Natural Killer (NK) cell receptor expression in cattle: development of a pipeline for accurate quantification of RNA Sequencing (RNA-Seq) data

Richard Borne

BSc. (Hons)

Submitted in fulfilment of the requirements for the Degree of PhD

College of Medical, Veterinary & Life Sciences

University of Glasgow

September 2018

Abstract

Cattle have undergone significant gene expansion within two of the major NK receptor encoding complexes, the leukocyte receptor complex (LRC) and natural killer complex (NKC). This expansion has resulted in a number of highly similar genes densely packed within each complex. The genes can be highly polymorphic and encode for fundamentally important receptors for controlling the functional response of NK cells. Understanding the nature of transcription of the LRC and NKC genes is required to confirm existing gene models, predicted functional status and also their distribution in cell and tissue types. RNA-Seq offers the potential for high-resolution analysis of LRC/NKC gene transcription. For this to occur, an analysis pipeline needed to be developed to account for the high sequence similarity and resulting multi-mapping of reads.

In this project an analysis pipeline, UniMMap, has been created that utilises the concept of mappability to weight the contribution of multi-mapping reads to the total read count. UniMMap was utilised to assess transcription of LRC/NKC genes in a number of RNA-Seq datasets. Transcription in peripheral blood mononuclear cells (PBMCs) and NK cells from two cattle was compared. Genes predicted to be non-functional based on the reference assembly were found to be consistently transcribed. Transcription was also compared between multiple immune cell types. CD8+ T cells were shown to transcribe killer-cell immunoglobulin-like receptor (KIR) at an equivalent level to NK cells. RNA-Seq data from multiple cattle tissues was analysed to produce an LRC/NKC gene atlas. Transcription of at least one LRC and NKC gene was observed in every tested tissue. Total transcription of the LRC/NKC genes was found to be highest in the lymph node, mammary gland and lung tissue. To assess the utility of UniMMap in a controlled study, as well as in a different species, the transcription of LRC/NKC genes were investigated in goats from a peste des petits ruminants virus (PPRV) vaccination study. A subset of the genes were found to respond differently in PBMCs between unvaccinated and vaccinated animals. This supports a role for these genes during PPRV infection.

UniMMap has facilitated analysis of transcription of LRC/NKC genes from multiple species, cell/tissue types and RNA-Seq library types. The consistently observed transcription of predicted non-functional genes suggests the presence of functional alleles or possibly a role in regulating transcription of other genes

within each complex. The changes observed in their transcription during infection, coupled with a similar distribution among tissues and immune cell types, suggests a functional similarity with their human counterparts.

Table of Contents

List of figures and tables

Figures

Tables

Acknowledgements

I would like to thank both of my Pirbright supervisors, John Hammond and Paolo Ribeca, for their guidance, support, and patience over the course of the project. I have learnt a great deal over the last four years and appreciate being able to take the project in the direction I found most interesting.

Thanks to my Glasgow supervisor Simon Babayan, who agreed to step in midway through the project and helped with having to navigate the University processes from hundreds of miles away.

The members of the Immunogenetics group past and present, who have provided much help and support. In particular, Alasdair Allan for his guidance during my first year. Dorothea Harrison and John Schwartz for their expertise and sharing in the joy of bioinformatics analyses of the LRC and NKC. Kevan Hanson and William Mwangi for their assistance with isolation of NK cells and subsequent RNA extractions. Veronica Carr for carrying out the PBMC isolations inside containment for the goat PPRV vaccination study. Graham Freimanis for his help creating the protocol to bring goat PBMC samples out of containment.

Thanks to Jessica Powell and Liam Morrison (Roslin Institute) and Azad K. Kaushik (University of Guelph) for sharing their unpublished RNA-Seq data.

Many thanks to Karin Darpel for allowing us to collect samples from the Goat PPRV vaccination study, and sharing information regarding the results of the study with us.

I would like to thank the BBSRC for providing the funding for the project.

Finally, many thanks to the Borne family for their support and encouragement over the years.

Author’s declaration

This thesis, and the work contained within it, was conducted from October 2014 to September 2018 by myself, unless stated otherwise. No part of this thesis has been submitted for another degree.

Richard William Borne

List of abbreviations

ADCC	Antibody-dependent cell-mediated cytotoxicity
FAS	FAS ligand
GEM	Genome multitool
HLA	Human leukocyte antigen
IFN-y	Interferon-gamma
Ig	Immunoglobulin
ITAMs	Immunoreceptor tyrosine-based activation motif
ITIMs	Immunoreceptor tyrosine-based inhibition motif
KIR	Killer-cell immunoglobulin-like receptor
KLR	Killer-cell lectin-like receptor
LILR	Leukocyte immunoglobulin-like receptors
LPS	lipopolysaccharide
LRC	Leukocyte receptor complex
MHC	Major histocompatibility complex
mya	million years ago
NCR1	Natural cytotoxicity triggering receptor 1
NK	Natural killer
NKC	Natural killer complex
NKR	Natural killer receptor
RPM	Reads per million
rRNA	ribosomal RNA
SRA	Sequence read archive
TCR	T-cell receptor
TRAIL	TNF-related apoptosis-inducing ligand

1 Chapter 1. Introduction

1.1 Natural killer cells

Natural killer (NK) cells are cytotoxic lymphocytes, derived from the same progenitor as B and T cells. The cytotoxic nature of NK cells is tightly controlled by a diverse array of germline-encoded activating and inhibitory receptors. Unlike B and T cells, these receptors do not undergo rearrangement to become antigen-specific. Consequently, NK cells are components of the innate immune system. As well as their cytotoxic abilities, NK cells are able to produce a number of cytokines and chemokines (Cooper, Fehniger, and Caligiuri 2001; Dorner et al. 2004). The release of cytokines and other immunomodulators, which influence other immune cells, allows them to serve as a bridge between the innate and adaptive immune system. NK cells also play a critical role in controlling viral infections, tumour surveillance and placentation.

1.1.2 Cell surface phenotype of NK cells

Human NK cells are typically defined as $CD3^-/CD56^+$ as the absence of CD3 excludes T cells and CD56 is found only on NK cells and a minority of T cells. Expression levels of CD56 allow the division of NK cells into functional subsets. $CD56^{bright}$ NK cells are more predominant cytokine producers and $CD56^{dim}$ are better at killing (Cooper, Fehniger, and Caligiuri 2001). Expression of CD16 is an additional marker of NK cells, although a small fraction of $CD56^{bright}$ cells are $CD16^-$ (Lanier et al. 1986). However, this definition of an NK cell does not apply across other species. In the rhesus macaque (*Macaca mulatta*), which diverged from the ancestors of humans 25 million years ago (Gibbs et al. 2007)CD56+ cells are a marker for monocytes and not NK cells (Carter et al. 1999), expression of CD56 is absent in mice (Hayakawa et al. 2006). Cattle NK cells have been commonly described as $CD3^-/CD2^+$ lymphocytes. More recently, use of the activating receptor NKp46/NCR1 as the defining marker of NK cells across species, including humans, mice and cattle has emerged (Walzer et al. 2007; Westgaard et al. 2004). Use of NKp46/NCR1 as an NK marker is not perfect however. In humans, NKp46/NCR1 is also found on a small subset of cytotoxic T lymphocytes and $CD3^-CD56^+$ cells have been observed with low or absent expression (Meresse et al. 2006; Caligiuri 2008). It is however currently the best known NK marker and has enabled the characterisation of NK cells in multiple

species. In cattle, both NKp46^{+}/CD2^{+} and NKp46^{+}/CD2^{-} NK cells have been reported (Storset et al. 2004). Both cattle NK types possess a full range of NK activity, NKp46^{+}/CD2^{-} NK cells are more proliferative and possess higher interferon-gamma (IFN-γ) activity (Boysen et al. 2006).

1.1.3 Major histocompatibility complex class I

The major histocompatibility complex (MHC) class I molecule is found on the cell surface of all nucleated cells in the jawed vertebrates (Kulski et al. 2002). Cytosolic proteins are degraded by the proteasome and then bound and presented by MHC class I, which is then transported to the cell surface. In humans, the MHC proteins are encoded by the human leukocyte antigen (HLA) gene complex. There are six different functional HLA class I genes (A-C and E-G) in total and an individual possesses at least three. These genes (particularly A-C) are highly polymorphic, the latest release (2018-07) of the IPD-IMGT/HLA database contains sequences of 13,680 HLA class I alleles (Robinson et al. 2015). Due to the large number of alleles, the vast majority of individuals are heterozygous, typically expressing six different MHC class I molecules.

CD8 expressed by cytotoxic T cells binds MHC class I in a manner that also allows binding of the T-cell receptor, which can sense antigenicity. This allows killing of cells presenting non-self-peptide, such as those of viral origin, by CD8+ T cells. Polymorphism of HLA class I introduces differences in the peptide binding of the encoded MHC class I molecules. This results in different binding specificities and efficacies to individual peptides as well as altering the contact between the molecule and T-cell receptors. MHC class I is also recognised by many of the receptors that inhibit NK cytotoxicity. Binding of an inhibitory receptor to MHC class I serves to prevent the NK cell from killing host cells. Many pathological conditions result in down-modulation of MHC class I molecules, removing the inhibitory effect and enabling NK cell activation. The process by which NK cells are ‘taught’ to recognise MHC class I is known as NK cell education. The influence of the MHC class I diversity upon the inhibitory NK receptors is not well understood. The HLA class I alleles present in an individual have been shown to increase the frequency of NK cells expressing cognate inhibitory receptors (Yawata et al. 2006). Variation in the strength of the inhibitory signal produced by the inhibitory receptors means that driving expression of a certain receptor can influence the ability of NK cells to respond.

1.1.4 Mechanisms of NK cell killing

NK cell cytotoxicity is tightly controlled by the balance of activating and inhibitory receptors on their cell surface. Activation of NK killing occurs through either a decrease in inhibitory signalling or an increase of ligand binding of the activating receptors. The presence of a large number of activating and inhibitory receptors on the surface of NK cells enables them to be controlled by a variety of external stimuli. A decrease in inhibitory signalling typically occurs through down-modulation of MHC class I on the target cell, as is the case in many tumour cells. Activating signals can be generated by engagement of activating receptors to their ligand, as happens in antibody-dependent cell-mediated cytotoxicity (ADCC). ADCC involves recognition of the Fc portion of an antibody bound to the surface of an infected cell by an Fc receptor such as CD16. Upon activation, NK cells employ one of two major mechanisms of killing, both of which require direct contact with the target cell. The first mechanism involves the release of cytotoxic granule proteins into the intracellular space. The two major granule proteins are perforin and granzyme. Perforin forms a pore in the target cell surface which enables entry of granzyme into the target cell (Catalfamo and Henkart 2003). Once inside the target cell cytosol, granzyme induces caspase-dependent apoptosis (Trapani and Sutton 2003). The second mechanism involves engagement of death receptor pathways. TNF-related apoptosis-inducing ligand (TRAIL) is a transmembrane protein able to transduce an apoptotic signal (Degli-Esposti 1999). FAS ligand (FasL) is also a transmembrane protein, belonging to the same family as TRAIL and induces apoptosis after binding to its receptor. Engagement of these death receptor pathways provides a perforin-independent method of cytotoxicity to NK cells (Wallin et al. 2003). NK cells have been shown to prime CD8+ T cells to kill *M. tuberculosis* infected monocytes. This occurs via interactions between CD40 expressed on antigen presenting cells and CD40 ligand on the surface of NK cells. Secretion of IFN-γ by NK cells also results in expansion of cytotoxic CD8+ T cells (Vankayalapati et al. 2004).

1.1.5 Protection against tumour cells by NK cells

Depletion of NK cells has been used to demonstrate their crucial role in defence against tumour cells (Ljunggren and Kärre 1985; Seaman et al. 1987). The presence of NK cells in tumour biopsies is linked to a favourable prognoses in patients diagnosed with cancer (Coca et al. 1997; Ishigami et al. 2000). NK cells

are a major producer of IFN-γ which has been shown to possess a number of anti-tumour functions. Tumour cells generated to be insensitive to IFN-γ have been shown to display enhanced tumourigenicity (Dighe et al. 1994).

1.1.6 Pregnancy and NK cells

Uterine NK (uNK) cells possess a unique cell surface phenotype and play a role in menstruation and pregnancy (Moffett-King 2002; Gaynor and Colucci 2017). Interactions between the foetal trophoblast and maternal NK cells help to form the placenta after implantation of the embryo into the uterine wall. Invasion of extravillious trophoblast cells (EVTs) occurs soon after, EVTs begin enlarging spiral arteries to provide adequate blood flow to the placenta for the foetus. Adequate EVT invasion is required to prevent pre-eclampsia in humans. The extent of placental invasion is thought to be controlled by uNK cells (Moffett-King 2002).

1.1.7 NK cells as immune regulators

NK cells have a multitude of regulatory roles within both the innate and adaptive immune systems (figure 1-1). NK cells have a synergistic relationship with dendritic cells (DC). Dendritic cells have been shown to promote cytolytic activity and IFN-γ production in resting NK cells (Fernandez et al. 1999). NK cells regulate both the maturation and functional activity of DCs in an IFN-γ-dependent manner (Pollok et al. 1993). NK cells also kill immature DCs in a perforin-dependent manner. Removal of immature DCs results in selection for a more immunogenic population of DCs with an improved ability to induce T cell proliferation (Morandi et al. 2012).

As with DCs, macrophages also have a synergistic relationship with NK cells. Macrophages have been shown to prime NK cells for cytotoxicity, activate proliferation and cytokine secretion and increase expression of activating NK receptors (Nedvetzki et al. 2007). Bacterial lipopolysaccharide (LPS) activates macrophages leading to the production of a range of inflammatory cytokines. However at high doses of LPS they become unable to respond to further stimulation. Too much cytokine produced in response to stimulation by LPS is thought to be the cause of septic shock (Fujihara et al. 2003). Macrophages stimulated by high doses of LPS upregulate stress-inducible class I MHC-like

ligands. These ligands are recognised by the activating receptor NKG2D on NK cells, resulting in killing of the macrophages (Nedvetzki et al. 2007).

NK cells prime and promote differentiation of T cells via the secretion of cytokines (Morandi et al. 2006; Wehner et al. 2009). NK cells are also capable of directly killing T cells. Killing of T regulatory cells expanded in response to infection with *Mycobacterium tuberculosis* has been shown (Roy et al. 2008). Killing of antigen-activated T cells by NK cells via NKG2D has also been observed (Cerboni et al. 2007). CD4+ T cells mediate chronic inflammation in the colon (colitis), transferring CD4+ T cells into recombination-activating gene (Rag) 1 and 2 knockout mouse recipients causes colitis in an NK-depleted recipient. However, colitis was not observed in the non-depleted recipients, demonstrating that NK cells control the responses of T cells to the gut microbiota (Fort, Leach, and Rennick 1998).

1.2 NK cell education

Due to the highly cytotoxic nature of NK cells, they must be educated so that they do not respond inappropriately towards host cells, they must acquire self-tolerance (Raulet and Vance 2006). Multiple models have been proposed to explain how this process occurs. The arming model suggests the ability of an NK cell to recognise self-major histocompatibility complex (MHC) class I is vital for it to become a functional effector cell (Kadri, Luu Thanh, and Höglund 2015; Yokoyama and Kim 2006). Recognition of MHC class I typically occurs via killer cell immunoglobulin-like receptors (KIR) in humans and the killer cell lectin-like receptor (KLR) Ly49 in mice. The disarming model is based on the observation that a lack of inhibitory receptors to balance the chronic activating stimulations results in hypo-responsive NK cells. NK cells from Ly49 or MHC class I deficient mice are unable to kill MHC class I deficient cells (Belanger et al. 2012). Subsequently, they are unable to carry out surveillance for downregulation of MHC class I, often occurring in tumour cells (Bubeník 2003). The rheostat model includes facets of both the arming and disarming models. This model suggests that the responsiveness of NK cells is controlled quantitatively by the strength of the inhibitory signal (Brodin et al. 2009). The most recently described model for NK cell education is the confining model. The name is derived from the observation that the activating receptors of educated NK cells are confined in nanodomains at the cell surface. In hypo-responsive NKs, receptors are dispersed

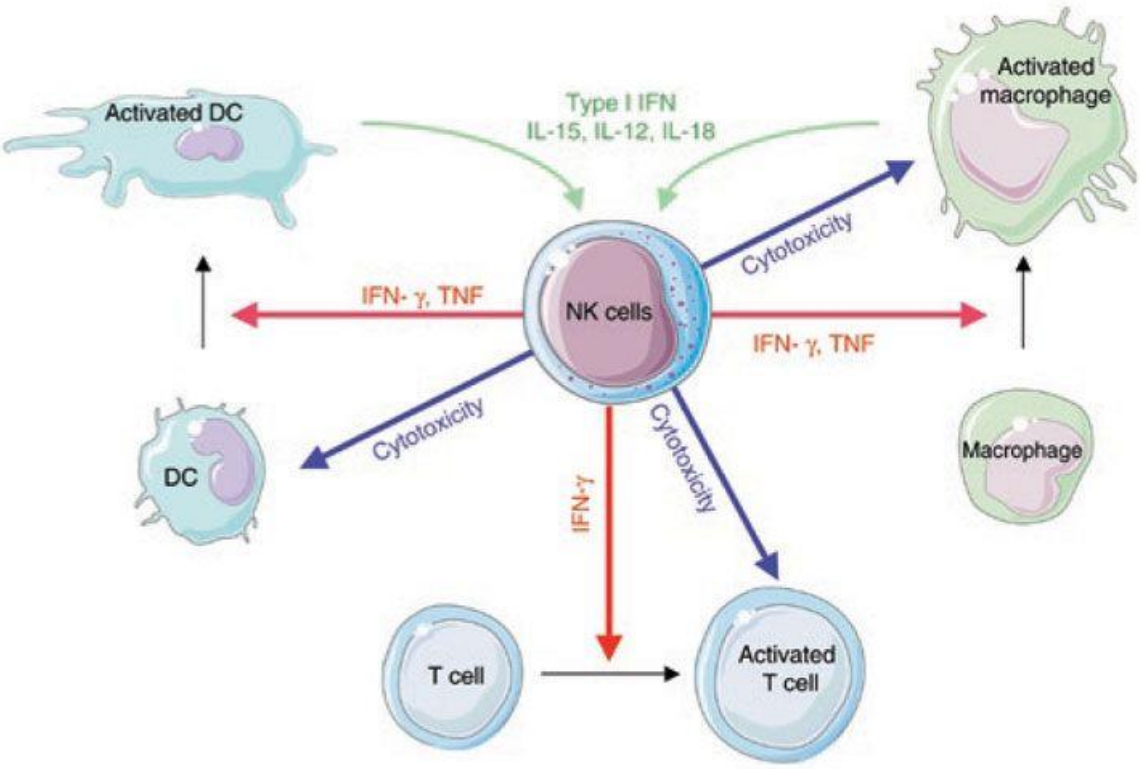


Figure 1-1. The influence of NK cells on immune responses. Blue arrows indicate killing of the cell by NK cells. After priming by other cell types (green arrows), NK cells regulate other immune cells via release of cytokines/chemokines. Taken from (Vivier et al. 2008).

into the actin meshwork (He and Tian 2017; Guia et al. 2011). Regardless of the correctness of the various models, it is clear that NK cell function is dependent on the process of NK cell education.

1.3 NK receptors in model organisms

The receptors responsible for recognition of MHC class I have been extensively studied in both humans and mice. These MHC class I specific receptors are typically encoded within either the leukocyte receptor complex (LRC) or the natural killer complex (NKC). The LRC, NKC and MHC genes are located on three separate chromosomes in all studied species. Convergent evolution has given rise to NK receptors of different structural families that recognise or are predicted to recognise MHC class I in multiple evolutionary distant species. This suggests that they have been subject to rapid evolution and highlights their vital importance.

1.3.1 The killer cell immunoglobulin-like receptors

Humans have undergone significant gene expansion within the LRC. Encoded within the LRC are the leukocyte immunoglobulin-like receptors (LILR), a subset of which recognise MHC class I (Jones et al. 2011). The LILR are primarily expressed on the surface of macrophages and dendritic cells as well as B cells and subsets of NK and T cells (Porwit, McCullough, and Erber 2011). Subsequently they are not considered NK specific receptors. Also encoded within the LRC are killer cell immunoglobulin-like receptors (KIRs), type I transmembrane glycoproteins that belong to the immunoglobulin (Ig) superfamily. KIR typically contain two or three Ig domains named D0, D1 and D2 based on their location from n-terminus to c-terminus. KIR receptors come in two forms, activating or inhibitory. Inhibitory KIR possess a cytoplasmic tail containing two immunoreceptor tyrosine-based inhibition motifs (ITIMs), which produce an inhibitory signal to the cell after receptor engagement. The activating KIR possess a shorter cytoplasmic tail and contain either a positively charged arginine or lysine in their transmembrane domain. The positively charged arginine/lysine aids with binding of a transmembrane accessory protein such as DAP10, DAP12, FcRγ and TCRζ which contain immunoreceptor tyrosine-based activation motifs (ITAMs) or YxxM motifs (Lanier et al. 1998; Lanier 2003). The number of domains (2D or 3D), signalling ability (L for long tail inhibitory, S

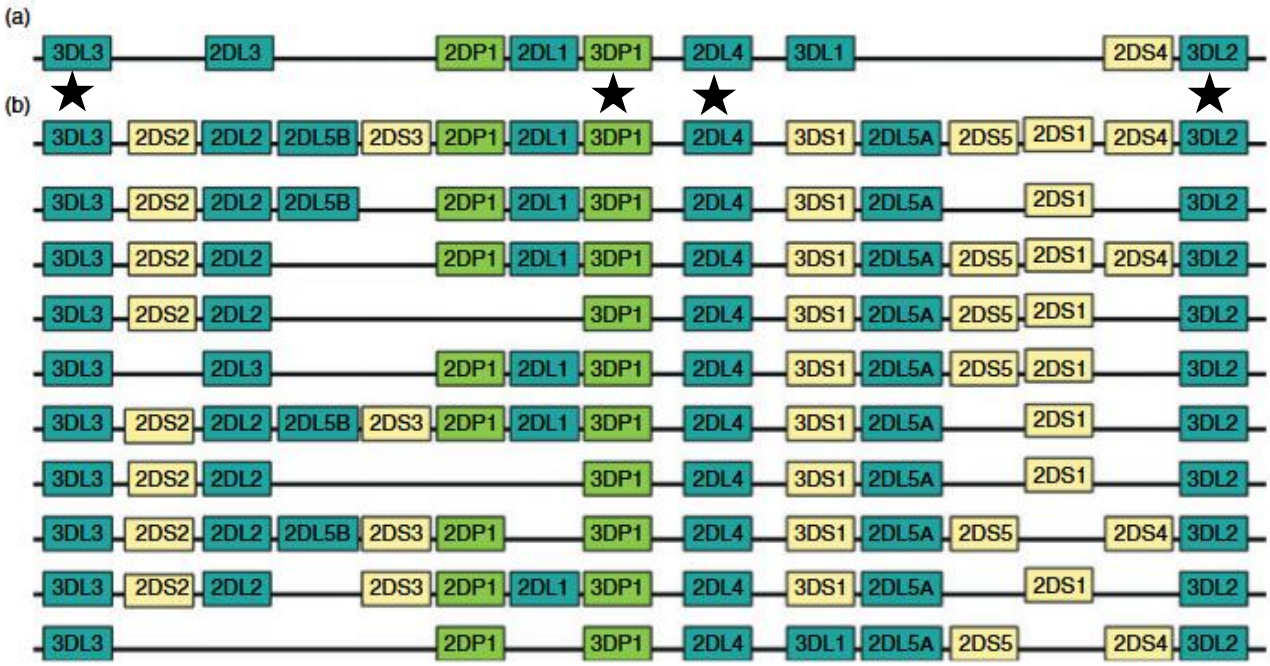


Figure 1-2. Variation of gene content within the two human haplotype groups (A and B). Black stars indicate the framework genes. Taken from (Middleton and Gonzelez 2010).

for short tail activating and P for pseudogene) and order of discovery give rise to the name of the receptor. For example, the two domain inhibitory KIR receptor which was first discovered, is named 2DL1. Mice and rats have not expanded the LRC to the same extent as humans, mice possess just two *KIR* genes, and rats only one (Hoelsbrekken et al. 2003) that also have a completely different function.

In humans, the *KIR* are a highly polymorphic gene family, rivalling the polymorphism of the HLA complex which encodes the MHC. The nature of their polymorphism is multi-faceted. Variation in the presence/absence of a subset or the *KIR* genes occurs between individuals. Four of the *KIR*, termed framework genes, are present in all currently genotyped individuals with very few exceptions.

There are two broadly defined *KIR* haplotype groups, A and B, which are based on gene content (figure 1-2). The A haplotype is largely non-variable in terms of gene content and is considered more inhibitory than B as it contains just one activating *KIR*. The B haplotype is much more variable in gene content and can contain genes for up to five activating KIR receptors. Homozygosity for the group A haplotypes in pregnant women carrying a foetus expressing HLA-C2 are at a high risk for pre-eclampsia. This is due to inhibitory signals generated by 2DL1 after recognition of HLA-C2 expressed by trophoblasts (Hiby et al. 2004). The risk is reduced if a group B haplotype is present when carrying a foetus expressing HLA-C2. The reduction in risk correlates with the number of activating *KIR* genes. Homozygosity for the group A haplotypes is beneficial in the context of viral infection however. Group A homozygosity provides resistance to acute Ebola virus infection and is associated with a favourable outcome when treated for chronic HCV infection when paired with HLA-C1 homozygosity (Wauquier et al. 2010; De Re et al. 2015). A model for maintenance of the two haplotype groups in human populations has been proposed. An epidemic infection will select for KIR A haplotypes that are more protective and provide a higher chance of survival. By the end of the epidemic, the percentage of the population possessing the A haplotype will be significantly increased. Once the epidemic has ended, survival of the population becomes

dependent on reproduction, resulting in selection for the B KIR haplotype (Parham and Moffett 2013).

Variation at the allelic level is a major contributor to the polymorphism of the *KIR*. Alleles of *KIR* can vary by orders of magnitude in the strength of their response and amount of expression on the cell surface. Alleles of *3DL1* vary in their NK cell expression pattern and inhibitory capacity, significantly influencing AIDS progression (Martin et al. 2007). A common allele of *3DL1*(*3DL1*004*) is poorly expressed at the cell surface due to substitutions in the D0 and D1 domains (Pando et al. 2003). In the latest release of the IPD-KIR database (July, 2017) there are 907 listed *KIR* alleles (Robinson et al. 2010).

1.3.2 The killer cell lectin-like receptors

The killer cell lectin-like receptors (KLR) are type II TM C-type lectin proteins, encoded by genes within NKC. Like the KIR, they are able to recognise MHC class I as well as MHC class I-like molecules. The NKC genes have two commonly assigned names based on different nomenclature systems. As both are often used, the alternate name is also listed here in brackets. The human NKC contains a single non-functional *KLRA* (*Ly49*) gene, four *KLRC* (*NKG2A*, *C*, *E* and *F*) genes, a *KLRK* (*NKG2D*) gene and a *KLRD* (*CD94*) gene. In contrast, mice and rats have undergone a much more significant expansion (figure 1-3). Mice and rats possess 23 and 34 *KLRA* genes respectively and a unique *KLRH* gene that has been lost in primates (Higuchi et al. 2010; Naper et al. 2002).

As with the KIR, KLRA receptors can either be inhibitory through ITIM, or activating by association with adapter molecules to a charged transmembrane domain containing ITAM. They exists as homodimers on the cell surface and interact with ligands via a natural killer receptor domain (NKD). Mice of different strains have varying numbers of *KLRA* and there are strain specific alleles. Variation in the presence/absence of *KLRA* has been implicated in resistance to *murine cytomegalovirus* (MCMV). The MCMV-encoded protein m157 has a strong interaction with an inhibitory KLRA in 129/J mice. C57BL/6 mice encode an activating KLRA, not present in 129/J mice. This activating KLRA also has a strong affinity to m157 and is thought to have arisen as a direct response to the immune pressure generated by the production of MHC-I decoys produced by MCMV (Smith et al. 2002; Sun and Lanier 2009).

