *ex vivo* stage 5 NK samples

Gene name	AC022182.1	AC079949.1	AC092580.4	AF131217.1	ANKRD20A11P	CASC15	EDDM3CP	FCGR2C	GS1-115G20.1	HCP5	LINC00299	LINC00861
ADAMTS1			-0.62	0.75			0.70		0.63		0.76	
ATP8B4				0.98			0.94		0.82	-0.71	0.68	
BNC2					0.71							
C1orf21				0.77			0.78		0.95	-0.88		
CAMK2A				-0.87			-0.91			0.63		
CLIC3	0.62											
CMKLR1											0.64	
CST7				-0.77			-0.77		-0.78	0.75		
CXCR1												0.64
DAB2			-0.73	0.83	0.73		0.84		0.74	-0.61	0.75	
DTX1					-0.80							
DUSP4				-0.81			-0.76		-0.76	0.75		
FCGR3A			-0.73	0.87			0.82		0.82		0.76	

GPR55								-0.67	-0.70			
GZMA				0.77			0.68		0.78	-0.69		
GZMB										0.68		
IL18RAP				0.67					0.66			
IL23R	0.66	-0.68										
ITGAM	0.73	-0.65										0.85
MYO7A						0.81						
N4BP3				-0.88			-0.82		-0.75	0.61	-0.69	
NCAM1				0.69			0.72				0.71	
NCR1												
NELL2					-0.77						-0.69	
PCDH1	0.76	-0.69										0.68
PDGFD				0.96			0.92		0.83	-0.72	0.70	
PDGFRB	0.84	-0.67						0.75			0.73	0.84
PRF1	0.66											
PYHIN1				0.84			0.79		0.92	-0.83		
SH2D1A	-0.63											-0.80
SLCO4C1			-0.67	0.86	0.68		0.85	0.64	0.72	-0.67	0.76	
SLC4A4				0.85			0.83		0.70	-0.69		
TGFBR3					0.75							

GNLY	FCGR3A	DUSP4	DTX1	DAB2	CXCR1	CST7	CMKLR1	CLIC3	CAMK2A	C1orf21	BNC2	ATP8B4	ADAMTS1	Gene name	TRGV9	FOSL2	XCL2	XCL1
													-0.62	**LINC01108**				
														MAGI2.AS3				
	0.65		-0.66	0.75		-0.75		-0.62		0.90		0.65		**RP11-1094M14.5**				
									-0.64					**RP11-222K16.2**		-0.77		
			-0.66			-0.97			-0.72	0.78		0.75		**RP11-277P12.20**				-0.76
	0.88	-0.90		0.84		-0.69			-0.78	0.86		0.94	0.74	**RP11-452D12.1**				
	0.86	-0.83		0.84		-0.82			-0.88	0.83		0.95	0.75	**RP11-456D7.1**		-0.72		
-0.81						-0.61		-0.79						**RP11-686D22.10**			-0.65	
-0.66								-0.64						**RP11-693N9.2**	-0.61	-0.72		
	0.68			0.63			0.77				0.69		0.70	**RP11-81H14.2**	0.62	0.78		
				-0.64		0.65			0.75			-0.75		**RP5-857K21.11**				

GPR55						-0.72					
GZMA				0.61		0.84	0.73				-0.62
GZMB			-0.67		-0.85			-0.74			
GZMH			-0.72					-0.79			
IL18RAP						0.66					
ITGAM		0.83									
KLRD1										-0.65	
MYO7A	-0.66		-0.68					-0.76	-0.69		
N4BP3	0.62		-0.61		-0.69	-0.82	-0.90			-0.75	
NCAM1	-0.66						0.72				
NCR1					-0.65			-0.73			
NELL2			-0.63							-0.69	
NKG7								-0.65			
PCDH1								-0.73	-0.65		
PDGFD			0.65		0.74	0.95	0.94				-0.75
PDGFRB		0.71									
PRF1								-0.79	-0.69		
PYHIN1			0.85		0.86	0.83	0.85				-0.66
SH2D1A		-0.71									
SLC4A10	-0.67							-0.63	-0.69		

SLCO4C1			0.65		0.71	0.85	0.89				-0.61
SPINK2			-0.66					-0.75			
SLC4A4			0.66		0.75	0.80	0.87			0.64	-0.70
TYROBP								-0.66			
VWDE					0.66						
FOSL2				-0.63		-0.87	-0.79				
TRGV9			-0.75		-0.68			-0.65			

19.7 Correlation coefficients between expression of protein-coding and non-protein genes within memory-like *ex vivo* stage 5 NK samples