The KLRC form disulphide bonded heterodimers with KLRD and KLRK forms homodimers. KLRC/KLRD heterodimers and KRLK homodimers are expressed on the cell surface. They are largely monomorphic and recognise non-classical MHC class I receptors HLA-E and Qa1 in humans and mice respectively. In the case of the KLRC/KLRD heterodimers, signalling is mediated by KLRC. KLRC cannot be expressed on the cell surface without KLRD acting as a chaperone (Carretero et al. 1997; O'Callaghan 2000; Phillips et al. 1996). Infection by *human cytomegalovirus* (HCMV) of fibroblasts results in the expression of the ligands of KLRK. HCMV is able to evade activation of KLRK by producing the glycoprotein UL16. UL16 captures a number of the expressed ligands, retaining them in the endoplasmic reticulum, resulting in diminished NK cytotoxicity (Welte et al. 2003).

1.4 Cattle NK cells

As in humans, cattle NK cells are defined by presence of NCR1 on their surface (Storset et al. 2004). Cattle NK cells can be differentiated into phenotypic subsets based on their level of CD2 expression. Approximately 80% of NK cells in peripheral blood are $CD2^{high}$. In comparison to $CD2^{low}$ NK cells, they are poorer producers of IFN-γ and do not respond as strongly to IL-2 stimulation (Boysen et al. 2006). $CD2^{-}/CD2^{low}$ NK cells are the main subset present within the skin-draining afferent lymphatic vessels and lymph nodes. The small number of NK cells found in the efferent lymph are also predominantly $CD2^{-}/CD2^{low}$. This suggests that NK cells can move from lymph nodes back into circulation and may be important during the immune response to infection (Hamilton et al. 2017).

1.4.1 NK cell receptors in cattle

Recent advancements in sequencing technologies, coupled with the falling cost of whole genome sequencing, have precipitated the sequencing and assembly of the genomes of multiple non-model species. The sequencing of livestock species is a vital component of genome-enabled breeding, which has been a particular priority due to concerns over food security. This has facilitated comparisons of the LRC and NKC across species. Expansion of the LRC appeared to be limited to the simian primates, with other species possessing either no or one *KIR* gene (Hammond et al. 2009; Futas and Horin 2013; Sambrook et al. 2006). Characterisation of cattle revealed an expansion of the LRC to an extent even

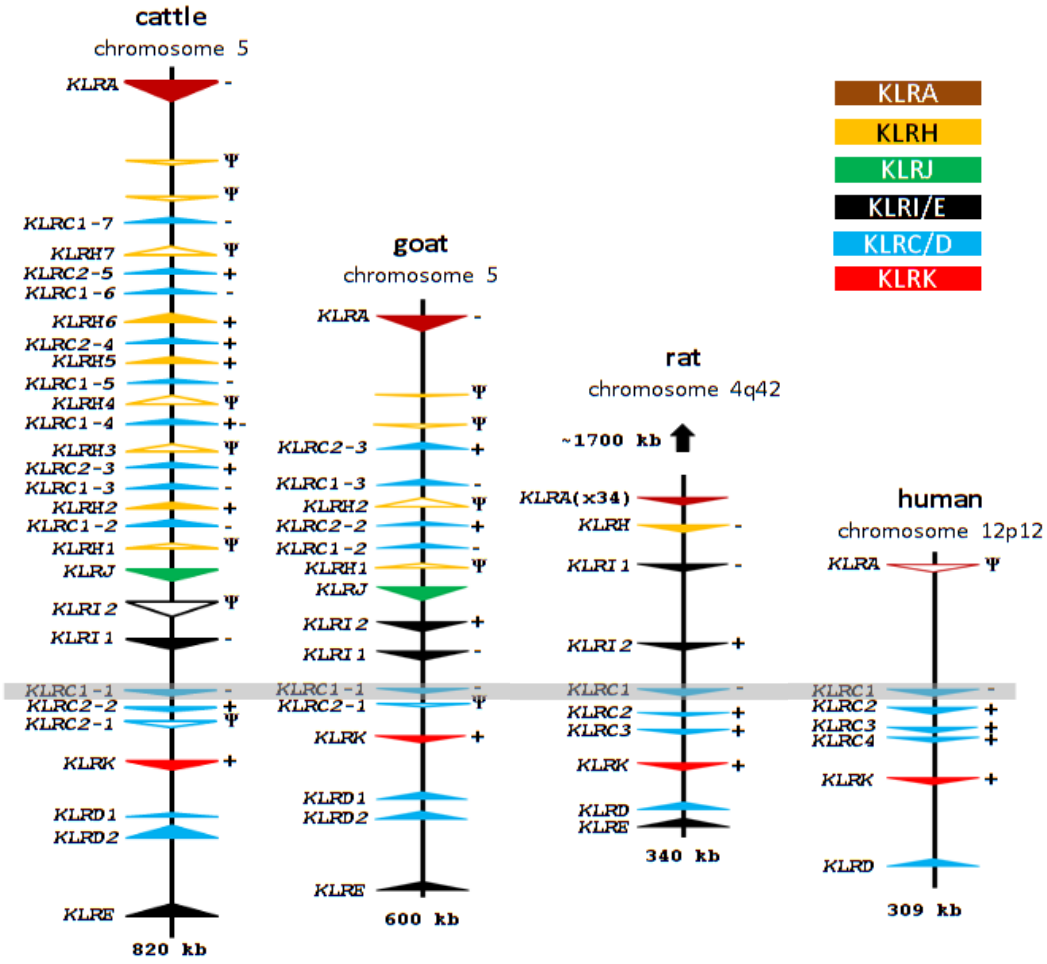


Figure 1-3. The structure of the NKC in the genomes of various species. Arrows indicate individual genes and arrow colour indicates *KLR* group according to the figure key. Filled arrows are putatively functional and empty are putatively non-functional (also indicated by a psi symbol to the right of the gene). Direction of transcription is shown by the orientation of the arrow. Taken and modified from (Schwartz et al. 2017).

greater than humans (McQueen et al. 2002; Storset et al. 2003; Sanderson et al. 2014). Expansion of the NKC has also occurred outside of the model organisms (figure 1-3) (Schwartz et al. 2017). Cattle are the only studied species to have significantly expanded genes within both the LRC and the NKC.

1.4.2 KIR in cattle and goats

The presence of *KIR* in cattle was first described after the detection of the cDNA of multiple KIR-like genes (McQueen et al. 2002). Since then, two cattle LRC haplotypes have been characterised (Sanderson et al. 2014). Cattle have expanded two lineages of KIR, *KIR3DL* and *KIR3DX*. In comparison humans have only expanded the *3DL* lineage. The divergence of the two lineages is estimated to have occurred ~135 million years ago (mya), occurring before the radiation of placental mammals. The nomenclature of the cattle *KIR* follows the same rules as the human *KIR*, with the exception of the addition of an X to denote if the gene is of the *3DX* lineage (e.g. *3DXL1*).

Due to the repetitive nature of the LRC, the quality of its assembly in the reference genome UMD3.1 is poor, lacking the majority of the known *KIR*.

Sequencing of BAC clones gave rise to the first two characterised haplotypes. The cattle LRC has arisen through a series of block duplications, each block contains *3DX* and *3DL* genes. The genes within these blocks have been assigned groups based on their phylogeny (figure 1-4). The most recent duplication appears to have given rise to blocks A and B, which are ~66 kb length. These blocks have high sequence identity, 191 nucleotide substitutions occur between them. Duplication of the *3DX* lineage gene has generated diversity of KIR in cattle, the *3DL* lineage has given rise to just one functional KIR (Guethlein et al. 2007). There are twelve *3DX* lineage genes, of which seven are predicted to produce a functional protein. Six of these encode for three Ig domains and an inhibitory long cytoplasmic tail. The exception to this is *3DXS1*, which encodes for three Ig domains and an activating short cytoplasmic tail. The five remaining *3DX* lineage genes are all predicted to be activating but non-functional. They have been inactivated in a consistent manner, deleterious mutations within the extracellular domains are identical between genes. This suggests that a single activating gene acquired these mutations before their duplication, but after the duplication event that gave rise to *3DXS1*.

Haplotype 1 (*H1*) represents a complete *KIR* haplotype of length 263 kb. Haplotype 2 (*H2*) from the same individual as *H1* is incomplete although contiguous, and is subsequently shorter at 203 kb. Within *H1* are 18 *KIR* genes and fragments and within *H2* are 14 *KIR* genes and fragments (figure 1-5). Within the overlapping region between the two haplotypes, there are 1008 SNPs. Despite the number of SNPs, the two haplotypes are identical in gene content in the overlapping region. However, the genes present on both *H1* and *H2* are allelically different on each haplotype. The nucleotide sequence of the alleles is highly similar. There are just 13 and 12 SNPs in the exons between haplotypes in block A and B respectively. Comparing block C alleles between the two haplotypes reveals 537 nucleotide substitutions, suggesting block C is significantly older than A and B.

There appears to be significant diversity in the cattle *KIR* system, as demonstrated by the allelic differences between the two known haplotypes. The nucleotide polymorphisms of the *KIR* are primarily nonsynonymous substitutions concentrated in the Ig-like domains, which in humans interact with MHC class I.

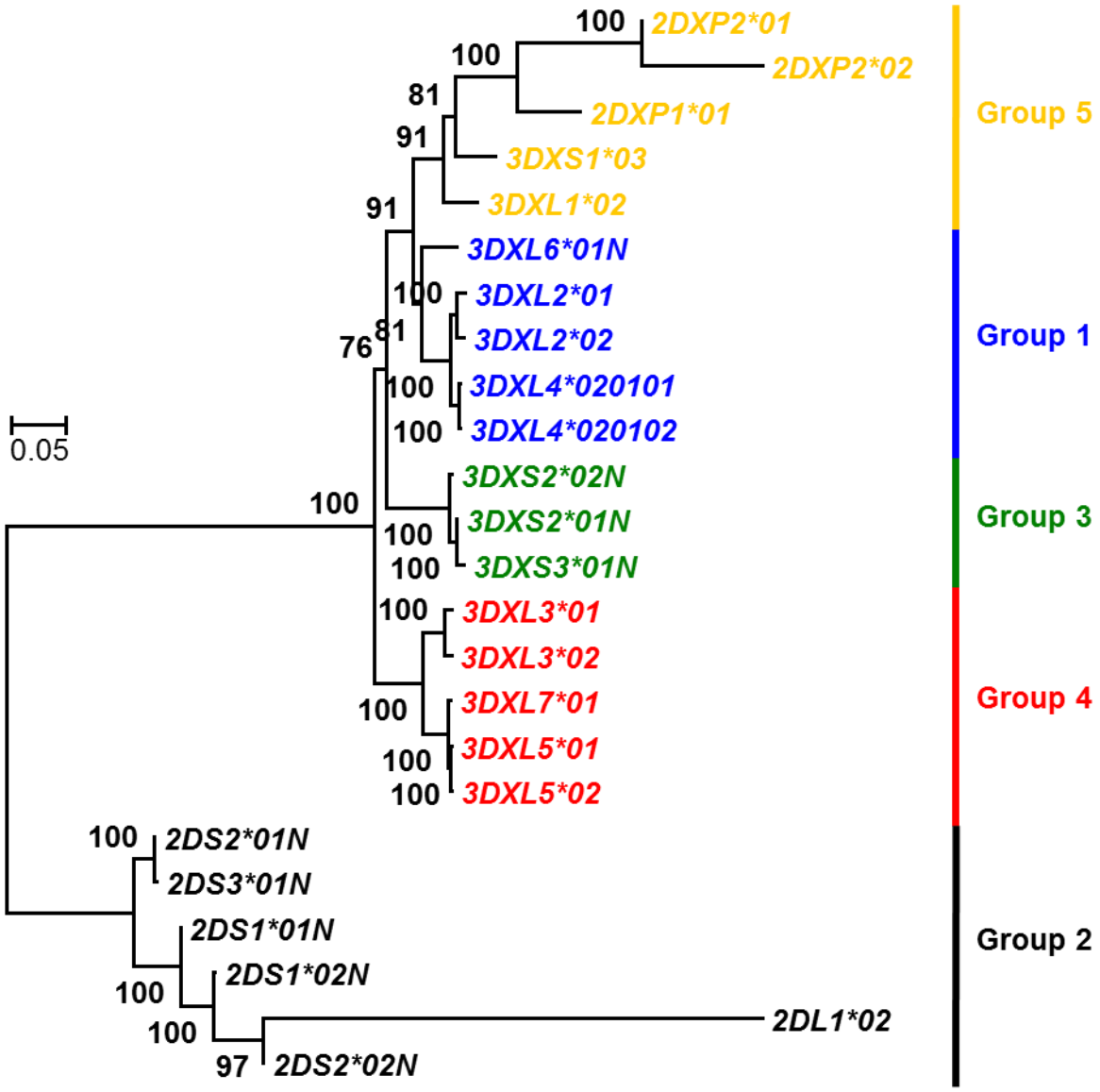


Figure 1-4. Phylogenetic analysis and subsequent grouping of the alleles present in *KIR* haplotype 1 and 2. Taken from (Sanderson et al. 2014).

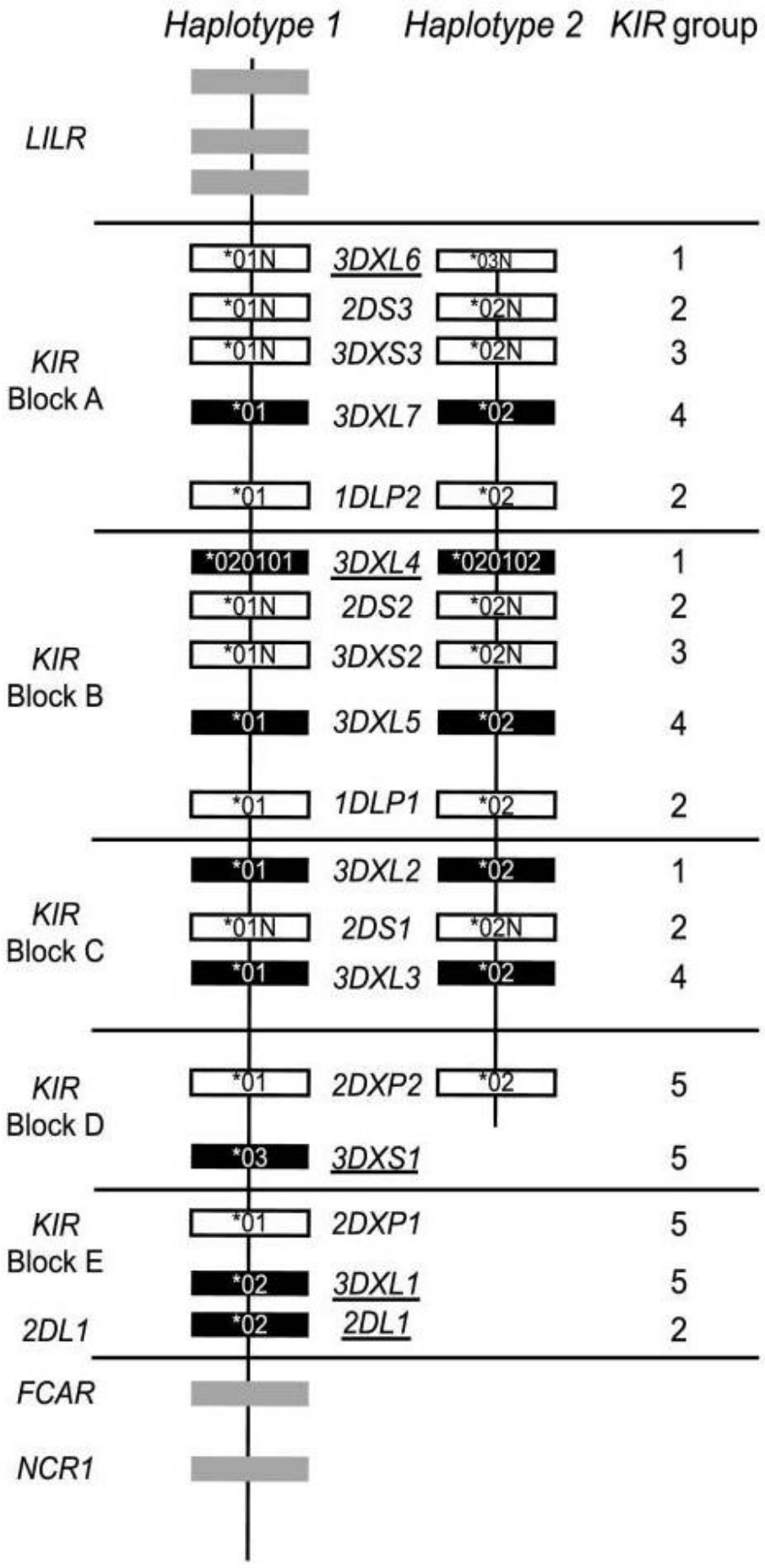


Figure 1-5. Comparison of the gene and allelic content of *KIR* haplotype 1 and 2. Black boxes indicate predicted functional *KIR* and white boxes predicted non-function *KIR.* Flanking genes are indicated by grey boxes. Taken from (Sanderson et al. 2014).

The most divergent of the cattle *KIR* is *3DXL6*, which is non-functional in *H1* and *H2*. Based on cDNA evidence, *3DXL6* appears to have several functional and non-functional alleles. Further cDNA evidence supports the existence of *2DXS1*, not present on either haplotype (Sanderson et al. 2014). The Ig domains of *2DXS1* appear to be allelic to *3DXL6*, whereas the *2DXS1* transmembrane domain and cytoplasmic domain share more similarity with the other activating *KIR*. This suggests that *2DXS1* is an allele of *3DXL6*, albeit with an activating tail rather than inhibitory (Sanderson et al. 2014). A similar situation is observed in humans, where *3DL1* and *3DS1* segregate as alleles of the same *3DL1/S1* locus (Kelley, Walter, and Trowsdale 2005; Parham 2005).

Expansion of the *KIR* has also occurred in goats, independently from other species. Located within the goat LRC are 15 *KIR* genes, seven of which are putatively functional. The functional *KIR* either clade with the group 4 cattle *KIR*, or clade into two novel groups defined as 6 and 7. Unlike cattle, goats do not possess a functional *3DL* lineage *KIR* gene. They do however encode for a novel functional four-domain inhibitory *KIR*, possessing two D1 domains (Schwartz, Sanderson, and Hammond, unpublished data).

1.4.3 KLR in cattle and goats

Unlike humans, cattle have also undergone a significant expansion of the NKC. Goats, another ruminant species, have undergone a similar expansion to cattle, although to a more limited extent (figure 1-3). The organisation of the NKC is largely conserved between the ruminants, rats and humans. The ruminants possess *KLRJ*, a gene lacking either an activating or inhibitory component, rendering its function unknown. The ruminants have also independently evolved a second *KLRC* locus and possess a novel *KLRH-like* gene with an activating tail. Duplication of this gene has occurred multiple times in cattle. The region located between *KLRA* and *KLRJ* contains the region of expanded *KLRC*-like and *KLRH*-like genes. In cattle, there are 9 *KLRC*-like and 7 *KLRH*-like genes. The *KLRH*-like genes are most closely related to *KLRH1* in rats based on phylogenetic analysis of the extracellular C-type lectin domain. Of the 7, 5 possess activating *KLRC2*-like cytoplasmic and transmembrane domains. This region consists of four *KLRC* genes and two *KLRH*-like genes. One of the *KLRH*-like genes in goats possesses exons 1 and 3 of a *KLRC2*-like cytoplasmic tail (Schwartz et al. 2017). This is evidence that both the *KLRC* expansion as well as *KLRH* recombination

occurred before the divergence of cattle and goats ~30 Mya (Hiendleder et al. 1998). The *KLRC* in this region are split into two groups, *KLRC1* and *KLRC2*. The six *KLRC1* genes are inhibitory and the three *KLRC2* are activating - all are putatively functional.

The second cluster of *KLRC* between *KLRK* and *KLRI* is much more conserved between the ruminants, humans and rats. Located in this region in both cattle and goats is an inhibitory *KLRC1* gene and a *KLRC2* pseudogene. The *KLRC2* pseudogene is rendered non-functional by the same mutations in both cattle and goats, suggesting the loss of function occurred before they two species diverged. Cattle also possess an additional putatively functional *KLRC2*. In humans, *KLRC* and *KLRD* form heterodimers at the cell surface. That cattle and goats have duplicated *KLRD*, suggests that this is also the case in cattle.

Allelic discovery for the *KLRC/H* region is made extremely difficult by the repetitive and highly similar nature of the genes within the region and its effect on the mapping of short reads. Outside of this region, mapping is much easier and subsequently the allelic variation of the rest of the NKC has begun to be elucidated. Sequencing of 20 *Bos taurus* and 3 *Bos indicus* genomic DNA samples revealed substantial allelic variation. The coding regions of *KLRA, KLRJ, KLRI2, KLRI1, KLRK, KLRD1, KLRD2, and KLRE* were shown to contain 77 SNPs, 55 of which resulted in a change at the amino acid level. This variation predominantly occurs within extracellular and ligand binding domains. In contrast *KLRK* was found to be highly conserved – no variation in the nucleotide sequence was observed in the *KLRK* across the sequenced animals. Within *B. taurus*, *KLRD2* and *KLRE* are monomorphic. Variation between the *KLRE* of *B. taurus* and *B. indicus* was observed however (Schwartz et al. 2017).

1.5 Transcription and expression of NK receptors

The entirety of the *KIR* genes present in a human individual are transcribed in their polyclonal NK cell population (Uhrberg et al. 1997). However, the KIR repertoire of an individual NK cell is limited to a subset of the total KIR encoded for in the individual's genome. The combinations of KIR expressed in an individual cell appear to be random and regulation appears to occur mainly at the transcriptional level. Examination of 100 NK clones from two donors, revealed 33 distinct receptor (KIR and KLRC/KLRD) phenotypes in one donor and

64 in the other. Every NK clone possessed either an inhibitory *KLR* or *KIR* (Valiante et al. 1997). The *KIR* and *KLR* gene families both have a pattern of variegated expression, 80% of NK cells express between one and three receptors out of the total *KIR* gene repertoire present in the genome (Valiante et al. 1997; Held and Kunz 1998; Kubota et al. 1999). The product of the frequencies of NK cells expressing each receptor can be used to estimate the frequency of NK cells co-expressing a given receptor combination. This is known as the "product rule" and applies to both *KIR* and *KLRA* (*Ly49*) (Raulet et al. 1997; Valiante et al. 1997; Kubota et al. 1999). Deviation from this occurs with NK cells co-expressing KLRC/KLRD and inhibitory KIR, the frequencies of cells expressing both is lower than expected (Valiante et al. 1997). The frequency of cells expressing a given *KIR* is highly dependent on the gene copy number of that *KIR*. The frequency of cells expressing 3DL1 is higher in individuals with two copies compared to those with one copy (Li et al. 2008; Beziat et al. 2013).

1.5.1 Transcriptional regulation of *KIR* and *KLR*

Transcriptional regulation of the *KIR* occurs via promoters, the *KIR* distal promoter and the *KIR* proximal promoter. The distal promoter has been identified upstream of all human *KIR*. It is not tissue specific, the distal promoter is active in NK cell and T cell lines (Stulberg et al. 2007). Stimulation with the cytokine IL-15 induces c-Myc binding to the distal promoter, driving distal transcription (Cichocki et al. 2009). This is thought to be a key stage of NK development as the IL-15 receptor is present only on $CD56^{bright}$ and mature $CD56^{dim}$ NK cells (Freud et al. 2006). Unlike the distal promoter, the proximal promoter is only active in NK cells. A strong correlation between stable transcriptional silencing of individual genes and DNA methylation within the genes proximal promoter has been demonstrated (Santourlidis et al. 2002; Chan et al. 2003). Both the distal and proximal promoter are located upstream of each gene and are bi-directional. The relative affinity of transcription factors required for sense antisense promoter activity determines the probability of producing the sense transcript required for gene activation.

More recently, transcription from an additional intermediate promoter (Pro-I) has been shown to be required for expression of *KIR2DL1* (Wright et al. 2014). A correlation between Pro-I transcripts and protein expression was observed. In contrast, no correlation was observed between proximal transcripts and protein

expression. Decreased Pro-I activity was associated with the presence of non-translatable splice variants of the proximal transcript. Four distinct classes of the Pro-I promoter have been identified. These classes correspond with four expression phenotypes of KIR sub-groups: 2DL4 which is expressed by $CD56^{bright}$ cells in a non-variegated manner, 2DL3 which is not expressed on circulating NK cells, 2DL2/S2/L3 of which expression is observed early in reconstituting NK cells after transplant, as well as the remaining KIR that are expressed by $CD56^{dim}$ subsets (Wright et al. 2014).

Methylation of the *KIR* promoters suppresses expression of the *KIR* and is responsible for the maintenance of allele-specific expression. The NK cell line NK-92 only expresses 2DL4, treatment with the demethylating agent 5-aza-2'-deoxycytidine results in expression of multiple KIR (Chan et al. 2003).

Bi-directional promoters (Pro1) are also utilised by the *KLRA* gene family. Pro1 is located upstream of the core Pro2 and Pro3 promoters responsible for production of *KLRA* transcripts. Forward transcription of Pro1 is thought to a produce a splice sense transcript that plays a role in opening the chromatin of the Pro2 and Pro3 promoters, enabling gene transcription. Regulation of *KLRA* by DNA methylation is thought to be unlikely, regulation is thought to occur at the histone level (Wilhelm et al. 2001; Gosselin et al. 2000). Pro1 has been shown to play a role in the gene activation of developing NK cells but a role in mature NK cells is less likely (McCullen et al. 2016).

1.5.2 Stability of the NK receptor repertoire

The life-span of an NK cell is relatively short, they have a turnover time of approximately two weeks (Jamieson et al. 2004). Despite this, the NK receptor repertoire remains remarkably stable over time – very low variability was observed over a period of 6 months. It can however undergo rapid modification in response to stimulation. Stimulation with either IL-2 or IL-15 results in a change of expression up to eightfold greater than observed in the same serial sampling of untreated NK cells. This stability is also observed at the individual NK level, the different combinatorial receptor phenotypes show low variability over time. Over the lifetime of an individual the diversity increases. This increase in diversity correlates with an increase in the NK terminal differentiation marker CD57 (Strauss-Albee et al. 2015). Expression of CD57 is

induced by cell or cytokine stimulation, resulting in an increase with age (Lopez-Verges et al. 2010; Le Garff-Tavernier et al. 2010). The increase in NK diversity is not simply a direct result of aging, but rather the NK response history of an individual. Higher NK diversity is associated with an increase likelihood of HIV-1 infection. $CD57^+$ terminally differentiated NK cells produce significantly more IFN-γ than $CD57^-$ NK cells. They also degranulate and divide less than $CD57^-$ NK cells. Short-term exposure of NK cells to cells infected with *West Nile virus* (WNV) and HIV-1 has been shown to result NK cell diversification. This results in a repertoire that becomes more individualised with each NK response (Strauss-Albee et al. 2015).

1.5.3 Expression of NK receptors on other immune cells

Expression of KIR has also been observed on $CD8^+$ T cells (Mingari et al. 1996), gamma delta (γδ) T cells (Battistini et al. 1997), and to a lesser extent $CD4^+$ T cells (Namekawa et al. 1998; van Bergen et al. 2004). T cells express the entire repertoire of *KIR* in an individual, expression of none of the KIR is limited to NK cells. A study comparing two donors of NK cells and NK receptor positive (NKR^+) T cells revealed that expression of KIR2DL1 is much higher in NK cells. Contrastingly, expression of KIR2DL2 and KIR3DL1 was more frequent in NKR^+ T cells. Low frequency expression of KIR2DL1 separates both γδ T cells and αß T cells from NK cells (Uhrberg et al. 2001). In healthy adult donors an average of 0.2% of $CD4^+$ T cells express KIR, increasing with age to 1.0% in the elderly. The majority of $CD4^+KIR^+$ cells are HLA class II-restricted effector memory Th1 cells and do not always possess an inhibitory KIR (van Bergen et al. 2004).

Expression of KLRC/KLRD has also been observed in γδ T cells and at a lower level in αß T cells. Although the KIR repertoire of NK cells and NKR^+ T cells is mostly similar, KLRC/KLRD expression is less frequent in NKR^+ T cells than in NK cells. KLRC1 is much more frequently expressed on γδ T cell clones than αß T cell clones. The expression of KLRC/KLRD is similar between NK cells and γδ T cells (Uhrberg et al. 2001). The NKR expressed by T cells have been shown to be functional. Inhibitory signals generated by inhibitory KIR or KLR can override the stimulatory signal of the T cell receptor (TCR), inhibiting the T cells cytolytic and cytokine release response (Nakajima, Tomiyama, and Takiguchi 1995; D'Andrea et al. 1996).

1.5.4 Tissue specific NK cells

A number of tissue and organ specific NK phenotypes and functions have been identified in humans and mice. These phenotypes are generated by cellular interactions and the local microenvironment unique to a specific tissue/organ.

In the liver, NK cells make up 30-50% of the lymphocyte population. Hepatic NK cells are hyporesponsive compared to NK cells in other tissues such as the spleen. They have a dampened response to both IFN-γ and IL-12/IL-18 stimulation (Lassen et al. 2010). This hyporesponsiveness is due to high expression of the inhibitory receptor KLRC1 (NKG2A) and decreased expression of the activating receptor KLRA. This alteration of expression is thought to promote liver tolerance but may also contribute to persistent infection within the liver. Tolerance is required as the exposure to gut-derived foreign antigens that occurs within the liver would otherwise lead to inflammation (Crispe 2003).

NK cells are also found in the lung, approximately 10% of the lymphocyte population in a healthy mouse lung is made up of NK cells (Grégoire et al. 2007). Bronchial epithelial cells are thought to contribute to the survival of NK cells in the lung through production of IL-15 (Ge et al. 2004). Large numbers of NK cells are recruited to the lung within days of infection and begin secreting cytokines (Stein-Streilein et al. 1983). As with liver NK cells, lung specific NK cells are less functionally active. Lung NK cells expresses higher levels of inhibitory receptors and lower levels of activating receptors. This implies the environment of the lung tightly regulates NK function, limiting functionality to periods of respiratory infection (Wang et al. 2012).

NK cells have been implicated in playing a role in a number of skin diseases including psoriasis. NK cells present in psoriasis lesions had reduced expression of CD57 and KLRC1 compared to controls (Batista et al. 2013). They also exhibit reduced degranulation and production of pro-inflammatory cytokines such as IFN-γ (Dunphy et al. 2017).

Gut NK cells are predominately of the $CD56^{bright}CD16^{-}$ phenotype (León et al. 2003). NK cells present in the gut-associated lymphoid tissues are a unique subtype, specialising in production of IL-22 (Colonna 2009). IL-22 has multiple

roles including cell proliferation and tissue regeneration, pathogen defence and protection of the intestinal barrier (Parks et al. 2015).

1.6 RNA sequencing

The ability to understand the identity of each RNA molecule in a cell along with the abundance of each RNA is widely considered the ultimate goal of RNA research. Research into this has been ongoing since the discovery of the role of RNA in gene expression. The first relatively high throughput method was the expressed sequence tag (EST). ESTs are short (200-800 nucleotide bases), randomly selected single-pass sequence reads from cDNA libraries. ESTs have been used extensively and have been used to identify a number of new genes in multiple species (Putney, Herlihy, and Schimmel 1983; Adams et al. 1991). However, ESTs provide limited information regarding expression levels, due to the high sequencing cost and the semi-quantitative nature of the data. Serial Analysis of Gene Expression (SAGE) was developed which reduced sequencing cost by only sequencing up to 21bp per cDNA (Velculescu et al. 1995). SAGE was superseded by microarray technology as it was much more affordable for large scale analysis. DNA microarrays utilises probes attached to a solid surface which hybridise with fluorescently labelled targets derived from transcripts. Whilst this method is large scale, enabling genome-wide analysis, it requires prior knowledge of the target transcripts to design the probes. This makes it unsuitable for discovery of novel genes/transcripts as well as for use with uncharacterised species (Schena et al. 1995; Lockhart et al. 1996). The invention of massive parallel sequencing (often referred to as Next Generation Sequencing (NGS)), has provided the next major advancement in the field of RNA research.

1.6.1 Illumina RNA sequencing

Illumina RNA sequencing (RNA-Seq) has quickly become a highly popular method to interrogate multiple aspects of RNA. RNA-Seq generates huge volumes of data in a short period of time. Its major advantage over microarray analysis is that it requires no prior knowledge of the underlying genome or transcriptome. However many applications of RNA-Seq are aided by the presence of a high quality reference genome.

The general principal of preparing RNA for RNA-Seq involves generating a cDNA library from RNA, fragmenting the cDNA to a certain size, and attaching

sequencing adapters to either side of the fragment which are required for subsequent amplification and sequencing.

Total RNA from a cell is dominated by ribosomal RNAs (rRNAs) which are typically of little interest (Rosenow et al. 2001). Two main approaches have been developed to attempt to eliminate rRNA from total RNA before sequencing. Poly(A) selection uses beads coated with oligo-dT molecules to select for the poly(A) tail present on most messenger RNA (mRNA) but not on rRNA. An alternative method is rRNA depletion. Depletion of rRNA is done by hybridizing sequence-specific probes to rRNA, after which depletion occurs using streptavidin beads (Cui et al. 2010). An alternative method of rRNA depletion uses targeted antisense DNA oligonucleotides and subsequent digestion by RNase H (Herbert et al. 2018).

Fragmentation is required due to limitations on read length of current Illumina sequencers. Fragmentation is carried out using either physical, enzymatic or chemical approaches and can be done on either RNA or cDNA in RNA-Seq protocols. The choice of fragmentation method introduces various biases into the RNA-Seq data, due to their non-random nature (Nicholson 2014; Poptsova et al. 2014). Dependent on the fragmentation method used, either before or after fragmentation, RNA is converted to cDNA. This conversion is an additional source of bias, as the random hexamers used introduces biases affecting both the nucleotide content of RNA sequencing reads as well as the uniformity of the locations of the reads along the transcript (Hansen, Brenner, and Dudoit 2010).

Adapters are then ligated to the cDNA, this process has the unfortunate side effect of losing information about which strand the RNA originated from. A number of strand-specific protocols have been developed to address this. Typically, cDNA synthesis is separated into two stages, first and second strand synthesis. First strand synthesis occurs in the normal manner using reverse transcriptase and random primers. After first strand synthesis, dTTP (thymidine) is swapped for dUTP (uracil) for second strand synthesis. The second strand containing uracils is then degraded using uracil-n-glycosylase. In the subsequent PCR to generate enough material for sequencing, only the first strand is then amplified, retaining the strand information of the libraries (Parkhomchuk et al. 2009).

1.6.2 Long read RNA-Seq

Further advancements in the field of DNA/RNA sequencing have primarily sought to improve on read length. Longer reads are beneficial to a variety of applications including genome assembly, mapping of reads to complex or repetitive regions, and transcript isoform discovery. The two major long read technologies for RNA-Seq are Pacific Biosciences (PacBio) Iso-Seq and Oxford Nanopore (Weirather et al. 2017). Both of these technologies offer the ability to sequence full length transcripts. This reduces or eliminates a number of technical challenges associated with RNA-Seq data analysis such as transcript assembly and mapping of short reads. Both PacBio and Nanopore sequencing are less mature platforms than Illumina RNA-Seq. The major problems associated with both platforms are a much higher error rate and a smaller total base output. The error rate of raw PacBio data is between 13-15% (Rhoads and Au 2015) and between 5-15% for Nanopore data (Jain et al. 2017; Jain et al. 2018). Error correction is available via bioinformatics methods for both platforms, resulting in accuracy of >99.9% and 95% for PacBio and Nanopore respectively. This increase in accuracy also results in a substantial decrease in sequence length (Mahmoud et al. 2017). The error rate of Illumina RNA-Seq is dependent on a number of factors such as sequencing platform, chemistry and read length, but is typically between 0.1-1% (Glenn 2011). Illumina platforms also provide much more data at a lower cost than PacBio or Nanopore. An Illumina HiSeq NovaSeq 6000 can output 2.4 Tb to 3 Tb per run at a cost of 10.26 USD per Gb(Daniel Johnson, Illumina Cambridge, personal communication). In comparison a PacBio Sequel outputs 5-10 Gb bases (85 USD per Gb) and a Nanopore MinION can output up to 20 Gb (24 USD per Gb)(van Dijk et al. 2018). This makes these platforms less useful and more expensive for quantification of transcripts that are not highly expressed within the global transcriptome.

1.6.3 RNA-Seq read mapping

An important step for many RNA-Seq analyses is the mapping/alignment of reads to a reference genome or transcriptome. A number of algorithms have been developed for the mapping of short sequencing reads. Mapping algorithms must

be ‘splice-aware’ and capable of handling strand-specific data in order to correctly map RNA-Seq data, consequently a good DNA-Seq aligner is not always suitable for RNA-Seq analysis. The accuracy at the mapping stage has a significant impact on the accuracy of downstream analyses. Accuracy is impacted by a large array of factors such as polymorphisms, intron-sized gaps, sequencing error, low-complexity sequences and alternative splicing. Poor quality annotations also negatively impact mapping accuracy, requiring aligners be able to align reads across unannotated splice junctions (Baruzzo et al. 2017).

Analysis of simulated data is commonly used as a benchmarking tool for RNA-Seq data as the precise nature of the data is known. The large number of RNA-Seq aligners available and the number of impacting factors that can be present in real RNA-Seq data makes comprehensive and accurate comparisons difficult however. Comparisons of RNA-Seq aligners often place particular emphasis on run time and resource usage. Depending on the computing resources available to the end-user and nature of analyses being carried out, small increases in accuracy may be worth the trade-off of longer run time. A benchmark of 14 mapping algorithms reveals that performance of the tools varies based on genome complexity (Baruzzo et al. 2017). They also found a poor correlation between accuracy and popularity of the tools, with the most widely cited tool (TopHat2) underperforming for most metrics using default parameters. These alignment algorithms require careful optimisation for the data being analysed. TopHat2 varied from 3% of total bases mapped correctly using default parameters to over 70% after optimisation.

Most RNA-Seq aligners work on a seed-and-extend basis, where a read is aligned by finding a seed (a short token of the sequence that also exists in the reference) and then the alignment is extended to the rest of the sequence. This approach allows few mismatches in the alignment and provides good performance at the cost of not being able to return all existing matches. Alternatively, the Genome Multitool (GEM) mapper uses a filtration-based approach to approximate string matching. This means that GEM always carries out an exhaustive search, finding all matches. This also occurs at a speed comparable to or faster than other aligners, whilst also being more accurate than many other mappers (Marco-Sola et al. 2012).

1.6.4 Multi-mapping reads

Each RNA-Seq aligner will typically produce different results from the same dataset, due to the variety of algorithms and scoring schemes that they utilise. One of the major problems RNA-Seq aligners face is how to account for reads that map multiple times (multi-mapping). This issue is particularly problematic when carrying out analyses into differential gene expression. Multi-mapping reads occur most often when genes have duplicated, resulting in multiple locations within the genome sharing high sequence similarity. Multiple methods for accounting with multi-mapping reads have been developed. Bowtie (Langmead et al. 2009) and Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009) both report a random match in their unique alignment mode if a read maps multiple times. Other tools uniformly distribute across all matches, sometimes weighting them based on the number of matches. All of these approaches result in significant bias in the results however (Zytnicki 2017).

1.6.5 RNA-Seq of NK cell receptor genes

RNA-Seq analysis has the potential to provide enormous insight into NK cell receptors of the lesser studied species such as cattle and goats. The extent and location of their transcription is currently unknown. Non-functional/functional status of genes is assigned often based on the gene structure in a single animal. Analysis of the regions of the genome encoding these receptors is complicated by gene duplication giving rise to genes with highly similar sequences (Sanderson et al. 2014). This makes any analysis utilising Illumina short reads extremely problematic. RNA-Seq in particular is made difficult due to multi-mapping, which can result in non-transcribed genes reported as transcribed (Mortazavi et al. 2008). Very little is known about the extent of their polymorphism or haplotypic variation in non-model species. This is because the majority of the current information originates from a very small number of animals. This further complicates RNA-Seq analysis as read mapping requires a suitable reference sequence. Reference free (*de novo*) RNA-Seq analysis typically involves assembly of transcript sequences, which is impractical due to the sequence similarity of the genes (Migalska et al. 2017).

Antibody-mediated analysis of cell surface expression of NK receptors is the commonly used tool in the human and mouse fields. NK receptors in these

species have been a subject of interest for decades. Subsequently a large panel of antibodies have been created over this time. These antibodies are not perfect, they are often cross reactive between individual receptors and some do not bind to all of a genes alleles (Huhn et al. 2018). For the non-model species there are no or very few NK receptor specific antibodies available. Their production is highly complicated by the lack of information on polymorphism and haplotypic variation. Expression analysis data is also fixed once generated. New antibodies must be designed should it be discovered that the current selection are unable to bind newly discovered genes or alleles. Subsequently expression analysis experiments must be repeated every time new antibodies arise. They also must be generated per species due to the species specific evolution of the NK receptor genes. RNA-Seq of total mRNA does not require a complete characterisation of the genes of interest prior to experimentation as it is not targeted. Rather it captures information about the total transcriptome. This means that datasets generated from RNA-Seq experiments can easily be reanalysed should new genes be discovered. RNA-Seq is also not species-specific, most RNA-Seq protocols do not required adapting to a particular species. Although transcription and expression are not the same, understanding transcription of NK receptors can answer numerous questions whilst also informing future expression experiments.

Development of a pipeline to accurately and confidently analyse transcription of NK receptors from RNA-Seq data would provide insight into their function, relative usage and distribution. This would have the additional advantage of not being specific to a particular species. As long as a reference genome containing a good quality assembly of the NK receptor gene complexes exists, an RNA-Seq pipeline could be utilised on any species of interest.

1.7 Aims of the project

Development of a pipeline to accurately quantify transcription of NK cell receptors in important livestock species from short-read RNA-Seq data is essential to further our understanding. The role of certain *KIR* in combination with *HLA* haplotypes in favourable disease outcomes in humans suggests that the same may be true for other species. Identifying these haplotypes in livestock species is one of the ultimate goals of the field as protective haplotypes could be preferentially selected through breeding programmes. The ability to

accurately investigate the transcription of NK receptors, particularly in response to infection, will provide information towards this ultimate aim. Before transcriptional analysis can be carried out, a method of accurately mapping RNA-Seq data to the NK receptor genes must be created. The current limitations of long read sequencing, coupled with the availability of a large number of short-read RNA-Seq datasets, necessitate the pipeline primarily be able to handle Illumina data. Therefore, the first aim of this project is to develop a pipeline that can accurately map short reads to the LRC and NKC genes, with high confidence. The pipeline will be developed using cattle genes as they are the focus of much of this work, but will be designed in such a way that is can be utilised on any chosen species.

The second aim of the project is to utilise this pipeline to characterise the transcription of NK cell receptors in *ex vivo* NK cells from multiple cattle. This will allow the determination of which genes are transcribed. It will also provide insight into the extent to which they are transcribed in the context of the global transcriptome. Understanding this is vital to informing the depth of coverage required for future RNA-Seq experiments as well as providing an indication of their stability.

The third aim is to characterise the transcription of NK receptors in multiple tissues types as well as immune cell types isolated from cattle. Understanding their distribution on immune cells will allow us to determine which, if any, of these receptors are specific to NK cells. Analysing their transcription in various tissue types can then provide insight into the distribution of NK cells and tissue specific phenotypes.

The fourth aim is demonstrate the use of the developed pipeline in another important livestock species, the goat. Tissue distribution of the NK receptors will be investigated in goats, which will at the same time provide information on which of their genes are transcribed. This will also confirm the ability of the pipeline to be utilised in other species.

The fifth and final aim of the project is to characterise the transcription of NK receptors during a viral infection as part of a vaccination study. This will provide insight into their role during infection, the variability of their transcription levels and the impact of vaccination on their transcription.

2 Chapter 2. Overcoming multi-map problems for RNA-Seq data

2.1 Introduction

The aim of almost all RNA-Seq experiments is to quantify gene transcription through gene counts and compare transcription levels of genes between samples. The most ubiquitous method of obtaining read counts is to align RNA-Seq reads to a reference genome or transcriptome. RNA-Seq reads mapped by alignment software can be placed into two different categories. Reads that map to a single location (uniquely mapping) and those that map to multiple locations (multi-mapping reads). Regions of the genome that contain duplicated genes or genes that contain a number of repeats have a high number of multi-mapping reads as the aligner is unable to determine the origin of the read. Multi-mapping is also directly related to the read length used in sequencing. Longer reads are more likely to contain a unique sequence allowing them to be correctly mapped.

A number of methods have been developed to account for the presence of multi-mapping reads in RNA-Seq experiments. One method is to simply discard multi-mapping reads (Liao, Smyth, and Shi 2014), this removes any ambiguity but can heavily skew downstream applications such as differential expression analysis as read counts to repetitive genes will be drastically reduced. Other aligners weight read counts based on the number of alignments each read generates or randomly pick one alignment (Li and Durbin 2009; Langmead et al. 2009). This approach prevents read counts massively outnumbering the number of reads mapped. It can however result in non-transcribed genes appearing to be transcribed if reads from a closely related gene multi-map to them and is not accurate for gene-complexes.

Mappability of the underlying genome/transcriptome can be used to identify regions that are likely to generate a high proportion of multi-mapping reads (Derrien et al. 2012). The first step of mappability calculations requires splitting the reference sequence into *k*-mers, unique subsequences of the reference of a specified length. Mappability is the inverse of the frequency with which a *k*-mer of read length n, originating at a position x, occurs within the two strands of the reference. If the sequence corresponding to the *k*-mer occurs nowhere else

within the reference, then the mappability at the position x is 1. If the *k*-mer occurs twice within the reference, then the mappability of the position is 0.5. If three times, then the mappability of the position would be 0.333̇ and so on.

The cattle LRC and NKC have both arisen through a series of gene conversion and duplication events (Schwartz et. al, 2017; Sanderson et. al, 2014) which results in a series of highly repetitive and polymorphic genes within each locus (figure 2-1, 2-2). Haplotypic variation of the LRC has also been observed, currently two full length haplotypes have been published that differ in both gene and allele content. Evidence from cDNA and genomic sequencing data suggest additional LRC haplotypes exist containing alleles not observed in either published haplotype (Sanderson et al. 2014). Consequently it is impossible to accurately determine the gene of origin for RNA-Seq reads from this region with current analysis methods. Using mappability calculations of a modified build of the cattle genome assembly (ARSv14) and later a modified cattle transcriptome (UMD 3.1), a custom analysis pipeline was developed. The pipeline uses reads that map to regions that are unique (mappability score of 1) to determine if a gene is transcribed. Reads mapping to non-unique regions are weighted based on their average mappability which is more biologically relevant as well as a more accurate representation of transcription levels compared to weighting on number of times mapped.

We validated our approach with simulations, which consistently show that our analysis pipeline produces better results than a simple mapping-based approach. The results are presented in the next sections.

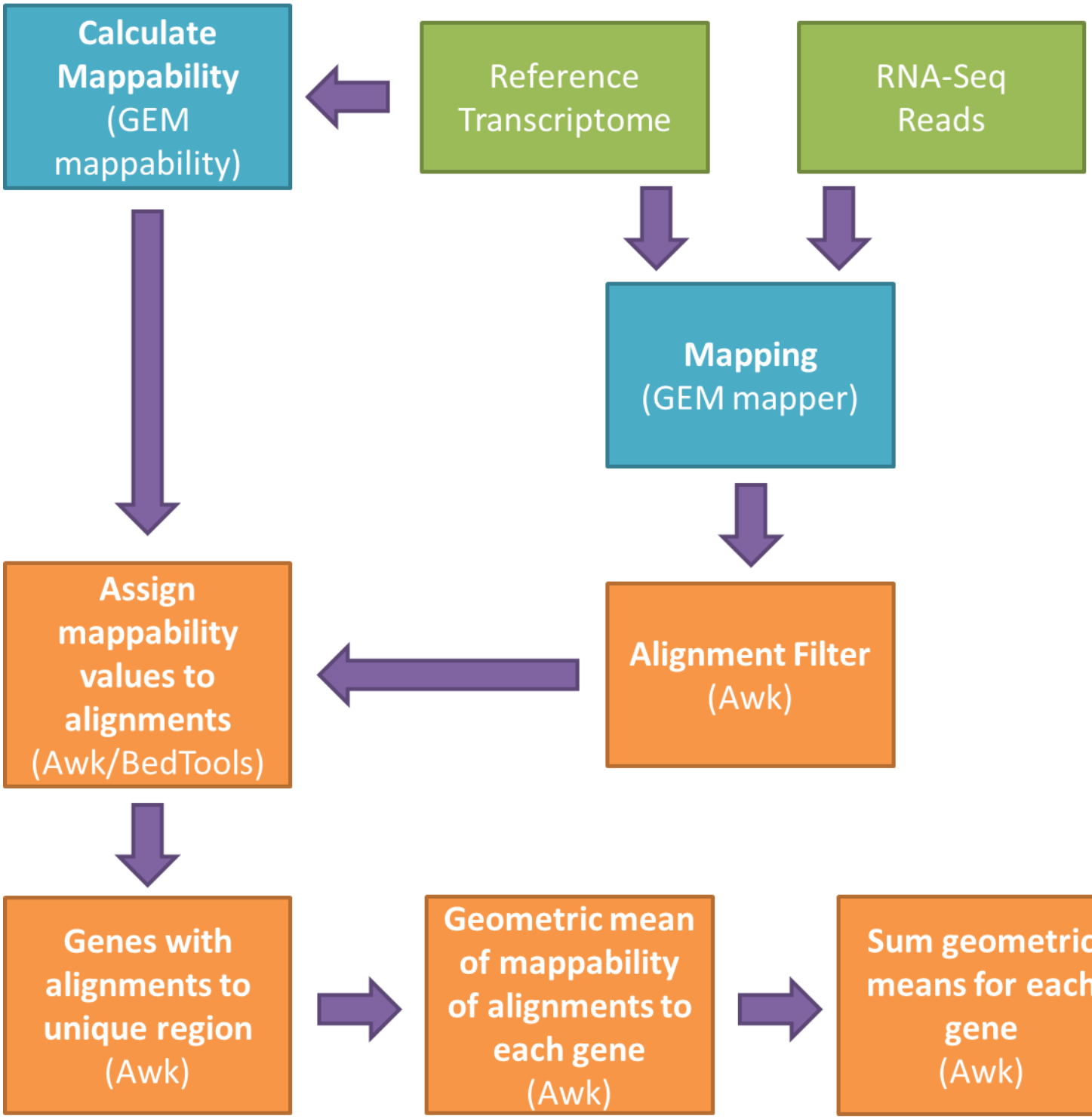


Figure 2-1. Overview of the UniMMap pipeline. Green boxes indicate pipeline input files. Pre-existing analysis tools used as part of the pipeline are indicated by blue boxes. Orange boxes represent novel steps downstream of mapping to generate UniMMap read counts.

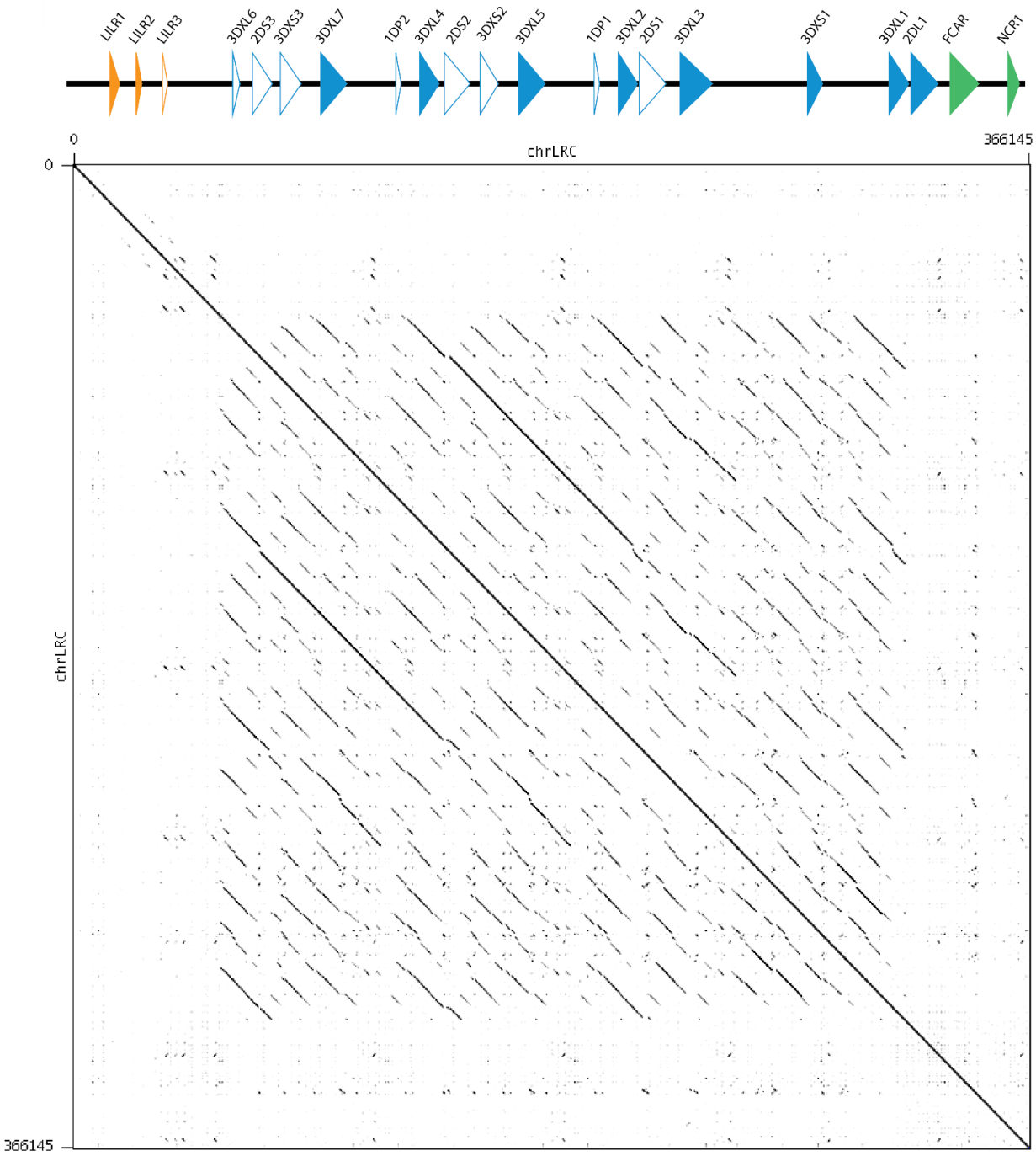


Figure 2-2. Dot plot of the cattle LRC against itself. The dot plot was generated using Gepard v1.40 (Krumsiek, Arnold, and Rattei 2007) with a word length of 10. The LRC sequence was generated from a BAC clone by (Sanderson et al. 2014). Orange arrows indicate *LILR* genes, blue arrows indicate *KIR* genes and green arrows indicate other genes. Open arrows are genes predicted to be non-functional and filled arrows are predicted to be functional.