Gene name	AC079949.1	AC092580.4	AF131217.1	ANKRD20A11P	CASC15	EDDM3CP	FCGR2C	GS1-115G20.1	HCP5	LINC00299	LINC00861	LINC01108
ADAMTS1			0.82									
ATP8B4	0.79											
C1orf21		0.82		0.86						0.86	0.86	
CAMK2A		-0.93		-0.82				0.82	-0.85	-0.85	-0.93	
CLIC3				0.86						0.82		
CMKLR1		0.79		0.79				-0.86	0.82	0.82	0.79	
CST7		0.86						-0.82	0.86	0.86	0.86	
CXCR1		0.82		0.82					0.89	0.89	0.93	
DAB2			0.86									
DUSP4		-0.86		-0.86					-0.86	-0.86	-0.96	
FCGR3A		1.00						-0.86	0.82		0.82	-0.79
FEZ1				0.86						0.79		
FOSL2		-0.96						0.93	-0.86			0.89
GZMA		0.93		0.79						0.82	0.89	
GZMB				0.93						1.00	0.93	
GZMH		0.93						-0.96	0.89			-0.82
IL18RAP			0.82									
ITGAM				0.86				-0.79	0.86	0.86	0.86	
KLRD1				0.96						0.86	0.86	
L3MBTL4				0.93						0.93	0.79	
MYO7A				-1.00						-0.93	-0.89	

N4BP3		-1.00						0.86	-0.82		-0.82	0.79
NCAM1		0.93							0.79	0.79	0.93	
NCR1		0.79		0.89					0.82	0.82	0.89	
NELL2				-0.79						-0.86		
NKG7		0.82		0.82					0.89	0.89	0.93	
PCDH1				0.96						0.96	0.86	
PDGFD		0.79						-0.86	0.93			-0.93
PDGFRB		0.79		0.86					0.82	0.96	0.96	
PRF1				0.96						0.96	0.86	
PYHIN1		0.79		0.79						0.79	0.93	
SH2D1A		-0.79						0.96	-0.93			0.82
SLC4A4		0.89						-0.89	0.82			-0.93
SLCO4C1		0.86		0.86					0.86	0.86	0.96	
SPINK2				-0.76		-0.80				-0.76		
TBX21					-0.85			-0.86	0.82			
TGFBR3				0.86						0.82	0.89	
TRGV9							0.86		0.79			-0.93
TTC38				0.96						0.96	0.86	
TYROBP		0.86						-0.82	0.96		0.82	-0.89

Gene name	MAGI2-AS3	RP11-1094M14.5	RP11-222K16.2	RP11-277P12.20	RP11-366L20.2	RP11-452D12.1	RP11-456D7.1	RP11-686D22.10	RP11-693N9.2	RP11-81H14.2	RP5-857K21.11
ADAMTS1								0.79			
BNC2								0.93			
C1orf21		0.79		-0.79		-0.86				0.82	
CAMK2A	0.78						0.85		-0.78	-0.93	
CLIC3	-0.79			-0.96		-0.86					
CMKLR1	-0.89						-0.96			0.86	
CST7	-0.79					-0.79	-0.89		0.79	0.93	
CXCR1							-0.86			0.96	
DUSP4							0.79			-0.93	
FCGR3A			0.82						0.89	0.79	0.86
FEZ1	-0.82			-0.86		-0.86					
FOSL2			-0.89				0.79		-0.93		
GPR55			-0.86						-0.79		
GZMA						-0.79				0.86	0.86
GZMB				-0.93		-0.93				0.96	
GZMH	-0.79		0.86				-0.86		0.96		
ITGAM	-0.82			-0.79			-0.93			0.89	
KLRD1	-0.79	0.86		-0.86		-0.79				0.79	
L3MBTL4				-1.00		-0.93				0.82	
MYO7A		-0.82		0.93		0.82				-0.86	
N4BP3			-0.82						-0.89	-0.79	-0.86

NCAM1										0.86	0.79
NCR1	-0.79						-0.86			0.86	
NELL2				0.86		0.79					
NKG7							-0.86			0.96	
PCDH1	-0.79			-0.96		-0.89	-0.79			0.89	
PDGFD							-0.86				
PDGFRB				-0.82		-0.86				1.00	
PRF1	-0.79			-0.96		-0.89	-0.79			0.89	
PYHIN1		0.89								0.82	
SH2D1A	0.86						0.96		-0.86		
SLC4A4			0.86					-0.79	0.86		
SLCO4C1							-0.79			0.93	
SPINK2		-0.76		0.80		0.80					
TBX21	-0.82						-0.96				
TGFBR3		0.96								0.79	
TRGV9				-0.79							
TTC38	-0.79			-0.96		-0.89	-0.79			0.89	
TYROBP							-0.79			0.79	
XCL2						0.79					

CV

Name: Olena Zaitseva

Residence: Gosslerstrasse 33, 37077 Göttingen, Germany

Telephone: +49 (0) 17661344781

E-Mail: ozaitseva89@gmail.com

Date and place of birth: 10.05.1989, Kyiv, Ukraine

EDUCATION:

10/2012 – 05/2016 PhD in Molecular Biology (expected date of defense 19.05.2016), Georg-August University Göttingen

- PhD thesis - "Expression profiling of natural killer lymphocytes"

09/2011 – 10/2012 M.Sc. in Molecular Biology, Georg-August University Göttingen, International Max Plank Research School in Molecular biology.