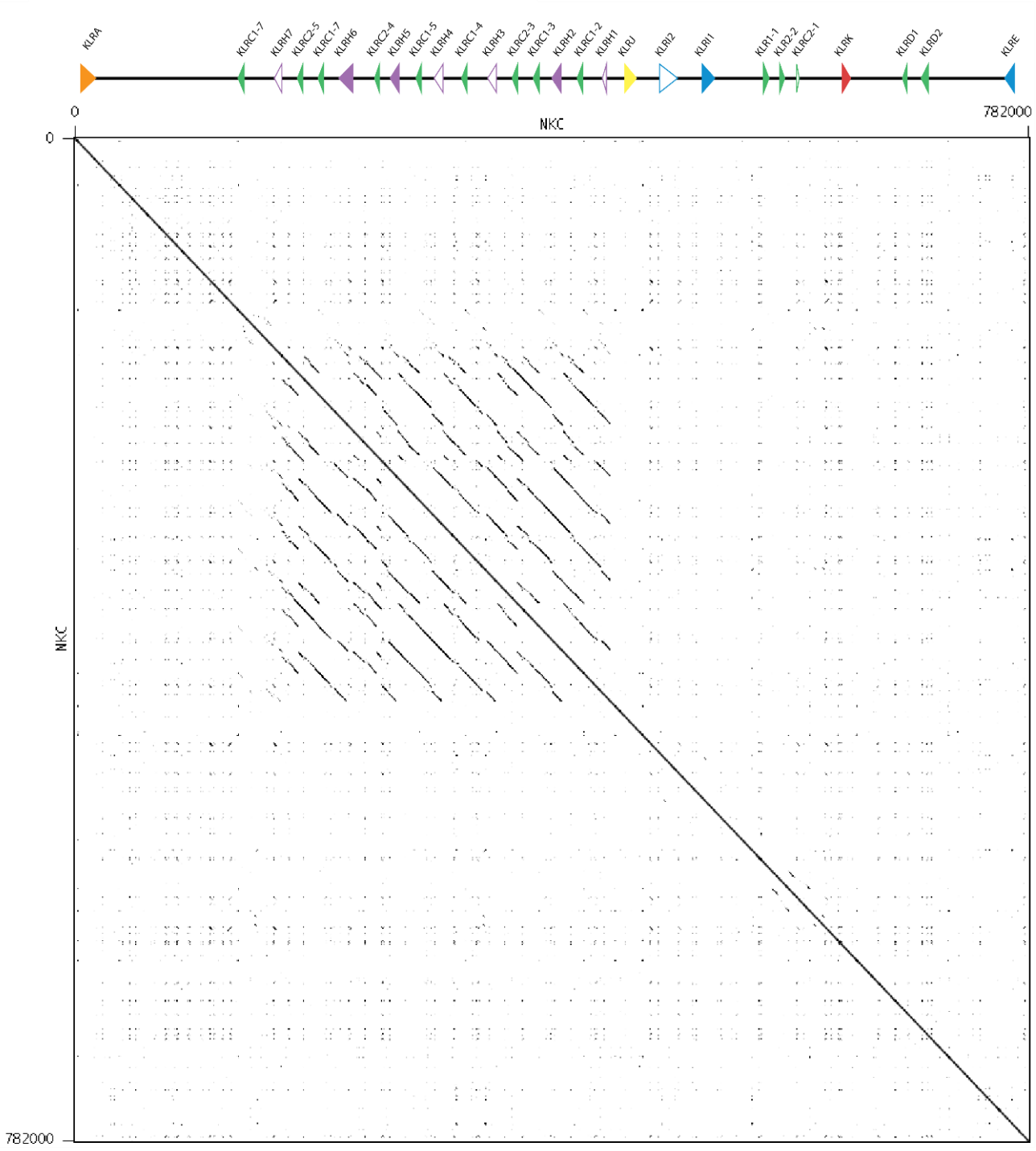


Figure 2-3. Dot plot of the cattle NKC against itself. The dot plot was generated using Gepard v1.40 (Krumsiek, Arnold, and Rattei 2007) with a word length of 10. The NKC sequence was extracted from UCD-ARSv0.1. Orange arrows indicate *KLRA* genes, green arrows indicate *KLRC/D* genes, purple arrows indicate *KLRH* genes, blue arrows indicate *KLRI/E* and red arrows indicate *KLRK*. Open arrows are genes predicted to be non-functional and filled arrows are predicted to be functional.

2.2 Methods

2.2.1 Custom genome construction

A pre-release of the *Bos taurus* genome assembly (ARS-UCD1.0.14), the first cattle assembly to utilise PacBio sequencing was obtained from the USDA-ARS- (Ben Rosen and Tim Smith). The region of ARSv14 containing the LRC (chr18: 62393497-63519120) was masked using bedtools maskfasta (Quinlan and Hall 2010). The haplotype 1 LRC sequence (Sanderson et. al, 2014) was inserted as an additional chromosome (chrLRC). This custom genome will be referred to as ARSv14Hap1.

Variations of ARSv14Hap1 were created by using bedtools maskfasta to mask selected *KIR*. ARSv14Hap1_nonfunc had predicted non-functional genes masked and ARSv14Hap1_genefamiles had both non-functional genes and all but one member of each of the *KIR* gene families removed, with the exception of *3DXL6* which remained as its own gene family. ARSv14Hap1_genefamiles_3DXL6 is identical to ARSv14Hap1_genefamiles except for the masking of *3DXL6* in ARSv14Hap1_genefamiles_3DXL6.

2.2.2 Custom transcriptome construction

The genome sequence and annotation for the UMD3.1 cattle genome assembly was obtained from Ensembl and transcript sequences generated using gem-retriever. A combination of BLAST (Altschul et al. 1990) and an annotation search was used to identify and remove NKC and LRC sequences and NKC transcripts from ARS-UCD1.0.14 and LRC transcripts from haplotype 1 were added.

2.2.3 RNA-Seq read simulation

Transcript sequences were generated from ARSv14Hap1 using gem-retriever (Marco-Sola et al. 2012) and an in-house annotation of the LRC and NKC. Transcript sequences were merged into a pool and separated by a known character. The pool was split into reads of specified length with 0 error in respect to the reference transcript and reads containing the known character discarded. Reads were converted into fastq format and each base given the maximium score in the Illumina Q33 quality encoding format. The exact location on the transcript that the simulated read originates from was written into the

fastq header for each read. The awk script used for read simulation is available in the appendix (7.1.1).

2.2.4 Mappability calculations

An index of the ARSv14Hap1 custom genome was generated using the GEM indexer (Marco-Sola et al. 2012). GEM-Mappability (Derrien et al. 2012) was run on the index with the approximation-threshold set to 23. The output from GEM-Mappability was converted to bed format using gem-2-bed mappability. Per exon mappability was calculated with awk and bash in a bespoke script (appendix – 7.1.2).

2.2.5 RNA-Seq read mapping

Reads were mapped to the ARSv14Hap1 with the GEMTools rna-pipeline (Lappalainen et al. 2013) using the GEM index generated previously for mappability calculations and the annotation used during RNA-Seq read simulation or the custom transcriptome. Default parameters were used with the exception of mismatches allowed set to 0.02 (2%). Weighted read counts generated from the GEMtools rna-pipeline were used as a standard in pipeline development. Mapping to the custom transcriptome described previously was carried out using GEM-mapper version 3 in sensitive mapping mode.

2.2.6 UniMMap read counts

When mapping to the genome, alignments passing the quality filter (GEM score ≥ 15360) were extracted from the map file created by GEMTools rna-pipeline. Reads mapped to the transcriptome were extracted if they had a score greater than 0 (equivalent to ≥15360 in the GEMTools pipeline). The mappability scores of k = read-length were converted to bed format and intersected with the extracted alignments to determine the mappability of each alignment. Using awk the geometric mean of reads with multiple alignments was calculated. Reads mapped to a chromosome containing a regions of interest were extracted and placed in separate files to reduce later file size and analysis time. These files were intersected with the annotation of the regions of interest to identify reads mapping to the exons of each gene. The sum of the average mappability of reads as well as the sum of reads with an average mappability of 1 (unique) mapping to each gene was calculated. Genes that did not have any reads

mapping with an average mappability of 1 were removed from the gene counts. This pipeline is subsequently referred to as 'UniMMap'. Read counts for this chapter are available at: https://github.com/richard-borne/PhD_thesis_data

2.2.7 Length analysis of datasets on the Sequence Read Archive (SRA)

The SRA database (Leinonen et al. 2011) was queried on 25/04/2018 to include only RNA-Seq data sets originating from RNA that also contained the *Bos taurus* species identifier and the run table downloaded. Runs were filtered with awk to only include RNA-Seq experiments. Average spot length was divided by 2 if the run contained paired ends. Runs were binned on read length and the number of runs belonging to each bin calculated in awk.

2.3 Results

2.3.1 Multi-mapping is highly prevalent in the LRC and NKC

Due to the presence of multi-mapping reads, mapping simulated 150bp single-end reads to ARSv14Hap1 results in an average of 3.2× alignments generated to the LRC for each read simulated, if each alignment is given a weight 1 (figure 2-4). All genes in the LRC have more reported alignments than number of reads simulated. Most of the predicted functional *KIR* genes show a large discrepancy between the number of reads simulated and the number of alignments generated. The functional genes that are particularly problematic are *3DXL7* (6.8×), *3DXL4* (4.9×), *3DXL6* (4.4×), *3DXL5* (3.9×) and *3DXS1* (3.2×). The lowest difference of the functional *KIR* is *KIR2DL1* (1.27×) which is close to *FCAR* (1.3×) and *NCR1* (1.2×), the non-repetitive genes of the LRC. Despite reads not being simulated from the non-functional *KIR,* alignments were observed to almost all with the exception of *1DP1* and *1DP2*.

Mapping of simulated reads to the NKC of ARSv14Hap1 (figure 2-5) results in less erroneous alignments on average (2.8×) compared to the 3.2× observed in the LRC. However *KLRC1-6* (10.4×) and *KLRC2-5* (8.5×) have more multi-mappers than any of the genes of the LRC. *KLRH2* (0.84×) and *KLRC1-2* (0.4×) are the only genes of either complex to generate less alignments than there were reads simulated. Of the non-functional KLR of which no reads were simulated, only *KLRH7* and *KLRH1* did not have any reads aligning to them. The non-functional *KLRH3* had 15710 alignments despite 0 *KLRH3* reads simulated.

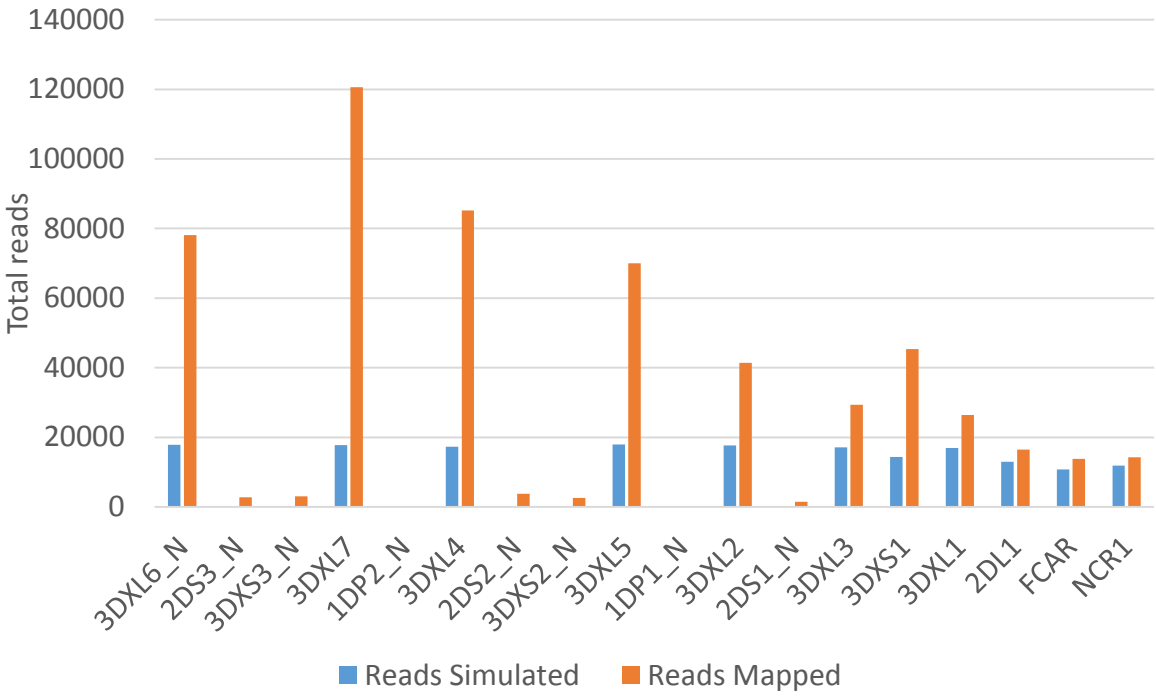


Figure 2-4. Total number of simulated reads generated and subsequently mapped to each gene of the LRC. The number of reads simulated for each gene is shown as blue bars. Orange bars indicate the total number of alignments generated when the simulated reads were mapped to ARSv14Hap1. Predicted non-functional genes are denoted with '_N' after the gene name.

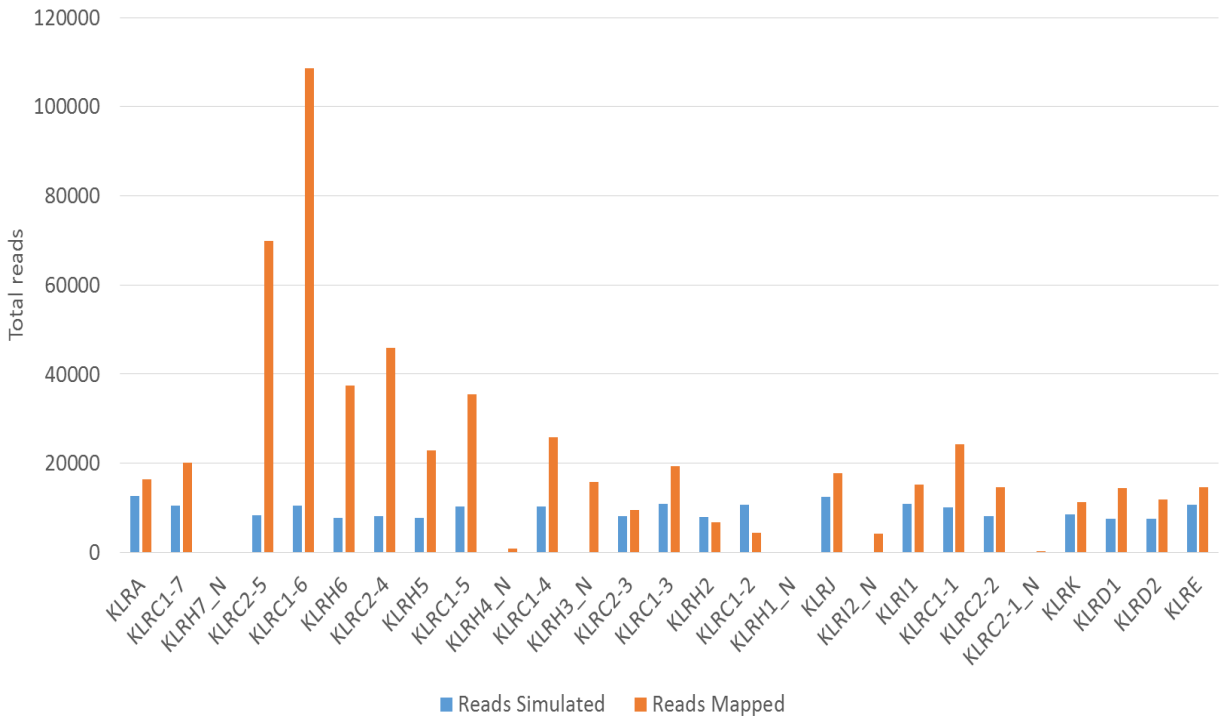


Figure 2-5. Total number of simulated reads generated and subsequently mapped to each gene of the NKC. The number of reads simulated for each gene is shown as blue bars. Orange bars indicate the total number of alignments generated when the simulated reads were mapped to ARSv14Hap1. Predicted non-functional genes are denoted with '_N' after the gene name.

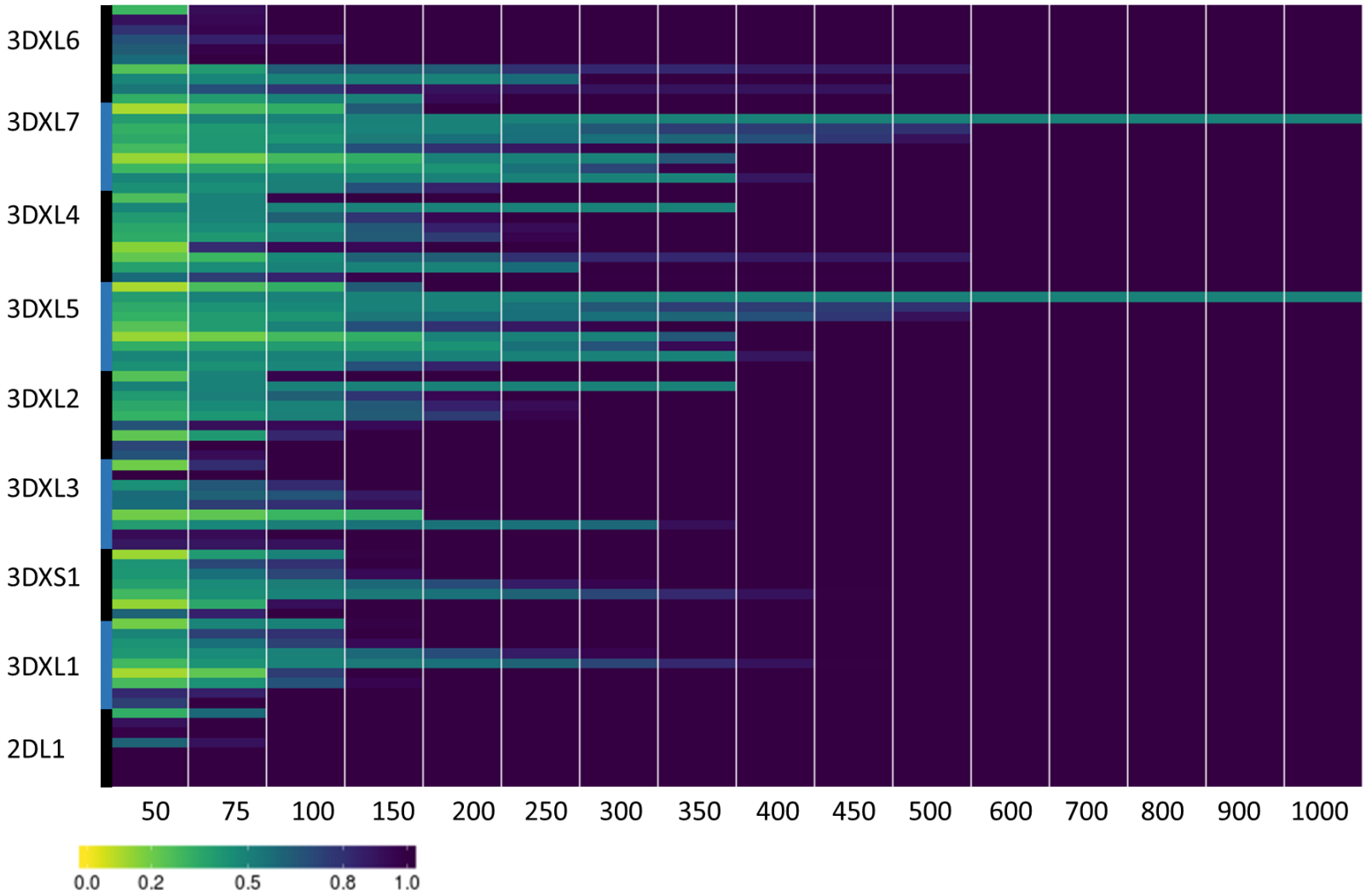


Figure 2-6. Mappability of the functional *KIR* genes in the LRC at different *k-mer* lengths. Mappability or 'uniqueness' is a score describing quantitatively how many times a read length of *n*bp occurs in a reference sequence, a *k-mer* with a mappability score of 1 occurs once, with a score of 0.5 occurs twice and so on. Per exon mappability calculated using GEM mappability is shown, separated by gene on the y axis. The *k-mer* length for which the mappability calculations were carried out is indicated on the x axis.

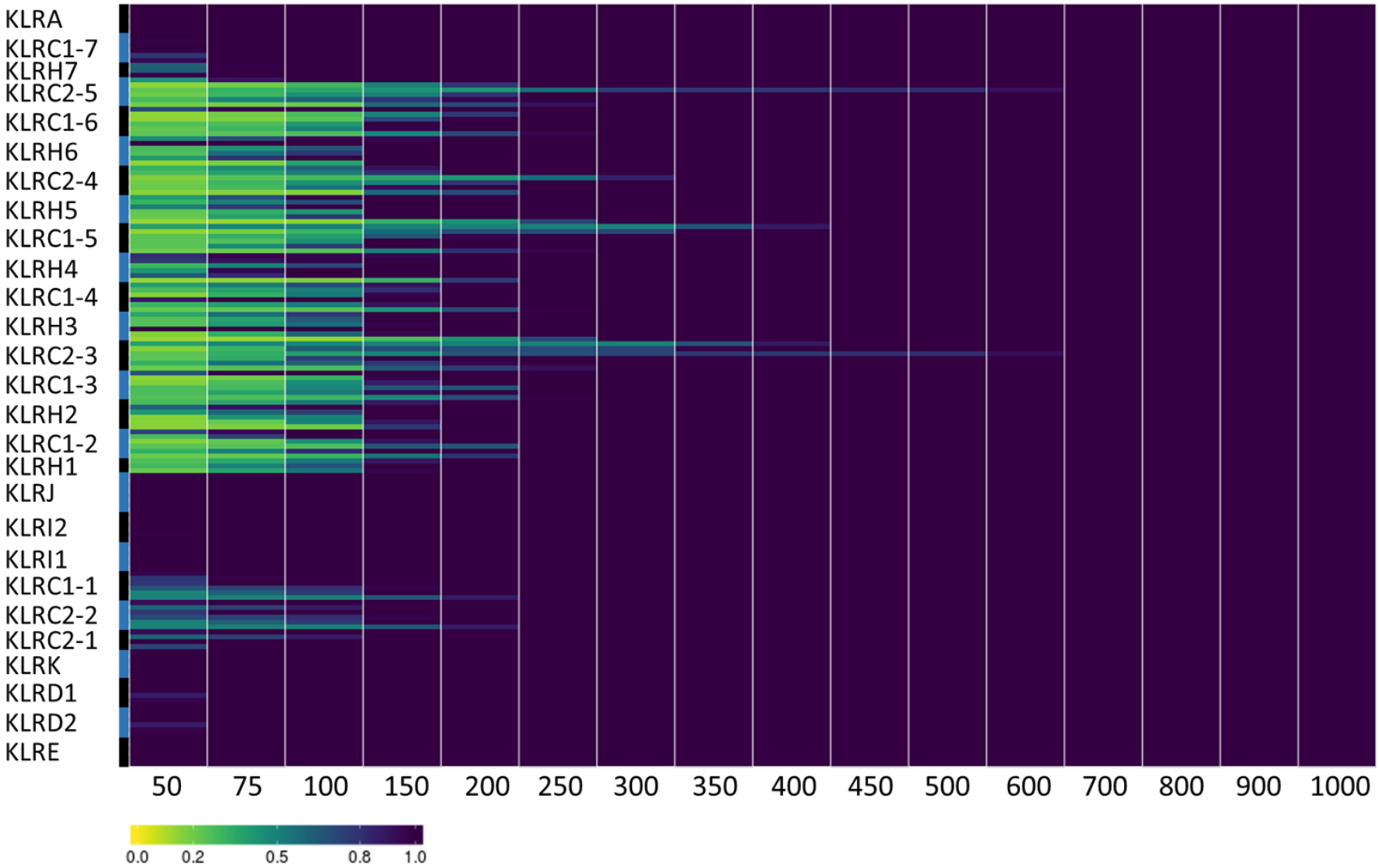


Figure 2-7. Mappability of the *KLR* genes in the LRC at different *k-mer* lengths. Mappability or 'uniqueness' is a score describing quantitatively how many times a read length of *n*bp occurs in a reference sequence, a *k-mer* with a mappability score of 1 occurs once, with a score of 0.5 occurs twice and so on. Per exon mappability calculated using GEM mappability is shown, separated by gene on the y axis. The *k-mer* length for which the mappability calculations were carried out is indicated on the x axis.

2.3.2 Maximum Illumina sequence length does not resolve mappability

To understand how the repetitive nature of the LRC and NKC was impacting multi-mapping, the mappability of the genome was calculated. Mappability is a score describing quantitatively how many times a *k-mer* of length *n* bp occurs within the genome. Calculating mappability allows an accurate estimation of how many alignments a sequencing read from that location will generate. With the exception of *2DL1*, none of the *KIR* are completely uniquely mappable at a *k-mer* length of 150bp (figure 2-6). Until recently, 150bp was the maximum read length of an Illumina NextSeq, HiSeq or NovaSeq, the only Illumina machines capable of whole transcriptome sequencing of a large mammalian genome. The Illumina NovaSeq 6000 is however capable of generating 250bp, paired end reads at a depth suitable for large mammalian transcriptome sequencing. At 150bp, *3DXL5* and *3DXL7* do not have any exons that are uniquely mappable. 200bp is the first *k-mer* length tested where every *KIR* has at least one uniquely mappable exon. At 300bp 60% of total exons have a mappability of 1 and *3DXL5/7* are the only *KIR* not to have the majority of exons uniquely mappable. By 600bp, all of the exons of the *KIR* with the exception of *3DXL5* and *3DXL7* have a mappability score of 1. At the longest tested *k-mer* length *3DXL5/7* still has the same mappability score as at a *k-mer* length of 600bp.

Figure 2-7 shows the mappability of the NKC at 16 different *k-mer* lengths between 50-1000bp. The genes between and including *KLRC2-5* and *KLRH1* are almost entirely non-unique between 50-100bp with 2.5% exons unique at 50bp and 13.6% at 100bp. At a *k-mer* length of 150bp mappability improves with 39.5% of the exons in this region uniquely mappable and 50% of the total exons have a mappability score of 1. 90.4% of exons with the NKC are uniquely mappable at a *k-mer* length of 300bp. At 600bp only two exons are not completely uniquely mappable, *KLRC2-5* and *KLRC2-3* each have an exon with an average mappability of 0.92. The first tested *k-mer* length where all exons have an average mappability of 1 is 700bp.

On average the LRC is more mappable than the NKC. However at the short read lengths typically used in RNA-Seq experiments, the highly repetitive region of the NKC is particularly problematic.

2.3.3 The Sequence Read Archive (SRA) contains a large number of *Bos taurus* RNA-Seq datasets

The run information of the *Bos taurus* RNA-Seq datasets available on the Sequence Read Archive (SRA) as of the 25/04/2018 was investigated to determine how many would be suitable for transcriptional analysis of the LRC and NKC. Of the 6974 *Bos taurus* RNA-Seq projects available on the SRA, 25.5% are ~50bp or less and 74.8% are ~100bp or less (figure 2-8). The 5 largest out of the 191 total bioprojects contain 33.6% of the total runs and the largest 10 contain 46.4%. Within the 5 largest bioprojects, 49% are ~100bp, 15.2% are ~75bp and 35.8% are ~50bp. 29.9% of the runs contain single end reads and 70.1% contain paired end reads. 98.9% of the *Bos taurus* RNA-Seq datasets were generated using an Illumina machine, 79.4% of the total runs were sequenced on an Illumina HiSeq and 4.7% on a NextSeq. The remainder of the runs were generated on an Applied Biosciences (0.5%), 454 Life Sciences (0.6%), Ion Torrent (0.27%) or PacBio machine (0.33%). Our mappability data suggests that any transcriptional analysis of the LRC and NKC genes that was carried out on these samples will most likely not have been accurate. Analysis of LRC transcription within these samples would also have been inaccurate due to the very poor quality of the LRC assembly in the reference genome.

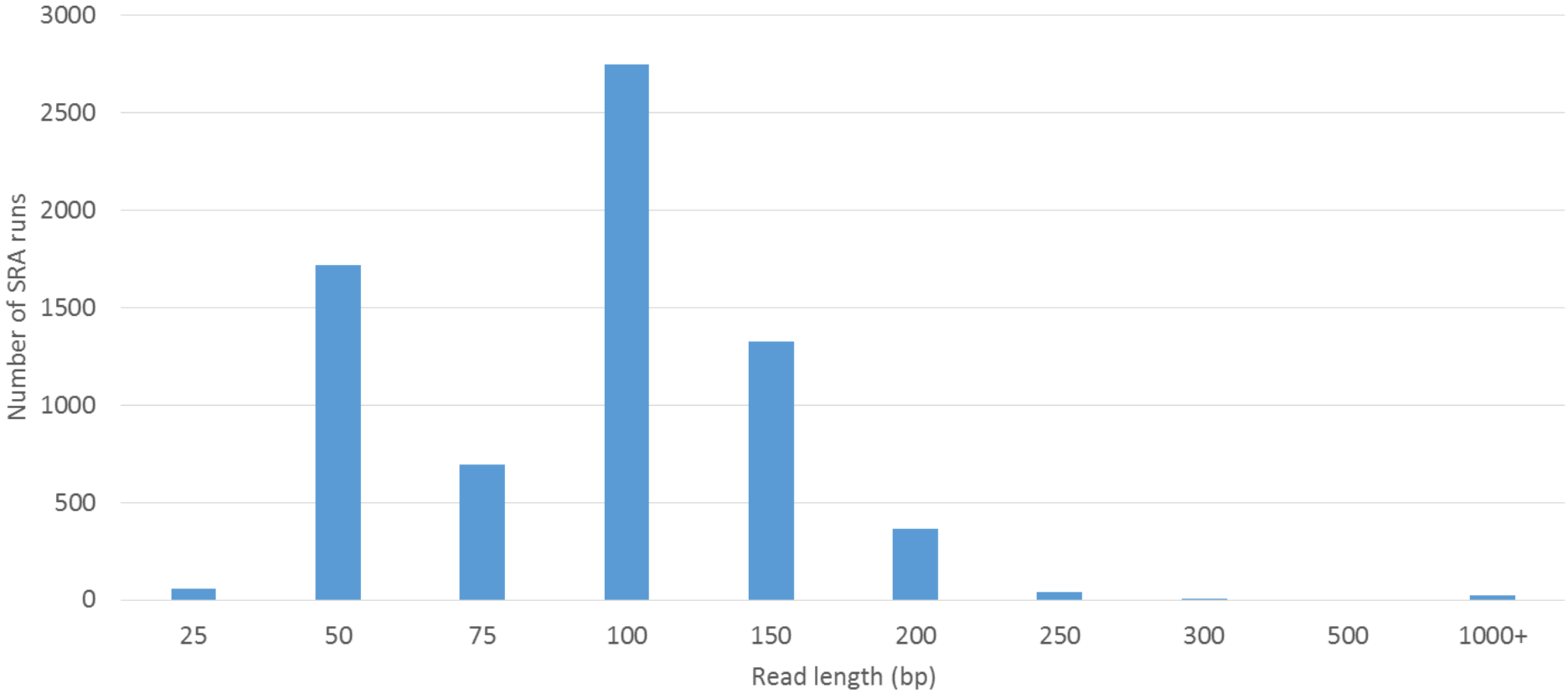


Figure 2-8. Histogram of read length of *Bos taurus* RNA-Seq datasets available on the Sequence Read Archive. Average spot length was divided by two in the case of runs containing paired-end data. Run read lengths are binned to represent the closest likely read length before trimming.

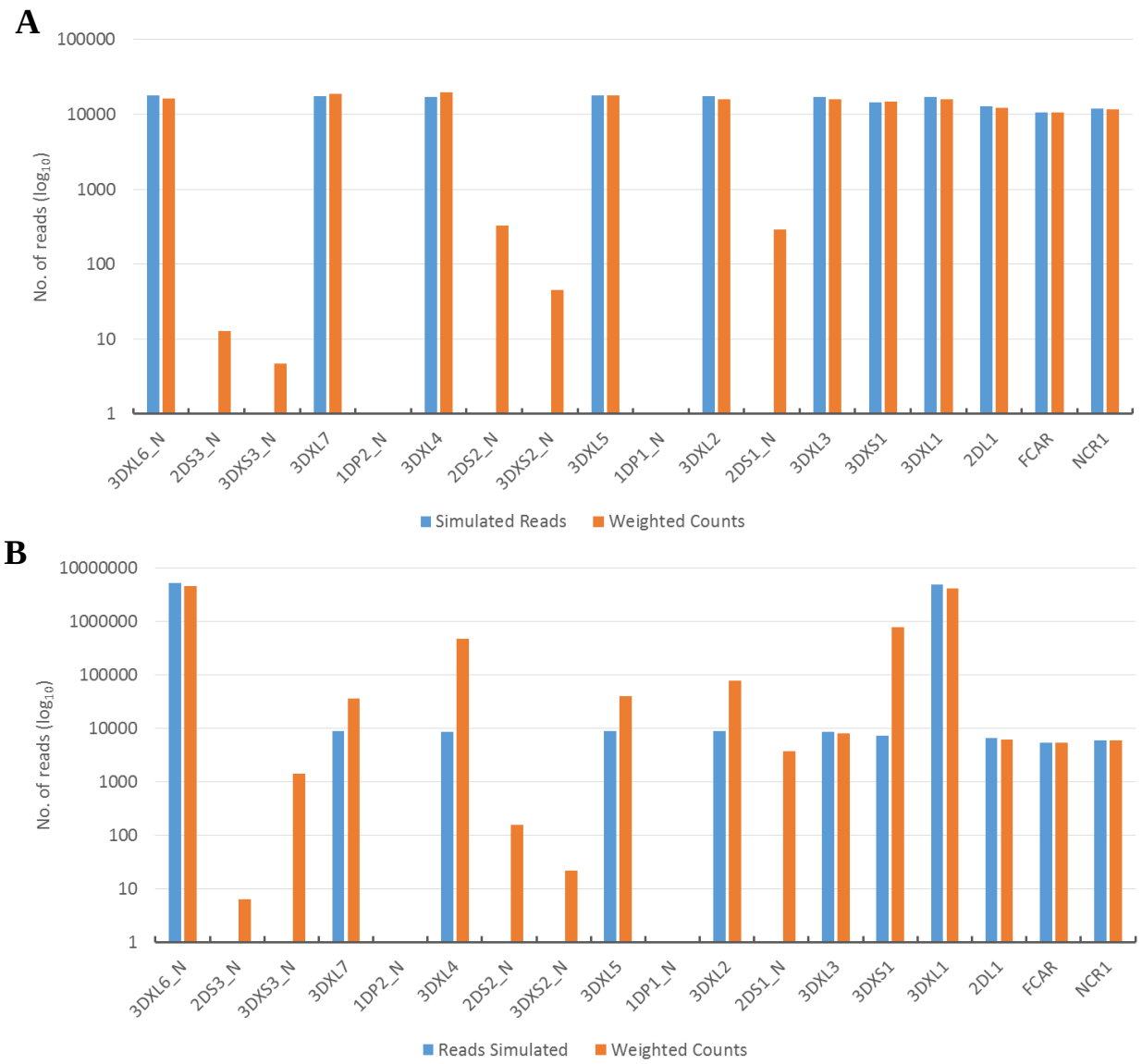


Figure 2-9. Number of reads simulated and weighted counts for each gene of the LRC. Reads of either even coverage (A) or mixed coverage (B) were simulated and mapped to ARSv14Hap1. The number of reads simulated for each gene are shown as blue bars and the weighted counts generated by GEMTools rna-pipeline as orange bars.

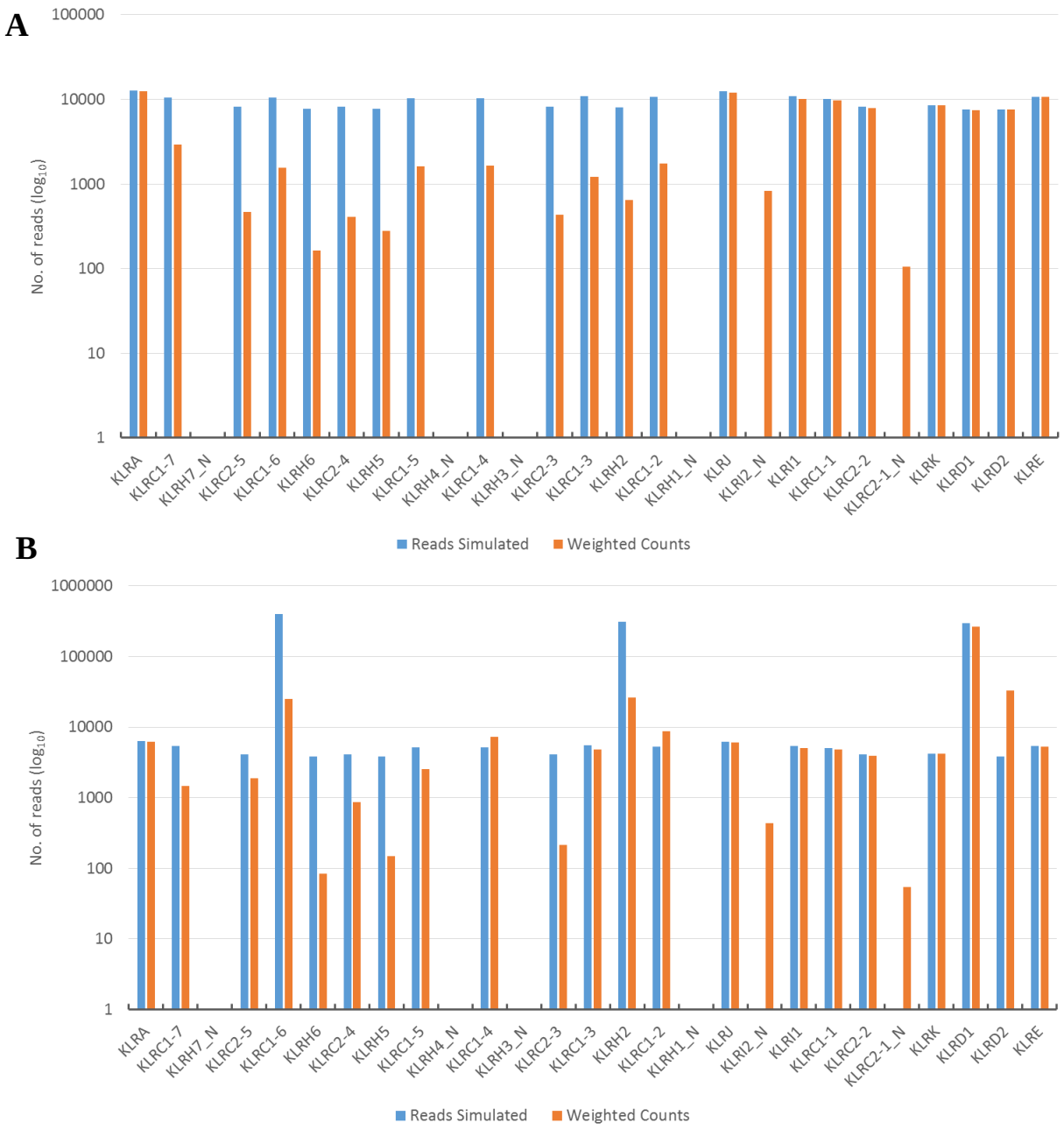


Figure 2-10. Number of reads simulated and weighted counts for each gene of the NKC. Reads of either even coverage (A) or mixed coverage (B) were simulated and mapped to ARSv14Hap1. The number of reads simulated for each gene are shown as blue bars and the weighted counts generated by GEMTools rna-pipeline as orange bars.

2.3.4 Accuracy of weighted counts is impacted by uneven read coverage

Weighted read counts are counts where each sequencing read has been weighted based on the number of different alignments it generates. For example a read mapping to two genes will add 0.5 to the read count of both genes. Weighted counts perform well when the number of sequencing reads simulated is even for all functional genes in the LRC (figure 2-9A). It should be noted that in this and the following sections we are performing the simulations without introducing errors, as they might also impact the accuracy of the weighted counts. Of the functional *KIR*, only *3DXL4* (1.14×), *3DXL7* (1.07×) and *3DXS1* (1.04×) had a read count higher than the number of reads simulated. The average difference between simulated and weighted counts for the functional *KIR* was low (0.99×). With the exception of *1DP1* and *IDP2* all of the predicted *KIR* pseudogenes were reported to have reads mapping to them. *2DS1* (287.5 reads) and *2DS2* (328.7 reads) were the two predicted non-functional with the highest read count and *3DXS3* was the lowest with a read count of 4.67.

To determine how weighted counts would perform in a more biologically relevant dataset, a mixed coverage dataset was created with 100× the number of reads originating from *3DXL6* and *3DXL1* compared to any other gene (figure 2-9B). For this dataset weighted counts perform poorly compared to the even coverage dataset, due to a larger number of reads leaking from one gene to another due to the presence of repetitive regions. The mixed coverage resulted in a 20.34× higher average of the weighted counts than the number of reads simulated for the functional *KIR*. This average is heavily skewed by *3DXS1* (107.9×) and *3DXL4* (54.5×) and without them the average drops to 2.93×, although this is still much higher than the 0.99x average of the even coverage dataset. Once again read counts are reported to all of the predicted non-functional *KIR* that reads were not simulated for, with the exception of *1DP1* and *IDP2*. A total of 5324.67 reads were counted as mapping to the non-functional genes, 3653.83 (68.62%) of which map to *2DS1*.

Unlike the LRC, weighted counts do not perform well for the genes of the NKC when reads of even coverage are simulated (figure 2-10A). In contrast to the LRC the reported read counts for the NKC are much lower than the number of reads simulated. The average difference between number of reads simulated and read counts for the functional *KLR* is 0.48×. None of the reported read counts were

higher than the number of reads simulated. The region of the NKC shown as having poor mappability scores (figure 2-7) were significantly underreported in the read counts. The average difference of the mixed coverage dataset (figure 2-9B) is again higher (6.16×) than the even coverage dataset (0.48×). *KLRD1* was the most problematic with a difference of 76.82× and *KLRC1-4* (12.48×), *KLRC1-3* (7.73×) and *KLRC2-5* (4.06×) were also massively over represented. The pseudogenes *KLRI2* and *KLRC2-1* had reported read counts despite no reads simulated.

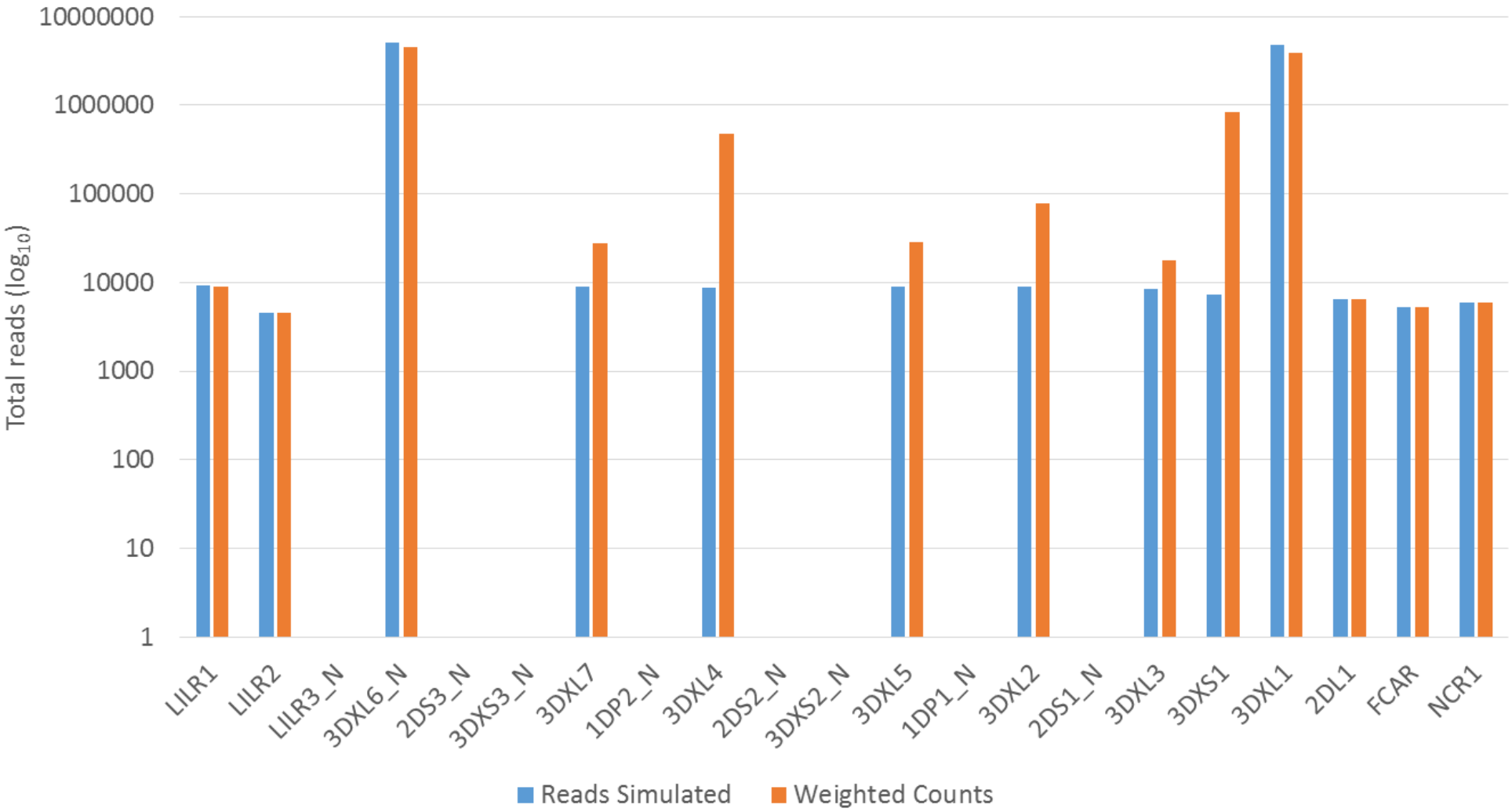


Figure 2-11. Number of reads simulated and weighted counts for each gene of the LRC after masking of non-functional *KIR*. Reads of mixed coverage were simulated and mapped to ARSv14Hap1_nonfunc. The number of reads simulated for each gene are shown as blue bars and the weighted counts generated by GEMTools rna-pipeline as orange bars.

2.3.5 Masking of non-functional *KIR* in the genome has little impact on functional *KIR* multi-mapping

Masking of the predicted non-functional *KIR* in ARSv14Hap1 was carried out to create ARSv14Hap1_nonfunc. The same simulated mixed coverage dataset as used in figure 2-9 was mapped to ARSv14Hap1_nonfunc and GEMTools rna-pipeline used to create weighted counts (figure 2-11). The difference between numbers of reads simulated and read counts for the functional *KIR* unsurprisingly slightly increases from 20.34× to 21.19× after the masking of non-functional genes. *3DXS1* changes from 107.9× to 116.2× after masking and *3DXL4* from 54.5× to 54.72×. As expected the number of reads mapping to the predicted non-functional *KIR* drops from 5324.76 to 0.

2.3.6 Grouping the functional *KIR* into gene families reduces total multi-mapping

Merging the *KIR* into their established gene families (Sanderson et al, 2014) was undertaken in an attempt to reduce multi-mapping as it was assumed most multi-mapping would occur within these families as some alleles share very high sequence identity. The sequence similarity between *3DXL6* and the other two members of its gene family, *3DXL2* and *3DXL4*, is lower than within other gene families. For this reason two variations of the ARSv14Hap1 genome were created, one with a gene family consisting of *3DXL2/4/6* (figure 2-12A) and another with *3DXL2/4* grouped together and *3DXL6* separated on its own (figure 2-12B).

Merging *3DXL6* into a gene family with *3DXL2/4* increases multi-mapping for every gene family with the exception of the newly created *3DXL2/4/6* gene family for this mixed coverage dataset. Before merging, *3DXL6* is fairly accurately quantified with a difference of 0.91x between *3DXL6* reads simulated and subsequently mapped to it. Multi-mapping to the *3DXL3/5/7* gene family increases from 2.58× to 31.2× after masking of *3DXL6* in the genome. These increases are likely due to the fact that *3DXL6* is one of the genes with a high number of reads simulated and subsequent masking means they generate poorer alignments to multiple other *KIR*. Although multi-mapping is ultimately reduced, a large degree of gene transcriptional information is lost by grouping into gene families.

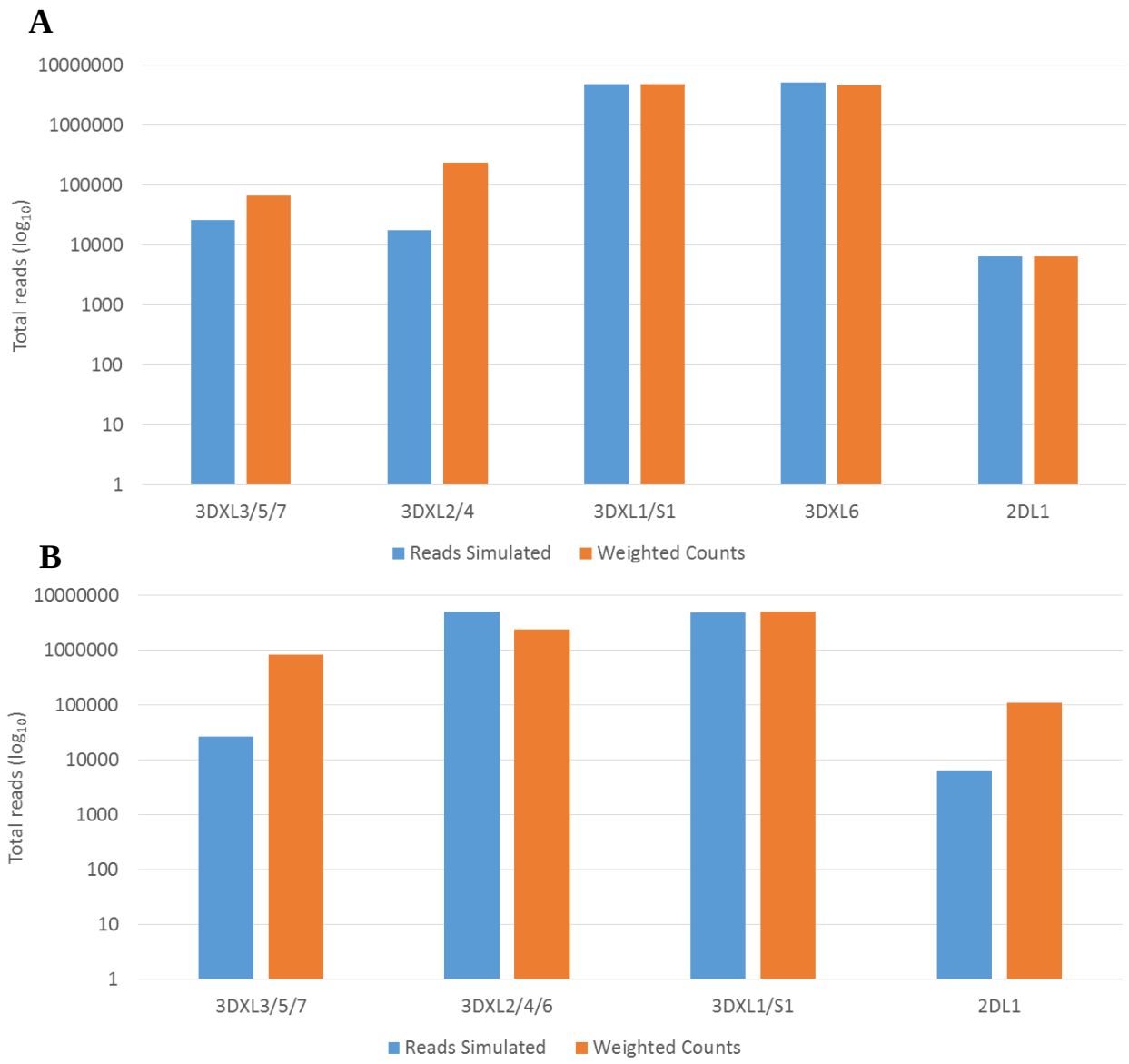


Figure 2-12. Number of reads simulated and weighted counts for gene families of the LRC. The mixed coverage simulated dataset was mapped to ARSv14Hap1_genefamilies (A) and ARSv14Hap1_genefamilies_3DXL6 (B) as described in the methods. The number of reads simulated for each gene family are shown as blue bars and the weighted counts generated by GEMTools rna-pipeline as orange bars.

2.3.7 Filtering alignments on alignment score or mappability reduces multi-mapping

Two alignment filtering strategies were implemented to reduce the number of erroneous alignments generated to the LRC and NKC. One method involved filtering alignments based on their GEMTools RNA pipeline score, which denotes the confidence of the pipeline in the uniqueness of the alignment. Alignments with a GEM score of 15360 or greater were selected so that only the highest quality alignments were carried forward. The other method examined was to filter for reads that aligned to a region with a mappability score of 1 at a *k-mer* length the same as read length. These two filtering techniques were also combined and compared to the individual filters.

Multi-mapping to the functional *KIR* of the LRC is substantially reduced between weighted counts and the GEM and combined filtering strategies (figure 2-13A, 14B). The average difference between simulated reads and read counts decreases from 20.4× in the case of weighted counts to 3.12× when filtering on GEM score and to 1.44× when the filters are combined. In contrast, the average difference in the case of the mappability filter was 34.3×, larger than the weighted counts. However, whilst the average difference reduces for the GEM filtering and combined strategies, the individual filters actually increases the number of reported reads by 1.2× for GEM and 1.7× for unique regions when compared to weighted counts. The 1.01× increase when applying both filters is negligible. These averages are also heavily skewed by a large amount of multi-mapping occurring to *3DXS1*. Removing *3DXS1* from the average results in a 9.4× difference for weighted counts, 6.4× for the mappability filter, 1.23× for the alignment filter and 0.60× when combining the filters. The large amount of multi-mapping to *3DXS1* is most likely due to the majority of *3DXL1* and *3DXS1* being almost identical, with the exception of the sequence encoding the intracellular tail, as well as the mismatches allowed during RNA-Seq mapping.