- Master's thesis - "Analysis of human KIR gene transcription at single cell level"

09/2006 –07/2010: **B.Sc.** in biology, **Minor Certificate** in ecology, "Kyiv-Mohyla academy" National University of Ukraine

- Bachelor's thesis - "Protein kinase D2, protein kinase Cβ and transcription factor IPO-38 as potential molecular markers of diffuse large B-cell lymphoma"

WORK EXPERIENCE:

10/2012 – 05/2016: Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Prof. Dr. Lutz Walter.

- Ph.D. study: Analysis of natural killer lymphocyte expression profiles; analysis of transcription regulation of killer-cell immunoglobulin-like receptor genes.

05/2012 – 06/2012: Gene Expression group, Max Plank Institute for Experimental Medicine, Göttingen, PD Dr. Moritz Rossner.

- Lab Rotation: Cloning and functional characterization of biosensors for cellular signaling pathways.

03/2012 – 04/2012: Gene Expression and Signaling group, Max Plank Institute for Biophysical Chemistry, Dr. Halyna Scherbata.

- Lab Rotation: The roles of HOW, ecdysone steroid hormone and their interaction during ovaria development and maintenance in *Drosophila melanogaster*.

01/2012 – 02/2012: Department of Molecular Oncology, Georg-August University Göttingen, Prof. Dr. Matthias Dobbelstein.

- Lab Rotation: The role of MAPKAP2 in DNA damage response.

10/2008 – 09/2011: Department of Cell Regulatory Mechanisms, Institute of Experimental Pathology Oncology and Radoibiology of National Academy of Sciences of Ukraine, Kyiv, Prof. Dr. Svetlana Sidorenko.

- Volunteer Research Assistant: Searching for new molecular markers of diffuse large B-cell lymphoma.

08/2007 – 07/2008: Department of Translation Mechanisms, Institute of Molecular Biology and Genetics of National Academy of Sciences of Ukraine, Kyiv, Prof. Dr. Boris Negrutskii.

- Volunteer Research Assistant: Non-conventional role of eukaryotic translation elongation factor 1A (eEF1A).

PROFESSIONAL SKILLS:

High-throughput microfluidic techniques: BioMark gene expression analysis, single cell qPCR, single cell mRNA sequencing.

Immunological techniques: isolation of human PBMCs, magnetic cell separation, Western blot analysis, immunohistochemical staining, flow cytometry.

Microbiology and cell biology techniques: cell culture, bacterial culture, plasmid and siRNA transfection, bacterial transformation, work in S1 and S2 laboratories.

DNA and RNA techniques: Sanger sequencing, Illumina sequencing library preparation, PCR, RT-PCR, molecular cloning, DNA and RNA purification.

Drosophila techniques: fly work, planning and conducting of crosses.

Microscopy techniques: light, fluorescent, confocal microscopy.

Analysis of mRNA sequencing data.

AWARDS AND SCHOLARSHIPS

- **2016**: GGNB Bridging Fund
- **2013**: GGNB Travel Grant.
- **2012**: GGNB Travel Grant.

- **2011 – 2012:** Stipend of the Excellence Foundation for the Promotion of the Max Planck Society.

CONFERENCES AND WORKSHOPS

2015: Natural Killer Cell Symposium: poster "Expression profiling of NK lymphocytes at the single cell level".

2013: International Student Conference "Horizons in Molecular Biology". Poster "Analysis of human killer immunoglobulin-like receptors' gene transcription at single cell level"

2013: Killer Immunoglobulin-Like Receptor Workshop. Talk "Analysis of human KIR gene transcription at single cell level"

PERSONALL SKILLS:

Teaching experience: tutorials for IMPRS Molecular Biology master's students; method courses for GGNB PhD students; supervision of master's student projects.

Organization skills: 2014-2016 – member of DPZ PhD Colloquium; 2007-2009 - head of Student's Biological Organization of "Kyiv-Mohyla academy" National University.

Languages: Ukrainian, Russian – native; English – fluent; German – C1 level.