As with the LRC, when no filtering is applied to the generated alignments, a very large percentage of the alignments are generated from a different gene of origin for the majority of the genes of the NKC (figure 2-13B). Applying the GEM filter results in an average difference between numbers of reads simulated and read

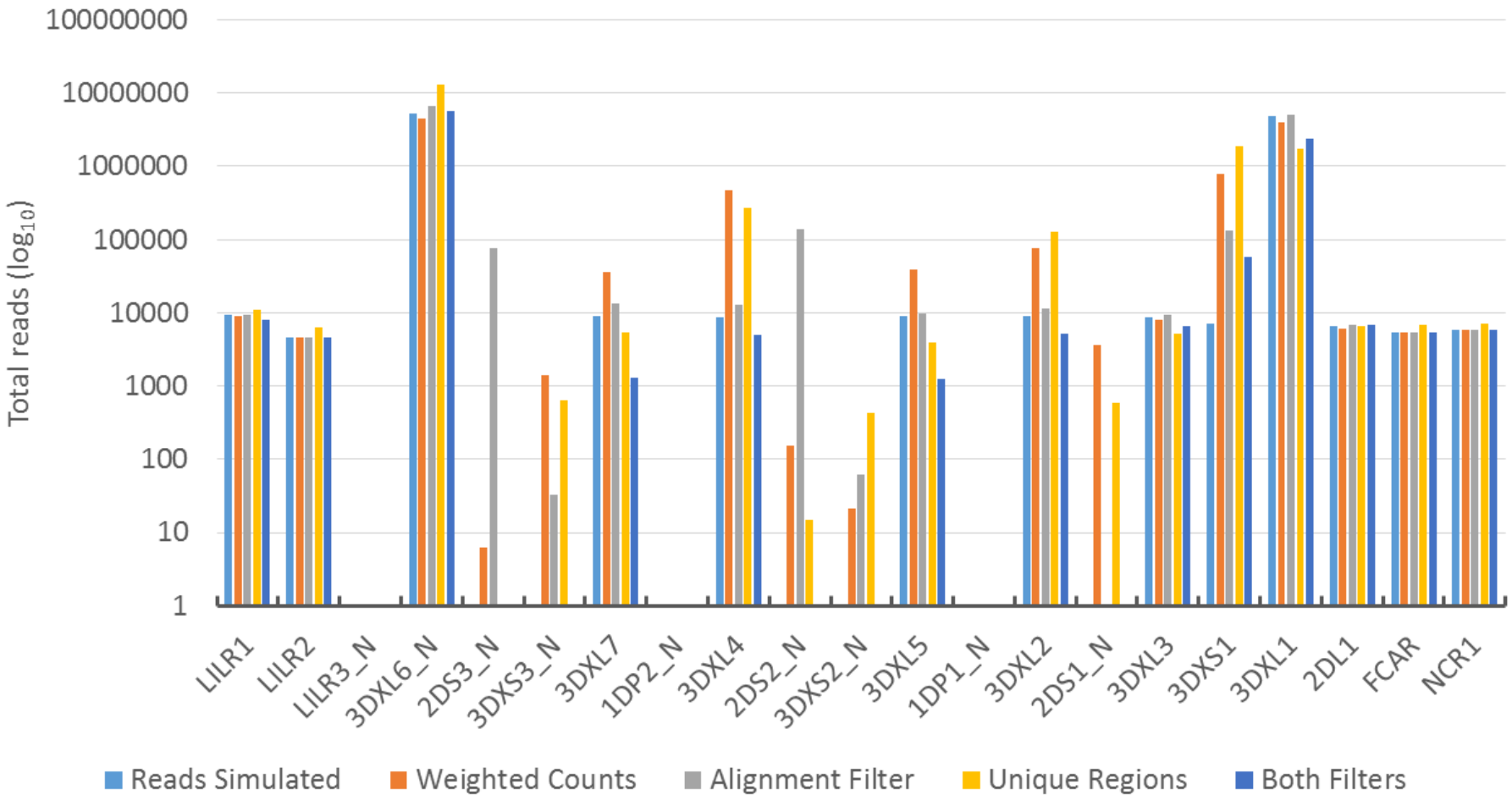
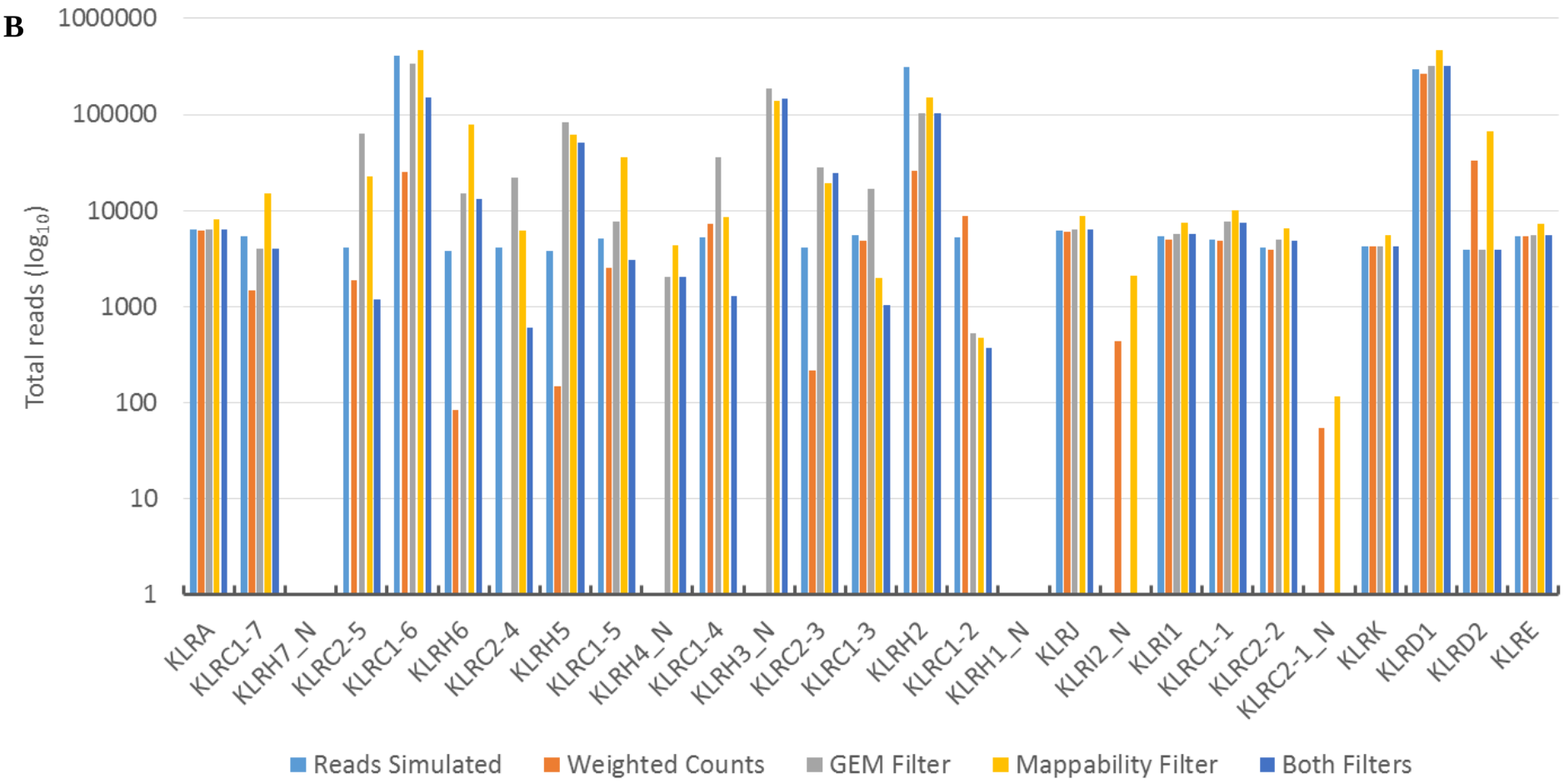


Figure 2-13. Comparison of different methods of eliminating multi-mapping. Reads of mixed coverage were simulated for the LRC (a) and NKC (b) and mapped to ARSv14Hap1. The number of reads simulated for each gene are shown as light blue bars. Weighted counts generated by the GEMTools rna-pipeline as orange bars. Alignments generated by GEMTools rna-pipeline were filtered based on their GEM score (=<15360) and the number of filtered alignments to each gene counted (grey bars). Alignments were also filtered based on whether they mapped to a region of mappability 1 at k=150 and those that did counted for each gene (yellow bars). Read counts after applying the GEM score filter followed by the mappability filter are shown as dark blue bars.

B


1000000
100000
10000
1000
100
10
1
Total reads (log10)
KLRA
KLRC1-7
KLRH7_N
KLRC2-5
KLRC1-6
KLRH6
KLRC2-4
KLRH5
KLRC1-5
KLRH4_N
KLRC1-4
KLRH3_N
KLRC2-3
KLRC1-3
KLRH2
KLRC1-2
KLRH1_N
KLRJ
KLRI2_N
KLRI1
KLRC1-1
KLRC2-2
KLRC2-1_N
KLRK
KLRD1
KLRD2
KLRE
Reads Simulated
Weighted Counts
GEM Filter
Mappability Filter
Both Filters

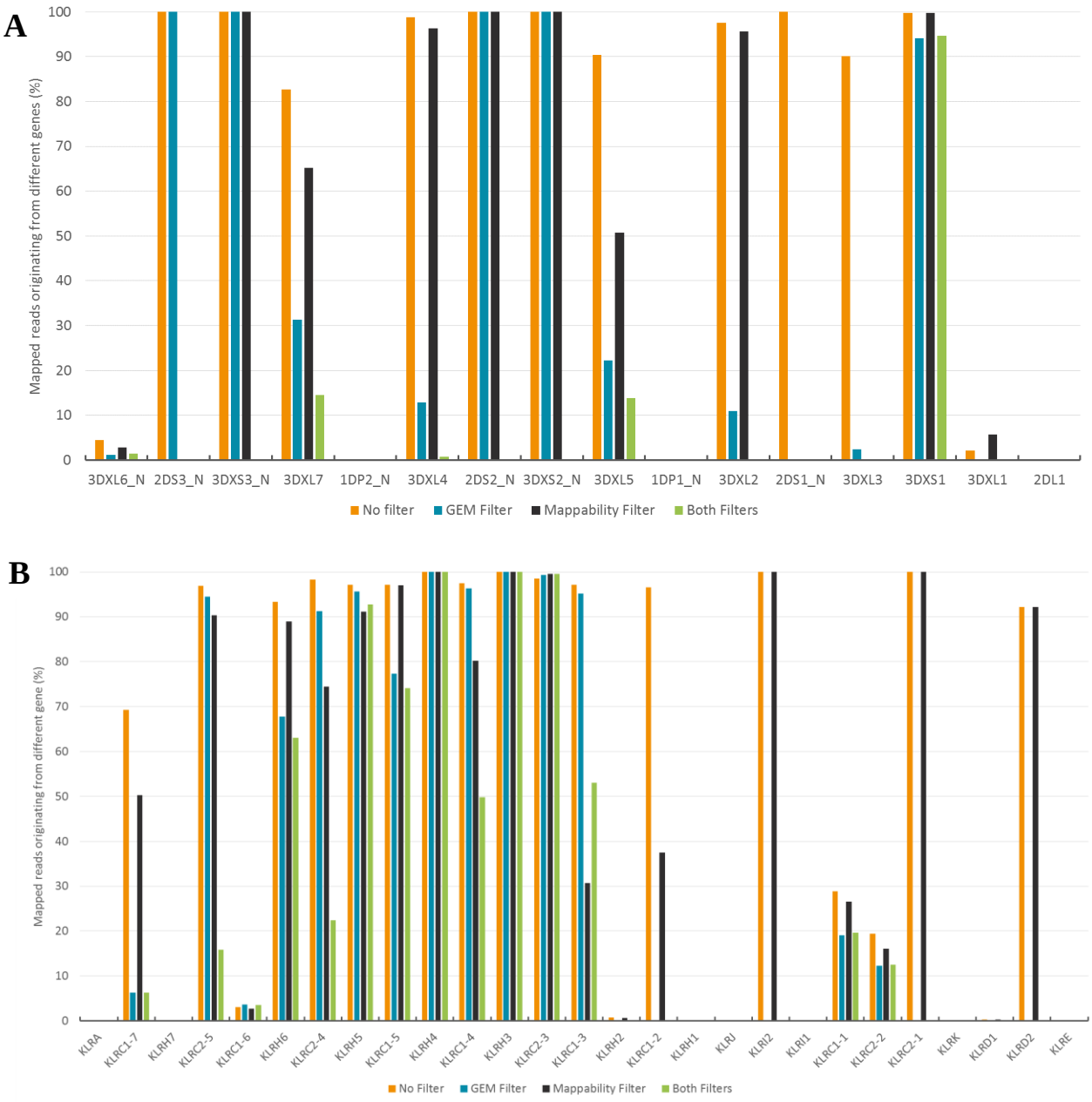


Figure 2-14. Percentage of reads mapping to each gene that originate from a different gene. Reads of mixed coverage were simulated and mapped to ARSv14Hap1 separately for each the LRC (A) and the NKC (B). The origin of the reads mapping to each gene was determined and the percentage originating from a different gene calculated when no filter was used and for each of the two alignment filters as well as the filters combined.

counts of 3.67×, the mappability filter results in a 4.33× difference and when applying both a 1.69× difference is observed (figure 2-13B). None of the filters eliminate the large number of incorrectly mapped reads to the non-simulated *KLRH3*.

2.3.8 UniMMap performs better than standard weighted counts for the LRC but is inaccurate for the NKC

We created a pipeline named UniMMap that filters reads based on their alignment score and weights counts on the average mappability of the regions they map to. It also discounts read counts for genes that do not generate any alignments to regions of unique mappability results. An absence of any such alignments means transcription cannot be accurately determined. Figure 2-15A shows that UniMMap is much more accurate than standard weighted counts for the LRC. As shown initially in figure 2-9B, standard weighted counts results in 20.4× the number of reads reported to functional *KIR* compared to the number actually simulated. In comparison UniMMap results in a 2.29× difference, which again is heavily skewed by multi-mappings between the almost identical *3DXL1* and *3DXS1*. If *3DXS1* is removed from the average, standard weighted counts drop to 9.4× and UniMMap drops to 0.92×.

The average difference between reads simulated and counts of functional KLR genes of the NKC is 6.16× when using weighted counts and 16.9× for the counts generated by UniMMap. Notably UniMMap is problematic for *KRH5*, *KLRC2-3* and *KLRC2-5* with 157.64×, 64.50× and 52.38× reported read counts compared to number simulated respectively. UniMMap also reports read counts of $1.6x10^6$ and $2.1x10^4$ for *KLRH3* and *KLRH4* respectively, two genes that did not have any reads simulated. We determined that the issue with the pipeline was caused by discrepancies in mappability due to the splicing that occurs during transcription and sought to alter the pipeline to account for this.

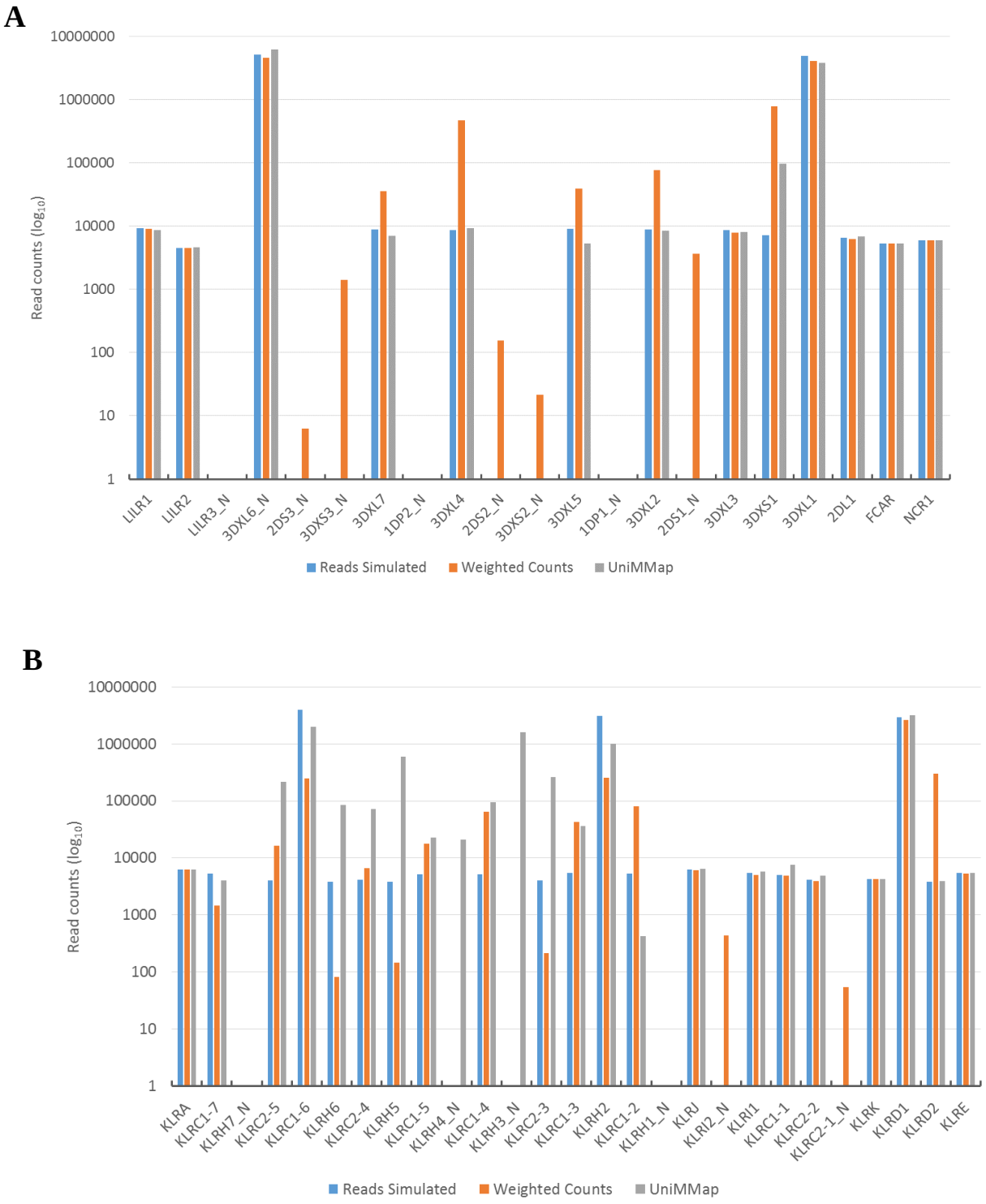


Figure 2-15. Number of reads simulated, standard weighted counts and UniMMap counts for each gene of the LRC (A) and NKC (B). Reads of mixed coverage were simulated and mapped using GEMTools to ARSv14Hap1. The number of reads simulated for each gene are shown as blue bars, weighted counts generated by GEMTools rna-pipeline as orange bars and UniMMap read counts as grey bars.

2.3.9 A precise annotation is crucial to the accuracy of UniMMap

During development of the pipeline using a rough draft LRC annotation there were missing read counts to *2DL1* which should be comparatively easy to map, an example of which can be seen on figure 2-16. With the draft annotation, 7 of the genes that had reads simulated had their read counts discarded by UniMMap as they did not contain any reads mapping to unique regions. This indicated there was a fundamental issue as included within these 7 were *NCR1* and *FCAR*, both of which have been accurately quantified at all stages. After creation of a precise annotation, UniMMap was tested again and generated read counts for the LRC in line with what was expected.

2.3.10 Mapping to the transcriptome results in a substantial accuracy increase for UniMMap counts

Due to the poor performance of UniMMap for the most repetitive genes of the NKC as well as *3DXL1/S1* of the LRC, a revised version of the analysis pipeline was created. Instead of the cattle genome for mapping and mappability calculations, the cattle transcriptome was used. Using the transcriptome instead of the genome provides a more accurate representation of the mappability of RNA-Seq reads that cross splice boundaries. The resulting UniMMap counts were compared to determine if transcriptome mapping resulted in an increase in accuracy (figure 2-17A). The average difference between reads simulated and UniMMap counts for the LRC drops from 2.29× to 0.76× and *3DXS1* decreases from 13.2× to 0.82×. The number of read counts caused by reads originating from a different gene is reduced to 0 for all genes of the LRC (figure 2-18A) when analysing reads that map to a location of unique mappability.

The difference between total number of reads simulated and total read count generated by UniMMap for the mixed coverage dataset is lower when mapping to the transcriptome (0.85×) than to the genome (2.47×) for the NKC (figure 2-17B). A large improvement can be seen for *KLRH5* (157.64× to 0.95×), *KLRC2-3* (64.50× to 0.70×) and *KLRC2-5* (52.38x to 0.71×) which were the most problematic NKC genes when mapping to the genome. In contrast to mapping to the genome, when mapping to the transcriptome none of the genes that did not have reads simulated have reported read counts. None of the genes are overrepresented in

the transcriptome mapping UniMMap counts. As with the LRC, there are no genes that have reported reads mapping to mappability unique regions originating from a different gene. This allows complete confidence in determining whether or not a gene is transcribed. These improvements in count accuracy, paired with confidence in determining if a gene is transcribed or not meant that transcriptome mapping was used for all subsequent analyses.

2.3.11 UniMMap can accurately quantify transcription of *KIR* gene families

Grouping the *KIR* had been previously shown as a method of reducing multi-mapping (figure 2-11) and was revisited and combined with UniMMap as a potential method of accurately quantifying transcription when unknown alleles are present in the RNA-Seq data. As before the merging of *3DXL6* with *3DXL2/4* decreases the accuracy of the UniMMap counts (figure 2-18B). The *3DXL2/4/6* gene family results in underreporting of the total counts (91.3%) whereas when *3DXL6* is separate the UniMMap counts account for 99.9% of the number of reads simulated (figure 2-18A). Gene family read counts are accurate for the gene families when *3DXL6* is separated, the average difference between reads simulated and read count is 4.4%, the highest is 15% (*3DXL3/5/7*) and the lowest 0.01% (*3DXL6*). When *3DXL6* is included with *3DXL2/4*, the average difference is 650%, the highest is *3DXL3/5/7* (2631%) and lowest is *2DL1* (0.34%).

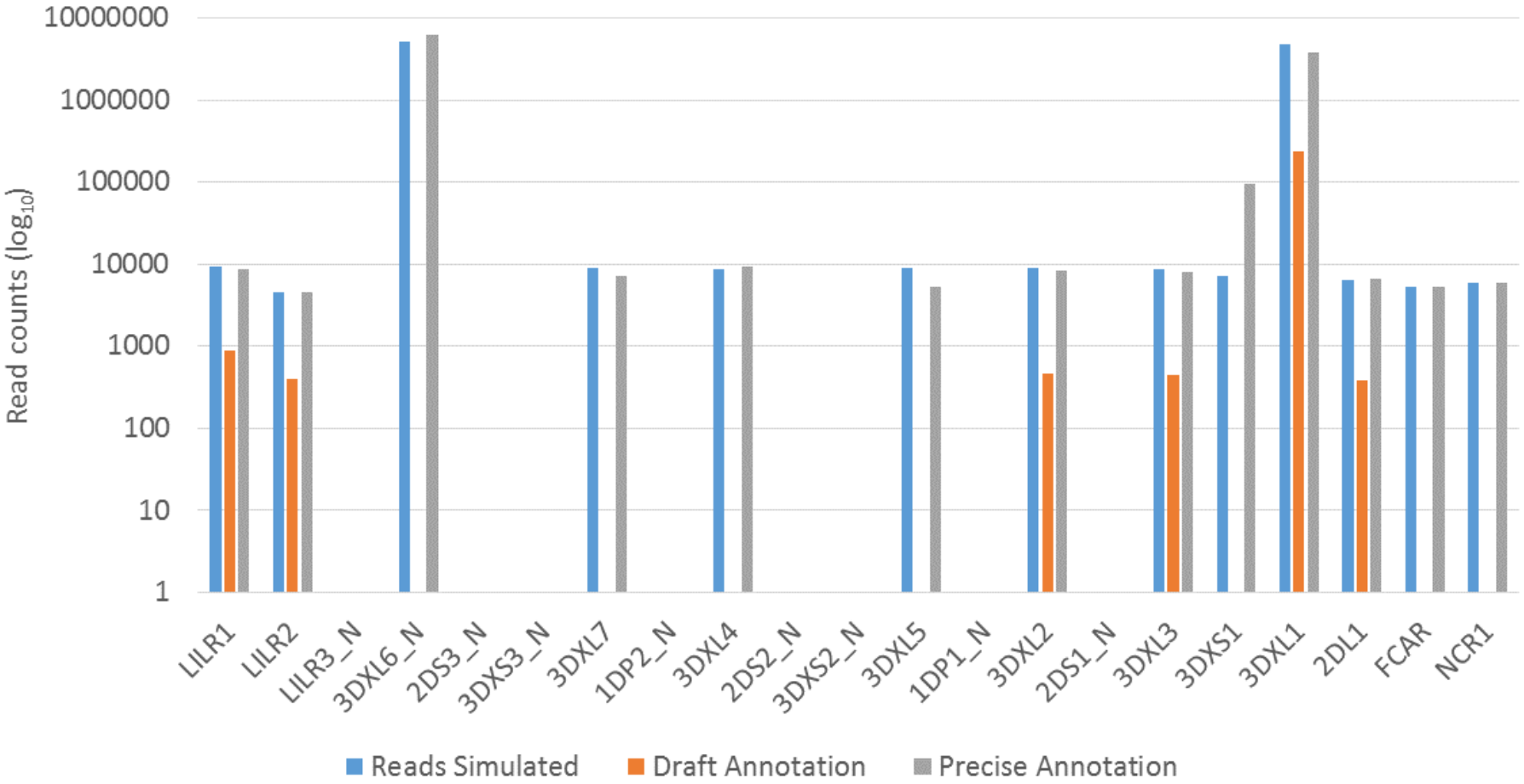


Figure 2-16. Number of reads simulated and UniMMap counts generated using either a draft or precise annotation for each gene of the LRC. Reads of mixed coverage were simulated and mapped using GEMTools to ARSv14Hap1. The number of reads simulated for each gene are shown as blue bars, UniMMap counts generated using the draft annotation as orange bars and UniMMap read counts generated from the precise annotation as grey bars.

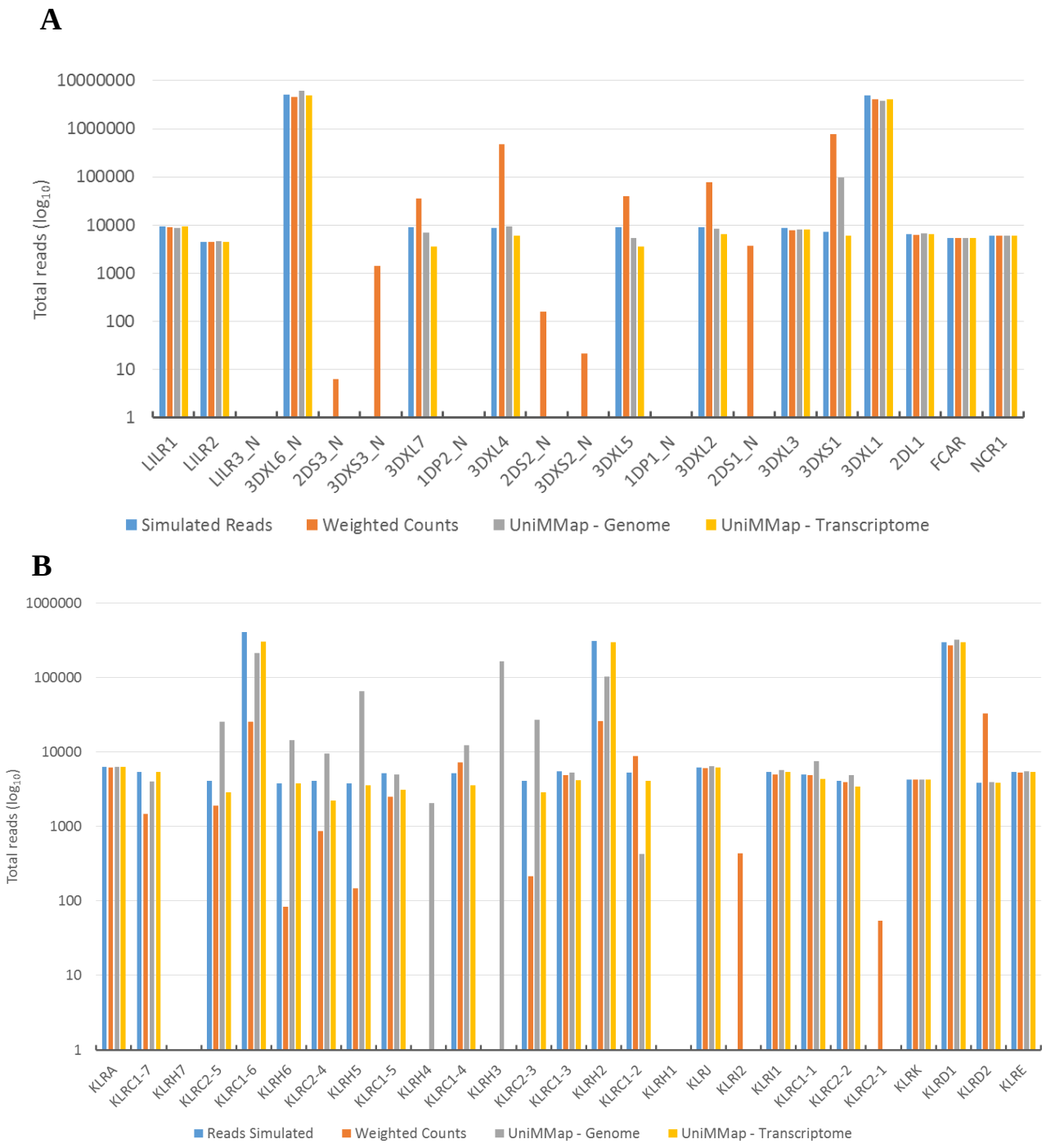


Figure 2-17. Number of reads simulated, standard weighted counts, UniMMap genome and UniMMap transcriptome counts for each gene of the LRC (A) and NKC (B). Reads of mixed coverage were simulated and mapped using GEMTools to ARSv14Hap1 or GEM-mapper to the UMD3.1 transcriptome.

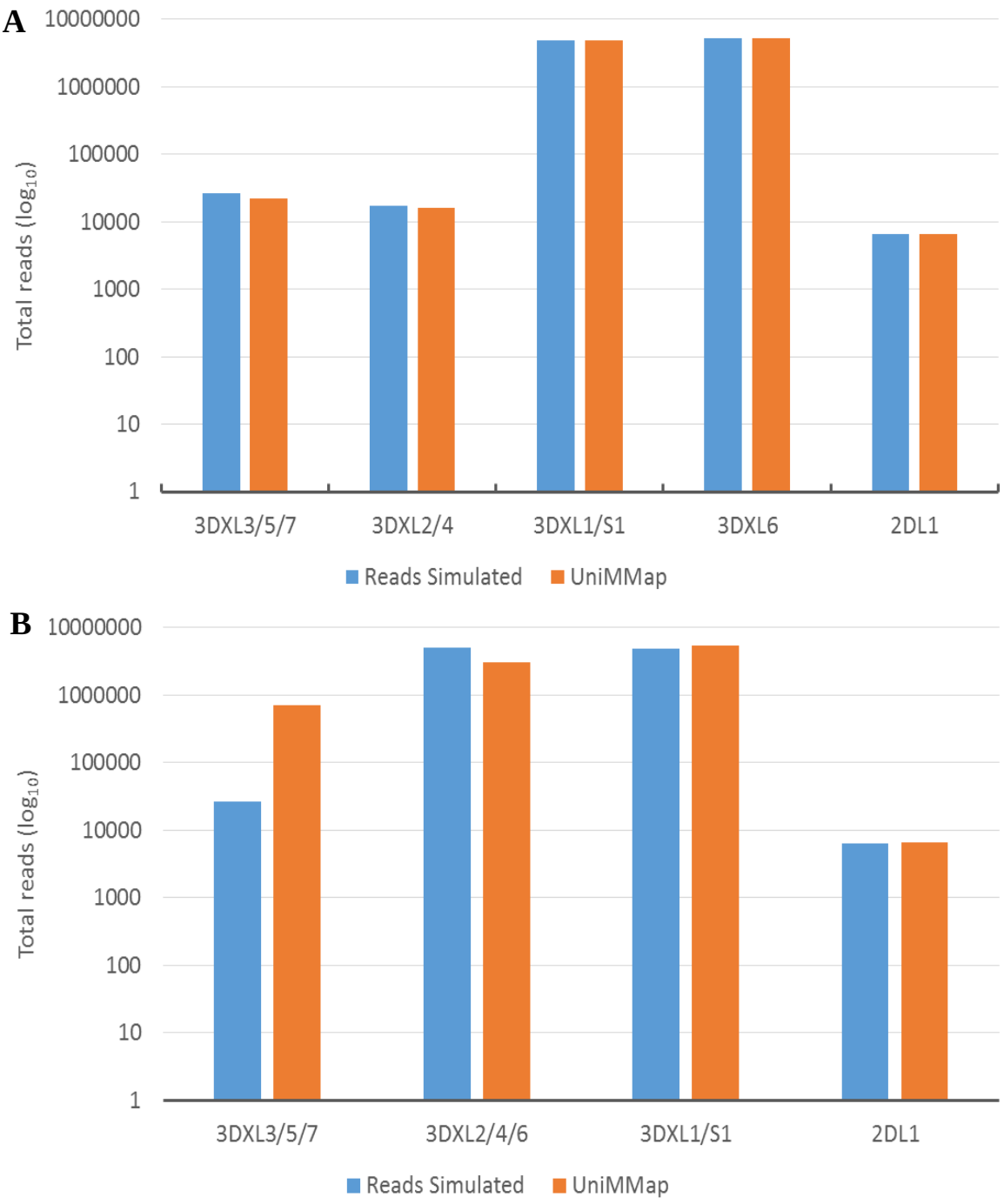


Figure 2-18. Number of reads simulated and UniMMap counts for gene families of the LRC. The mixed coverage simulated dataset was mapped to ARSv14Hap1_genefamilies (A) and ARSv14Hap1_genefamilies_3DXL6 (B) as described in the methods. The number of reads simulated for each gene family are shown as blue bars and the weighted counts generated by UniMMap as orange bars.

2.3.12 Characterisation of *KIR* transcription in humans with UniMMap

To confirm the effectiveness of UniMMap on real RNA-Seq data, as well as its adaptability to other species, it was used to generate read counts from publically available RNA-Seq datasets (Küçük et al. 2016). The datasets were produced from resting and 48 hour IL-2 stimulated NK cells from a single human donor. Read counts were generated only for the LRC genes as the human NKC has not expanded to the same extent as in cattle. Subsequently, it would not provide an accurate representation of mapping to the cattle NKC.

Of the 18 human *KIR*, we observed transcription of 9 (figure 2-19). As this data was generated by Küçük et al 2016, we compared the results of our analysis with theirs. However a complete comparison is not possible as they did not analyse the full repertoire of human *KIR* genes (14/18 *KIR*). The two analyses vary in the presence/absence of transcription of a number of genes; UniMMap does not report transcription of *2DL2*, *2DS1*, *2DS2*, *2DS5* and *3DS1*. Of the four genes not analysed by Küçük et al, we observed transcription of all (*3DL2*, *3DP1*, *3DL3* and *2DP1*). Two of the four are pseudogenes (*3DP1* and *2DP1*), transcription of *3DP1* is very low with an average of 0.09 reads per million reads mapped (RPM). For the genes contained in both sets of analyses, we observe similar results. Particularly with regards to the increase of *2DL4* and decrease of *2DS4* transcription in the cytokine stimulated population.

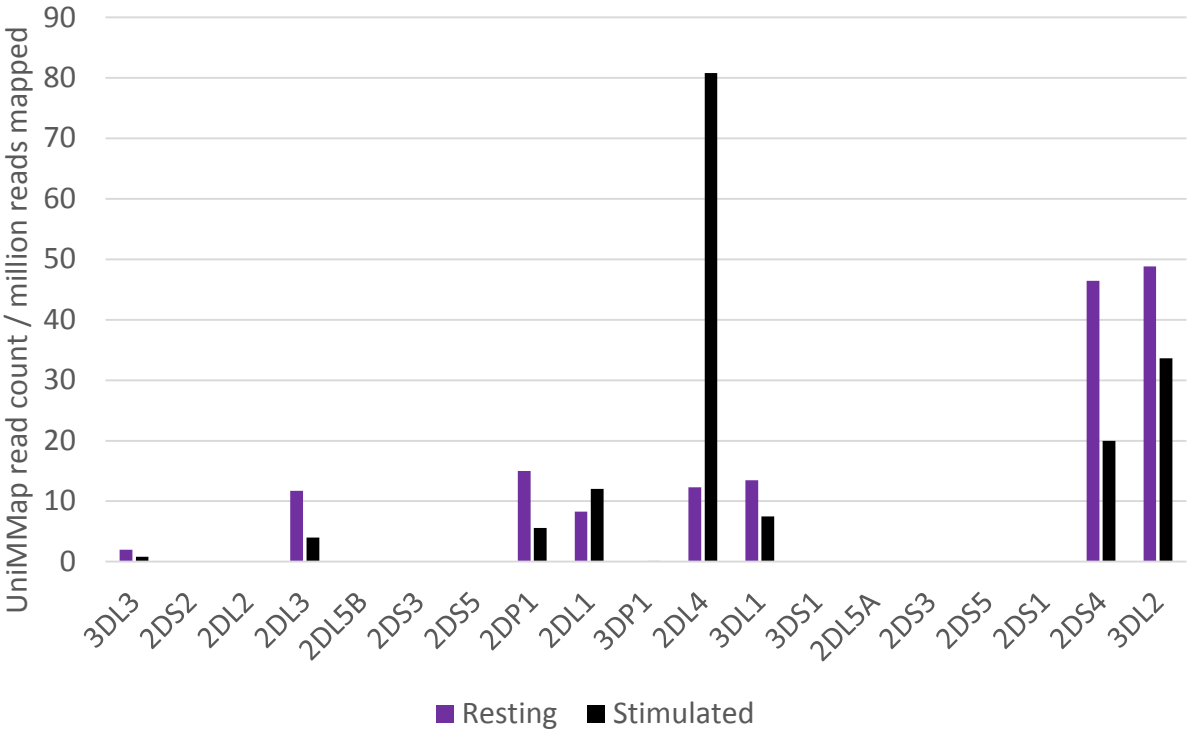


Figure 2-19. UniMMap results from two human NK RNA-Seq datasets. Reference sequences were generated by clustering all alleles present on the IPD-KIR database for each gene and generating a consensus sequence. RNA-Seq data from resting and 48-hour IL-2 activated NK cells was acquired from the SRA using accession number SRA200820. UniMMap was used to determine read counts from the RNA-Seq data. Read counts were normalised based on the number of reads mapped for each sample. Purple bars indicate the normalised read count for resting NK cells and black bars indicate 48-hour IL-2 activated NK cells.

2.4 Discussion

The handling of multi-mapping reads between regions of high sequence similarity is a common problem in RNA-Seq analysis. Multi-mapping reads can cause over-reporting of gene read counts as well as non-transcribed genes appearing to be transcribed. Due to the highly repetitive nature of the bovine LRC and NKC, generating accurate read counts to the genes of these complexes is not possible with current methods. The mapping of simulated 150bp RNA-Seq reads was used to compare and develop multiple methods of generating read counts to accurately measure gene transcription within the LRC and NKC. The ability to accurately determine transcriptional status (whether or not a gene was transcribed) was also assessed. Using these results, a custom pipeline named UniMMap was created that generates accurate read counts for the genes of the LRC and NKC, as well as increased accuracy in determining transcriptional status. Grouping the genes of the cattle LRC into pre-established gene families before applying the pipeline will enable greater accuracy when encountering haplotype variation as well as helping to account for the polymorphic nature of genes within the LRC. UniMMap will enable the analysis of existing and newly generated short-read RNA-Seq datasets at a level of accuracy not previously possible.

2.4.1 LRC and NKC sequences in Illumina RNA-Seq datasets cannot be accurately mapped using current methods

Mappability calculations for the LRC and NKC pinpoints the highly repetitive regions of these gene complexes and provides an explanation for the large scale multi-mapping observed. Longer reads solve most of the multi-mapping issues but currently no sequencing technology is available that provides both long reads and the high-read coverage needed to accurately measure global gene transcription. Also motivating the development of a new analysis pipeline, was the availability of a large number of publically available *Bos taurus* RNA-Seq datasets generated using Illumina short reads (≤150bp). Although none of these datasets originate from an enriched NK population, the retroactive analysis of these datasets with UniMMap will allow the creation of an atlas of NK receptor transcription in cattle (discussed in chapter 4). Use of these datasets is also important as part of the principles of the 3Rs (Replacement, Reduction and

Refinement) as it greatly reduces the need to generate a large amount of new datasets to analyse tissue-specific transcription of genes within the LRC and NKC.

2.4.2 Filtering strategies are vital to reduce erroneous alignments

Mismatches must be allowed when aligning RNA-Seq reads to a genome to account for sequencing error and differences between the genome of the individual/s used to produce the reference assembly and the individual the RNA-Seq data originated from. Allowing mismatches also has the undesired effect of increasing the number of alignments to the wrong genomic location, particularly when the sequencing reads originate from repetitive gene complexes. As these alignments are often (but not always) of lower quality, stringent filtering on alignment quality is vital to reducing the number of multi-map alignments. As multi-map alignments are not always of lesser quality, a second alignment filter is needed based on the underlying mappability. Filtering on mappability allows the isolation of reads mapping to a region known to be unique to an individual gene and consequently creates high-confidence in the accuracy of the alignment.

2.4.3 RNA splicing requires mappability to be calculated on the transcriptome rather than genome

Mappability of the genome at any given position does not always directly correlate to that of an RNA-Seq read alignment originating at the same position. Splicing means that an RNA-Seq read may not span the same region as the *k-mer* used to calculate mappability at that position. This issue was particularly noticeable in the NKC as the average exon length is shorter and so the likelihood of a read covering a splice event is increased. Calculating the mappability and subsequent mapping to the transcriptome eliminates this issue as the mappability is calculated on sequence that has already undergone splicing. Mapping to the transcriptome resulted in subtle changes to the final UniMMap read counts in the LRC with the exception of *3DXS1*, the accuracy of which was substantially improved. Improvements to read count accuracy are much more immediately noticeable in the NKC, in particular the highly repetitive region between *KLRC2-5* and *KLRH1*. In contrast to genome mapping, determination of transcriptional status was completely correct when mapping to the

transcriptome. Calculating mappability of the transcriptome instead of the genome has the added benefit of increasing the number and size of unique regions as the impact of non-transcribed sequence on mappability is removed. If reads map to the unique regions of a transcript we can be confident the gene is transcribed as reads originating from that region should map only to that region. Increasing the size and number of unique regions allows a greater number of reads to be used and increases overall accuracy, especially in low coverage datasets. The downside of using the transcriptome for mapping in general is that it relies on a very accurate reference sequence and annotation to generate the transcripts, however both of these elements are already required by UniMMap regardless of if the genome or the transcriptome are used for calculating mappability.

2.4.4 Gene families are a pre-emptive method of dealing with haplotype variation

As well as accurate quantification of individual genes, UniMMap is accurate when grouping the *KIR* genes into their established gene families. This can be critical due to the haplotypic variation of the LRC, the extent of which is currently unknown. The genes of the LRC are also highly polymorphic. This polymorphism can be observed when comparing haplotype 1 and haplotype 2, the two known haplotypes. Although not full length, haplotype 2 is identical in gene content whilst allelically completely different over the length of the overlapping region. The most striking example of polymorphism within the LRC is *KIR3DXL6*, for which both functional/non-functional and activating/inhibitory variants have been discovered. Gene families will allow the capture of reads from *KIR* that are divergent from those present in the known haplotypes in the event that the current method does not, although at the cost of individual gene resolution. Until new finalised haplotype assemblies are available, it is not possible to determine the necessity of the gene family methodology as haplotype 2 does not contain any genes not present in haplotype 1, on which the pipeline was developed.

2.4.5 UniMMap reports a valid *KIR* haplotype in human RNA-Seq data

Analysing real RNA-Seq datasets, allows us to determine the accuracy of UniMMap with the addition of various errors and biases present in RNA-Seq data. Using human data allows us to utilise the large amount of information regarding

haplotypes and individual gene transcription/expression to determine accuracy. The genes reported as transcribed by UniMMap are of a valid human *KIR* genotype (Hsu et al. 2002; Pyo et al. 2010), indicating the individual sequenced is homozygous for haplotype A. Transcription of *3DS1* was reported by Küçük et al (2016) but not by UniMMap. It is most likely a result of multi-mapping, *3DS1* is a B haplotype specific *KIR* and is always present with *2DL5A*, which was not observed. Interestingly, they did not include *3DL3* or *3DL2* in their analysis, two *KIR* that are present in all sequenced humans. UniMMap reports *3DL2* as the highest transcribed gene in the resting population and second highest in the stimulated. Potentially the *3DL2* reads were mapped to other non-transcribed *KIR* by Küçük et al (2016). Transcription of both pseudogenes (*3DP1* and *2DP1*) was reported by UniMMap. The mRNA of the pseudogenes *2DP1* and *3DP1*, is frequently observed in NK cells (Professor Paul Norman, University of Colorado Denver, personal communication). The results of the analysis provide high confidence in the ability of UniMMap to accurately resolve read counts for complex regions of a genome, particularly as it reports a valid genotype.

2.4.6 UniMMap conclusions

Current RNA-Seq methodologies are not sufficient for the analysis of transcription of the LRC and NKC. UniMMap is able to provide accurate read counts and more importantly, provides confidence in determining the transcriptional status of the genes of the LRC and NKC. We have prepared a modified pipeline that utilises UniMMap to provide accurate read counts for *KIR* gene families in the event that haplotype variation increases the complexity of mapping to individual genes. As the poor mappability and resulting multi-mapping is due to the gene duplication and conversion events that gave rise to the NKC and LRC, many other gene complexes that evolved through similar mechanisms are likely to share the difficulty of mapping short RNA-Seq reads. It is therefore possible to use UniMMap to accurately assess transcription of other gene complexes of *Bos taurus* as well as other species, providing the gene complex in question has been accurately assembled and annotated.

Chapter 3. Transcriptional analysis of cattle LRC and NKC genes

3.1 Introduction

Natural killer cells are tightly controlled by the balance of activating and inhibitory receptors on their cell surface. The most important receptors for control of the NK functional response are encoded by genes in either the LRC or the NKC. These receptors are highly polymorphic and the majority recognise MHC class I (Lanier and Phillips 1996). Genotypic, and allelic variation within these loci has been shown to contribute to differential outcomes for various viral infections, pregnancy and autoimmune diseases (Kim et al. 2008; Alter et al. 2007; Colucci 2017; Liang, Ma, and Tan 2017). The level of KIR expression on the cell surface can also influence disease outcome. KIR expression levels vary between alleles, some are not expressed on the cell surface but are detectable intracellularly (Pando et al. 2003). Individual NK cells also differ in their receptor repertoires; Yawata et al (2006) examined expression of six KIR in NK cells from 58 humans and observed 64 subsets expressing all possible receptor combinations in all individuals. This variegated expression between individual NK cells creates cell populations that are differentially responsive, which is more difficult for a pathogen to overcome.

In comparison to the more extensively studied human and mouse models, little is known about genotypic or allelic diversity in cattle. However, it has been shown that cattle have undergone an expansion of these loci to an extent not observed in other species. This significant expansion could potentially lead to even greater diversity in the NK receptor repertoire of individual NK cells and subsequently NK populations. We can predict which of the genes within these loci are likely to be transcribed based on their nucleotide sequence in the relatively few well-characterised haplotypes (Sanderson et al. 2014). However, it is possible functional alleles of predicted non-functional genes exist in uncharacterised haplotypes as some alleles are considered null based on very few mutations. Also unknown is how many of these receptors are specific to NK

cells, rather than more general leukocyte receptors, or are transcribed only in specialised NK populations such as liver NK cells.

As a first step in understanding the extent to which the genetic diversity in cattle results in a diverse NK cell population, Allan et al (2015) compared NK cell receptor transcription between MHC homozygous cattle of two contrasting genotypes. Although they were able to examine only a subset of the total receptors, they used PCR to show variable transcription of five receptor genes including *3DXL6,* independent of MHC class I genotype. Additionally they investigated transcription in individual NK cells and found four *KLR* genes were consistently observed in the population data but varied at the individual animal level. They also observed variable transcription of *3DXL1, 3DXL6* as well as the *3DXL3/5/7 KIR* group between the individual NK cells.

To further examine this differential transcription, UniMMap was used in combination with four high coverage RNA-Seq datasets from two animals to compare the transcription profiles of the LRC and NKC between peripheral blood mononuclear cells (PBMCs) and NK cells. Whole transcriptome sequencing enabled the first analysis of transcription of the entire LRC and NKC and the investigation of transcription of these genes in the context of global transcription.

3.2 Methods

3.2.1 Preparation of samples for RNA-Seq

All animal experiments were approved by The Pirbright Institute Ethics Committee and carried out in accordance with the U.K. Animal (Scientific Procedures) Act 1986. This work was carried out under home office license number P7A651E99 'Supply of ruminant and porcine blood for virus infection research and diagnostic'. Two overtly healthy animals were selected from the Pirbright MHC homozygous herd and 200ml of blood taken from each animal on two separate days. PBMCs were isolated from blood by underlaying histopaque-1083. PBMCs were mixed with GR13.1, an antibody for NCR1, and then incubated with MACS IgG beads. NCR1+ cells were then isolated using an autoMACS Pro Separator running the positive selection protocol. The population purity was then assessed using a MACSQuant Analyzer 10 (supplementary figure 3-1). The isolated cells were resuspended in Trizol and stored at -80°C. Total RNA was extracted from the cells using an Invitrogen PureLink RNA Mini Kit and quantified using an Agilent TapeStation. Samples with an RNA intergrity number (RIN) of 8.0 or above were taken forward. RNA from NK cells originating from the same animal but isolated on different days was pooled to generate enough material for sequencing as well as to minimise day-to-day variation.

3.2.2 RNA-Seq library preparation and sequencing

Total RNA was sent to the University of Liverpool Centre for Genomic Research for sequencing. Dual-indexed, strand-specific libraries were generated using NEBNext polyA selection and the Ultra Directional RNA library preparation kit. Libraries were sequenced on a lane of an Illumina HiSeq 4000 generating paired-end, 150bp from >280M clusters.

3.2.3 UniMMap analysis of samples

The UniMMap pipeline is described fully in chapter 2. Briefly, GEM-Mappability was run on the custom UMD3.1 transcriptome with a k-mer length of 150bp and the approximation threshold disabled. RNA-Seq reads from each sample were mapped to the UMD3.1 custom transcriptome using GEM-Mapper. The resulting alignments were intersected with the mappability of the transcriptome and the average mappability of each read determined. Read counts of each gene were weighted on the average mappability of the reads aligning to them. The read

counts for genes that have no alignments to a region of mappability=1 were set to 0. Read counts were normalised based on the number of reads mapped in each sample to produce a reads per million reads mapped (RPM) value for each gene. Read counts for this chapter are available at: https://github.com/richard-borne/PhD_thesis_data

3.2.4 RNA-Seq aligner comparison

Salmon (Patro et al. 2017) was run using the UMD3.1 transcriptome, as used with UniMMap. The custom ARSv14hap1 genome created and used previously (chapter 2) was the reference for both STAR (Dobin et al. 2013) and Tophat2 (D. Kim et al. 2013). Salmon was run in quant mode with default parameters except for defining library type and read counts output automatically. STAR was also ran using default parameters and read counts obtained by setting the quantMode parameter to GeneCounts. Tophat2 was again run using default parameters except to define library type, read counts were obtained by running featureCounts (Liao, Smyth, and Shi 2014) from the Subread package on the output alignment file. Genes were ranked in order of highest to lowest read count for each of the analysis methods.

3.2.5 Whole transcriptome analysis

RNA-Seq reads from each sample were mapped to the UMD3.1 genome in combination with the UMD3.1.92 annotation using the GEMTools-rna pipeline with default settings. Weighted read counts were generated by the pipeline for each of the genes in the annotation. Gene names were obtained for genes of interest by converting the Gene IDs using Ensembl BioMart (Kinsella et al. 2011). Read counts were normalised to reads per million reads mapped (RPM) for comparison between samples. Gene IDs of the genes for which read counts were at least one fold higher in the NK dataset were extracted using awk, imported into Reactome (Fabregat et al. 2018) and mapped to known *Bos taurus* identifiers.

3.3 Results

3.3.1 UniMMap read counts for a majority of NKC and LRC genes are higher after positive selection for NCR1+ cells

NCR1 expression is currently the best marker of cattle NK cells, and facilitates cell enrichment using positive selection with NCR1 specific antibodies. To understand whether LRC/NKC gene transcription is specific to NK cells, PBMC and subsequently NK cells were isolated as described in the methods of this chapter from two animals belonging to an MHC homozygous cattle herd. The population of cells isolated from PBMCs was found to be consistently between 60-65% $NCR1^{+}$.

Transcription of *NCR1* was assessed to confirm selection of an NCR1+ population (supplementary figure 3-2) and a 16.5x increase was observed in the NK population compared to PBMCs. For the LRC genes, transcription of genes considered to encode null alleles based on the annotation of haplotype 1 (Sanderson et al. 2014) was detected in both the PBMC and NK populations, accounting for 53.86% of the observed transcription in the average PBMC population and 58.85% in the average NK cell population (figure 3-1). Read counts across the LRC are higher in all cases for the NK population compared to PBMCs. The average read count of the two PBMC populations across the LRC is 5.43 RPM and increases to 23.39 RPM in the NK population. The smallest difference in transcription between PBMCs and NK cells is for *3DXL7* where the normalised read count is 1.80x higher in the NK population. The largest difference is observed for *2DL1* which has a read count 12.85x higher in the NK population. The average difference between the two populations is 5.15x.

NKC gene transcription between the two cell populations illustrated that like the LRC, transcription of genes previously found to encode null alleles was observed in the averaged PBMC population (figure 3-2). The gene with the largest increase in the average NK population was *KLRC2-2*, which was 16.80x higher. The smallest difference for a gene with a higher read count in the NK population was *KLRC2-5* (1.31x) and the average difference was 2.57x. *KLRH6* and *KLRC2-3* were only detected in the average NK population and *KLRH3* was only detected in the

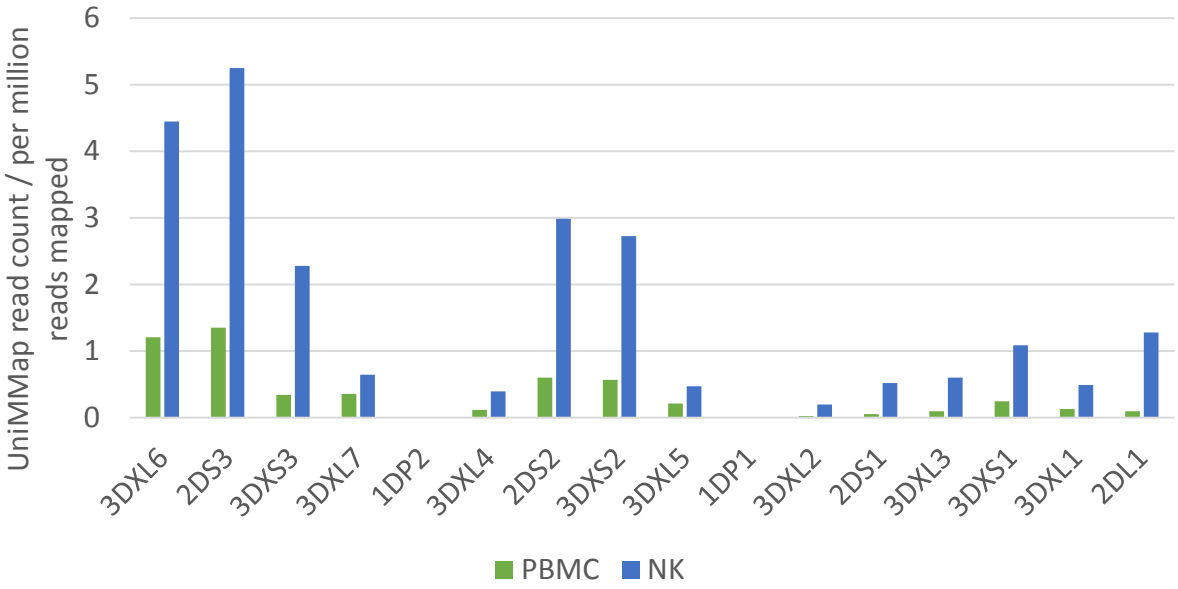


Figure 3-1. Normalised average UniMMap read counts for genes of the LRC from PBMC and NK RNA-Seq. UniMMap was run on two cattle PBMC and two cattle NK 150bp, paired-end RNA-Seq datasets and the resulting read counts normalised to RPM. Read counts across each cell type were averaged. Green bars indicate the average read count from PBMCs and the blue bars the average from NK cells.

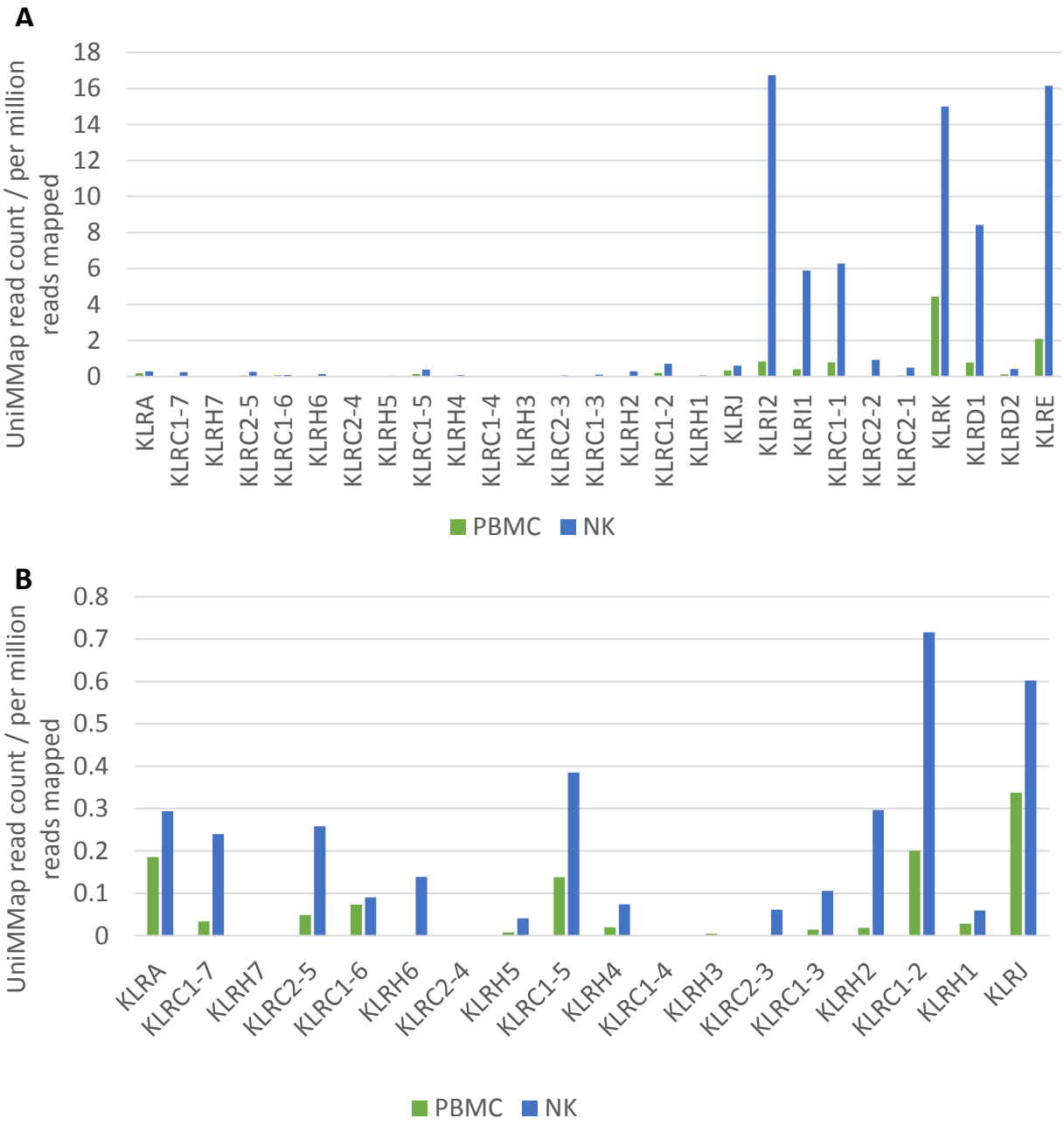


Figure 3-2. Normalised average UniMMap read counts for genes of the entire NKC (A) and the region between *KLRA* and *KLRJ* (B) from PBMC and NK RNA-Seq. UniMMap was run on two cattle PBMC and two cattle NK 150bp, paired-end RNA-Seq datasets and the resulting read counts normalised to RPM. Read counts across each cell type were averaged. Green bars indicate the average read count from PBMCs and the blue bars the average from NK cells.

average PBMC population. Read counts were very low for many of the genes of the NKC, 17 out of the 24 detected genes have a read count lower than 1 RPM in either of the populations. It is therefore possible that the genes that were not observed were transcribed at a level we were unable to detect.

3.3.2 There is large variation in LRC/NKC gene read counts between the two NK populations but not in transcriptional status

The two NK populations were compared to determine the level of variation in read counts for each of the genes of the LRC and NKC between the two animals as to determine if any genes were differentially transcribed. To enable comparison between the animals 1020 and 1021, the read counts generated by UniMMap were normalised to RPM based on the number of reads mapped for each animal.

In the LRC, only *2DS1, 3DXL7, 3DXL3* and *2DL1* are transcribed more highly in animal 1021 than 1020 (figure 3-3). The read counts of *3DXL5* (0.47 RPM and 0.48 RPM) and *3DXL2* (0.20 RPM in both) are almost identical in the two animals. The normalised read count to the entire LRC is 30.01 RPM for animal 1020 compared to 16.77 RPM in animal 1021. The largest difference occurs when comparing *3DXL3* between animals where there are 9.72x more reads per million in animal 1021. Transcription of null alleles accounts for 64.79 % of the total transcription of the LRC in 1020 and 48.23 % in animal 1021.

The difference in total reads per million mapped between the two animals is much greater for the NKC than the LRC, transcription in animal 1020 (100.30 RPM) is double that of animal 1021 (47.8 RPM) (figure 3-4). Of the additional 52.5 RPM mapped to the NKC in animal 1020, 37.8 RPM are from just three genes (*KLRI2, KLRK* and *KLRE*). In animal 1020, these three genes account for 67.13% of the total transcription observed for the 27 genes of the NKC. The same three genes represent 58.14% of the total NKC transcription observed for animal 1021. The majority of transcripts identified from the NKC were from genes located between *KLRI2* and *KLRE*, 96.13% in animal 1020 and 91.57% in 1021.

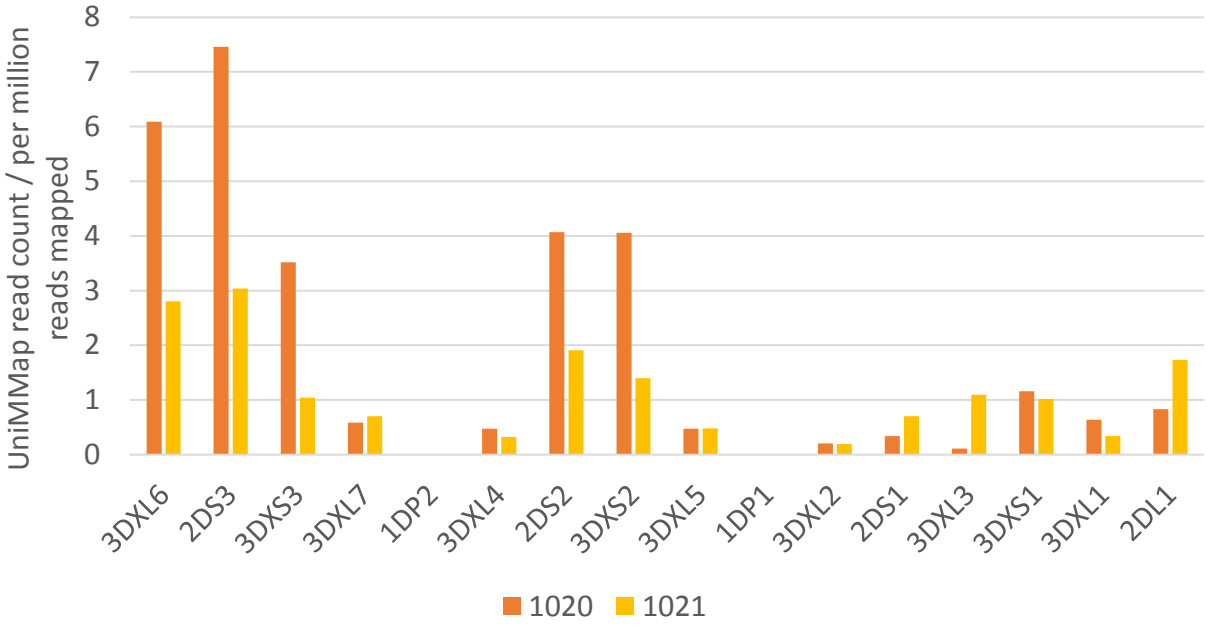


Figure 3-3. Average UniMMap read counts for genes of the LRC from NK cells isolated from two animals of different MHC class I haplotypes. UniMMap was run on two cattle NK 150bp, paired-end RNA-Seq datasets and read counts across each cell type averaged. Read counts were normalised to RPM. Orange bars indicate the UniMMap RPM from animal 1020 and the yellow bars from animal 1021.

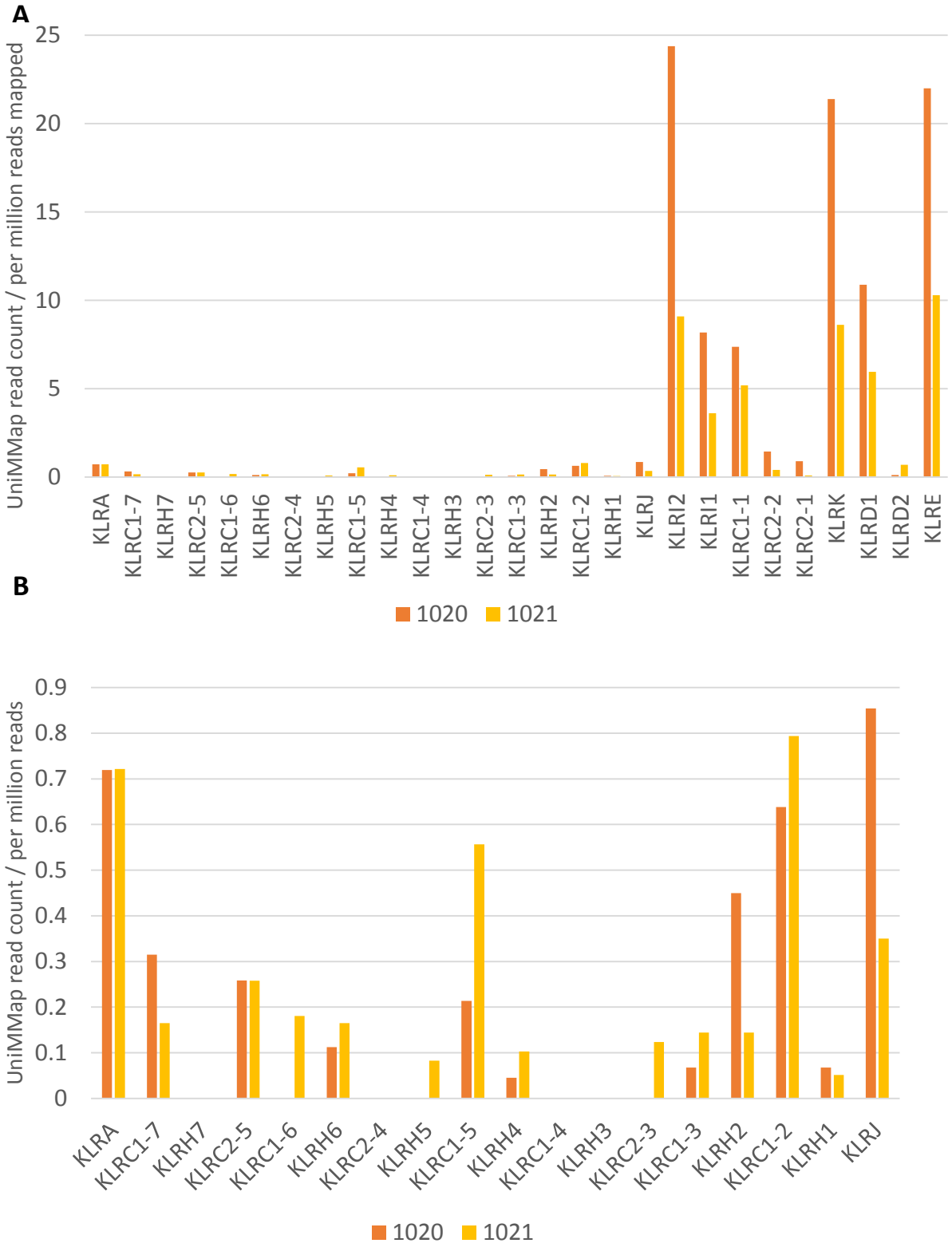


Figure 3-4. Average UniMMap read counts for genes of the NKC (A) and the region of the NKC between *KLRA* and *KLRJ* (B) from NK cells isolated from two animals of different MHC class I haplotypes. UniMMap was run on two cattle NK 150bp, paired-end RNA-Seq datasets and read counts across each cell type averaged. Read counts were normalised to RPM. Orange bars indicate the UniMMap RPM from animal 1020 and the yellow bars from animal 1021.

In the region of much lower transcription, between *KLRA* and *KLRJ*, 10 of the 18 genes are more highly transcribed in animal 1021. The only gene in the NKC in which transcription does not differ between the samples is *KLRC2-5*, although the transcription of *KLRA* is only 1.03x higher in 1021 than 1020. Comparing transcriptional status of the genes of the NKC between the samples shows that *KLRC2-3* and *KLRH5* are observed in only animal 1021. However, the pre-normalised read count for *KLRC2-3* in animal 1021 was just six and so it is difficult to say with certainty if it is not transcribed in animal 1020 or if it is not detected due to low coverage.

We demonstrate a clear transcriptional bias to one end of the NKC, with the majority of the reads originating from just three genes in both of the animals. Although double the reads mapped to the NKC in one animal compared to the other, there is little difference in read counts for most of the NKC genes, 71% of the extra reads originate from the three most transcriptionally active genes.

3.3.3 There is a large variation in read counts generated to the LRC/NKC genes between analysis methods

As discussed in the previous chapter, standard analysis methods are unable to provide accurate read counts for transcripts originating from many genes located within the LRC and NKC. UniMMap was developed using simulated data and has been shown to provide accurate read counts even when the number of reads originating from different genes vary by an order of magnitude. To gauge the impact UniMMap has on read counts from RNA-Seq datasets, the read counts obtained from UniMMap were compared to three commonly used RNA-Seq analysis tools.

Read counts were generated by each RNA-Seq analysis tool for both of the NK populations (animals 1020 and 1021) and compared for the LRC (figure 3-5). For animal 1020, Salmon and Tophat2 produce similar read counts for many of the *KIR*. When generating read counts from animal 1021 however, Salmon often produced read counts much higher than Tophat2. Tophat2 produced the highest total read count to the LRC for animal 1020 (2884 total reads), Salmon produced 2602.74 and STAR and UniMMap produced 1516.49 and 1335.17 respectively. UniMMap reported the lowest total read count for the LRC for animal 1021

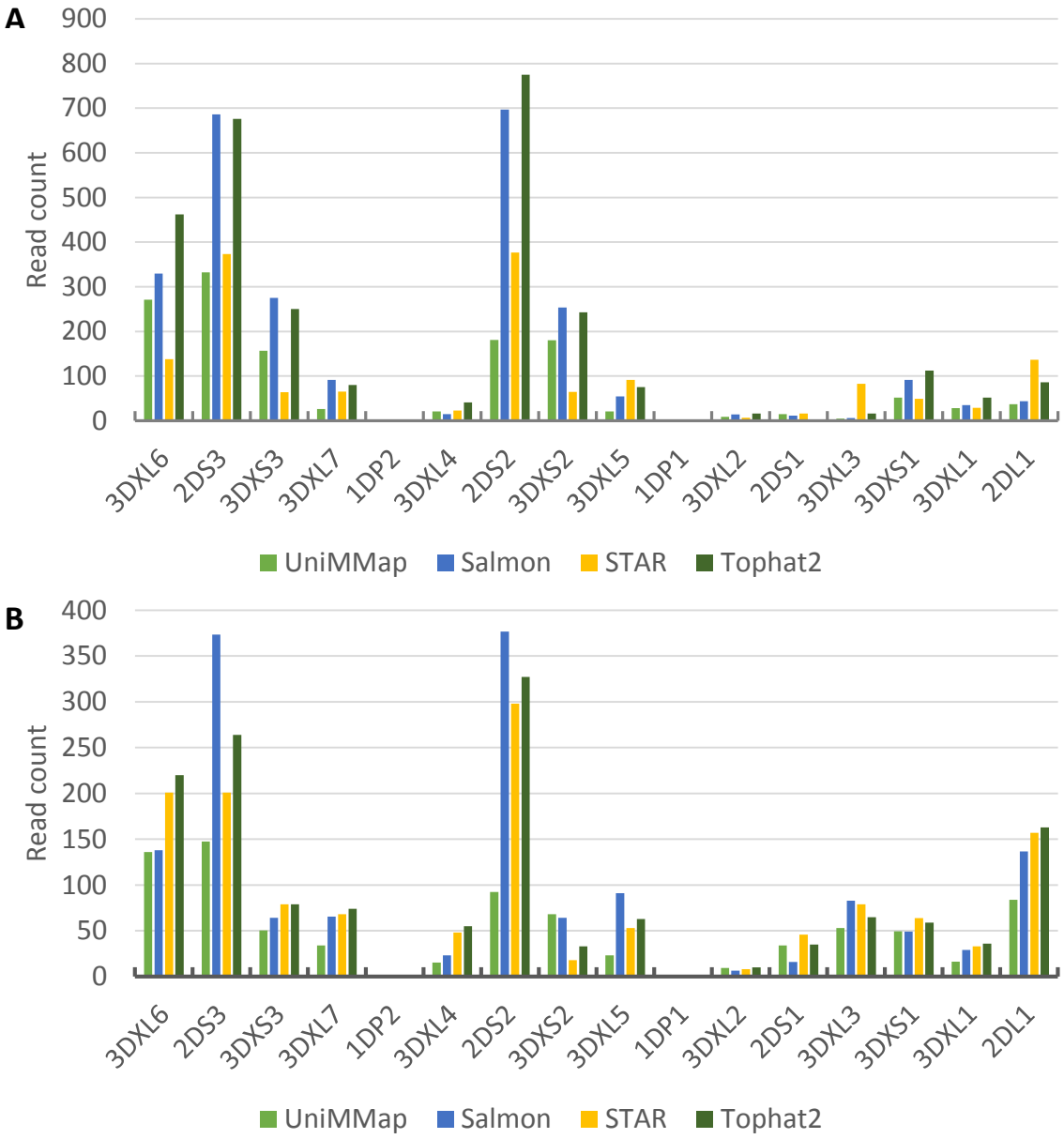


Figure 3-5. LRC gene read counts from cattle NK RNA-Seq from animal 1020 (A) and 1021 (B) generated by four RNA-Seq analysis methods. The same sample was analysed using UniMMap (light green bars), Salmon (blue bars), STAR (yellow bars) and Tophat2 (dark green bars). Salmon, STAR and Tophat2 were all run using default parameters.

A

Animal 1020	UniMMap	Salmon	STAR	Tophat2	Largest Difference
3DXL6	2	3	3	3	1
2DS3	1	2	2	2	1
3DXS3	7	9	6	5	4
3DXL7	9	7	7	6	3
1DP2	15	15	15	15	0
3DXL4	13	12	10	10	3
2DS2	3	1	1	1	2
3DXS2	5	8	13	13	8
3DXL5	11	5	9	8	6
1DP1	16	16	16	16	0
3DXL2	14	14	14	14	0
2DS1	10	13	11	12	3
3DXL3	6	6	5	7	2
3DXS1	8	10	8	9	2
3DXL1	12	11	12	11	1
2DL1	4	4	4	4	0

B

Animal 1021	UniMMap	Salmon	STAR	Tophat2	Largest Difference
3DXL6	2	3	3	3	1
2DS3	1	2	2	2	1
3DXS3	5	4	9	4	5
3DXL7	9	6	7	8	3
1DP2	15	15	15	15	0
3DXL4	10	11	12	11	1
2DS2	3	1	1	1	2
3DXS2	4	5	8	5	4
3DXL5	11	8	5	9	6
1DP1	16	16	16	16	0
3DXL2	13	12	14	13	2
2DS1	12	13	13	14	2
3DXL3	14	14	6	12	6
3DXS1	6	7	10	6	4
3DXL1	8	10	11	10	3
2DL1	7	9	4	7	5

Figure 3-6. Comparison of transcriptional rank of the genes of the LRC generated from multiple analysis methods from animal 1020 (A) and 1021 (B). The 150bp paired-end RNA-Seq dataset of an NK population (animal 1020) was used to compare transcriptional rank of the genes of the LRC from read counts generated from UniMMap as well as three commonly used RNA-Seq analysis tools. Colours indicate a heat map of green (highest) to red (lowest) based on the ranking.

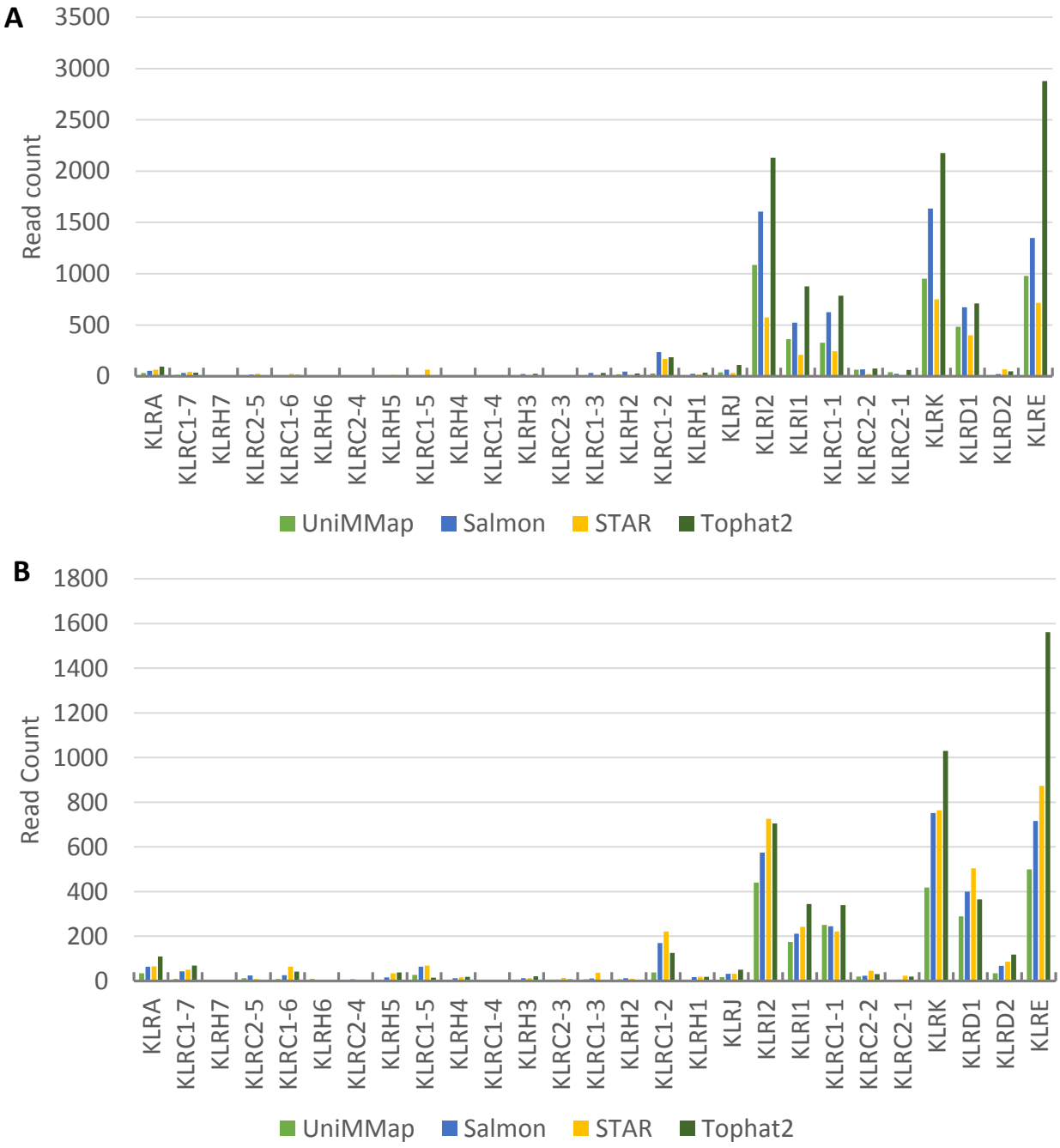


Figure 3-7. NKC gene read counts from cattle NK RNA-Seq from animal 1020 (A) and 1021 (B) generated by four RNA-Seq analysis methods. The same sample was analysed using UniMMap (light green bars), Salmon (blue bars), STAR (yellow bars) and Tophat2 (dark green bars). Salmon, STAR and Tophat2 were all run using default parameters.

A

Animal 1020	UniMMap	Salmon	STAR	Tophat2	Largest Difference
KLRA	10	10	10	9	1
KLRC1-7	13	12	11	14	3
KLRH7	21	26	27	27	6
KLRC2-5	14	18	14	22	8
KLRC1-6	22	23	13	18	10
KLRH6	16	25	26	26	10
KLRC2-4	23	27	22	25	5
KLRH5	24	24	17	24	7
KLRC1-5	15	20	9	23	14
KLRH4	20	22	20	21	2
KLRC1-4	25	21	25	19	6
KLRH3	26	17	19	17	9
KLRC2-3	27	19	23	20	8
KLRC1-3	18	13	21	15	8
KLRH2	12	11	18	16	7
KLRC1-2	11	7	7	7	4
KLRH1	19	15	16	13	6
KLRJ	9	9	12	8	4
KLRI2	1	2	3	3	2
KLRI1	5	6	6	4	2
KLRC1-1	6	5	5	5	1
KLRC2-2	7	8	15	10	8
KLRC2-1	8	14	24	11	16
KLRK	3	1	1	2	2
KLRD1	4	4	4	6	2
KLRD2	17	16	8	12	9
KLRE	2	3	2	1	2

B

Animal 1021	UniMMap	Salmon	STAR	Tophat2	Largest Difference
KLRA	8	10	10	9	2
KLRC1-7	15	11	12	10	5
KLRH7	24	26	27	27	3
KLRC2-5	13	14	23	21	10
KLRC1-6	14	13	11	12	3
KLRH6	16	25	26	26	10
KLRC2-4	25	22	25	25	3
KLRH5	21	17	15	13	8
KLRC1-5	10	9	9	19	10
KLRH4	20	20	19	18	2
KLRC1-4	26	27	24	23	4
KLRH3	27	19	21	15	12
KLRC2-3	19	23	20	20	4
KLRC1-3	17	21	14	22	8
KLRH2	18	18	22	24	6
KLRC1-2	7	7	6	7	1
KLRH1	23	16	18	17	7
KLRJ	12	12	16	11	5
KLRI2	2	3	3	3	1
KLRI1	6	6	5	5	1
KLRC1-1	5	5	7	6	2
KLRC2-2	11	15	13	14	4
KLRC2-1	22	24	17	16	8
KLRK	3	1	2	2	2
KLRD1	4	4	4	4	0
KLRD2	9	8	8	8	1
KLRE	1	2	1	1	1

Figure 3-8. Comparison of transcriptional rank of the genes of the LRC generated from multiple analysis methods from animal 1020 (A) and 1021 (B). The 150bp paired-end RNA-Seq dataset of an NK population (animal 1020) was used to compare transcriptional rank of the genes of the LRC from read counts generated from UniMMap as well as three commonly used RNA-Seq analysis tools. Colours indicate a heat map of green (highest) to red (lowest) based on the ranking.

(813.83) compared to Salmon (1516.50), STAR (1355) and Tophat2 (1483). Null alleles account for 64.79% of the total counts generated by UniMMap compared to 73.87% by Salmon, 58.98% by STAR and 67.41% by Tophat2. Ranking the order of the genes based on read count for each of the analysis tools (figure 3-6), shows which genes are the most variable between the analysis tools. For animal 1020, the only genes that have the same ranking in all methods are *1DP1* and *1DP2*, which have read counts of 0 in all cases. Animal 1021 has two additional genes that have identical rankings, *3DXL2* is ranked 14th and *2DL1* is 4th for each analysis method. For animal 1020, *3DXL6*, *2DS3* and *3DXL1* and for animal 1021, *3DXL6, 2DS3* and *3DXL4* had a maximum difference of one in the rankings. Most difficult to rank for animal 1020 was *3DXL5*, which differed six places in the rankings between UniMMap and STAR, and *3DXL3*, which differed six places between STAR and UniMMap/Salmon. In animal 1021, *3DXS2* differed eight places between UniMMap and STAR/Tophat2. As with animal 1020, *3DXL5* differed by six places for animal 1021, in this case however the largest difference was observed between UniMMap and Salmon. UniMMap agreed with the ranking of another analysis method 13 times out of the 32 total rankings of each sample combined.

Comparing the read counts generated by the various analysis tools to the NKC genes reveals just how variable the reported read counts can be (figure 3-6). The total read count for the NKC in animal 1020 generated by UniMMap was 4467.90, Salmon reported 7114.78 total reads, STAR reported 3513.31 reads and Tophat2 reported 10,348 reads, a read count more than double two of the other tools examined. When analysing animal 1021, Tophat2 produces a total read count of 5041 to the NKC, again higher than UniMMap (2317.27), Salmon (3513.31) and STAR (4143). The percentage of read counts to null alleles in the NKC is variable in animal 1020, the highest percentage is UniMMap (25.29%) and lowest is STAR (17.72%). Less variability is seen in the percentage of reads counts to null alleles in animal 1021, the highest is again UniMMap (19.53%) and lowest is Tophat2 (15.55%). As with the LRC, the transcription rankings of the NKC generated by each analysis tool were calculated (figure 3-8). The larger number of genes in the NKC allows for greater variation in the rankings and this can be observed with *KLRC2-1* which is ranked 16 positions higher by UniMMap than STAR in animal 1020 and eight places higher in animal 1021 by Tophat2

than Salmon. The positional ranking for *KLRC1-5* is also inconsistent in both samples, a difference of 14 between STAR and Tophat 2 in animal 1020 and a difference of 10 between Salmon/STAR and Tophat2 in animal 1021. The only gene ranked identically by all aligners was *KLRD1* in animal 1021 which was ranked as the 4th highest transcribed gene of the NKC. In animal 1020 and 1021, eight and 10 of the samples had a maximum difference of two or less respectively. We show that there is large variation in both the total and relative read counts produced by each tested analysis method. The variation in relative read counts leads to inconsistent rankings of gene transcription between the methods. UniMMap also provides confidence that that the haplotype 1 null-alleles are transcribed, rather than a result of multi-mapping.

3.3.4 There is little variation in the top 100 genes transcribed in each cell population from either animal

To understand which genes are the most transcriptionally active in both PBMCs and the NK cell subset, the top 100 transcribed genes from each sample were compared. Despite polyA selection during library preparation, on average 18 of the top 100 genes for each sample were ribosomal related genes, these genes were removed from subsequent analyses. The top 100 non-ribosomal related genes were identified for each cell population and combined, producing 138 unique genes IDs. The read counts for these 138 genes were then extracted for each population of cells, normalised to reads per million and transformed to their square root values to reduce the size of the range of values (figure 3-9). Out of the 138 genes compared, 35 were not identified by Ensembl BioMart and were labelled as 'Uncharacterised'. Further investigation into Uncharacterised_19, of which transcription was much higher in PBMCs and NK cells from animal 1021 than 1020, reveals it is major histocompatibility complex class II, DQ beta (BOLA-DQB). Uncharacterised_38 and Uncharacterised_3 have higher transcription in the two NK samples and are located on chromosome 21 and 8 in UMD3.1 respectively, and each produced a transcript coding for multiple immunoglobulin C1-set domains. As the majority of the top 100 genes from each sample are shared between all of the samples, global transcription was next analysed to identify transcription patterns unique to either population.

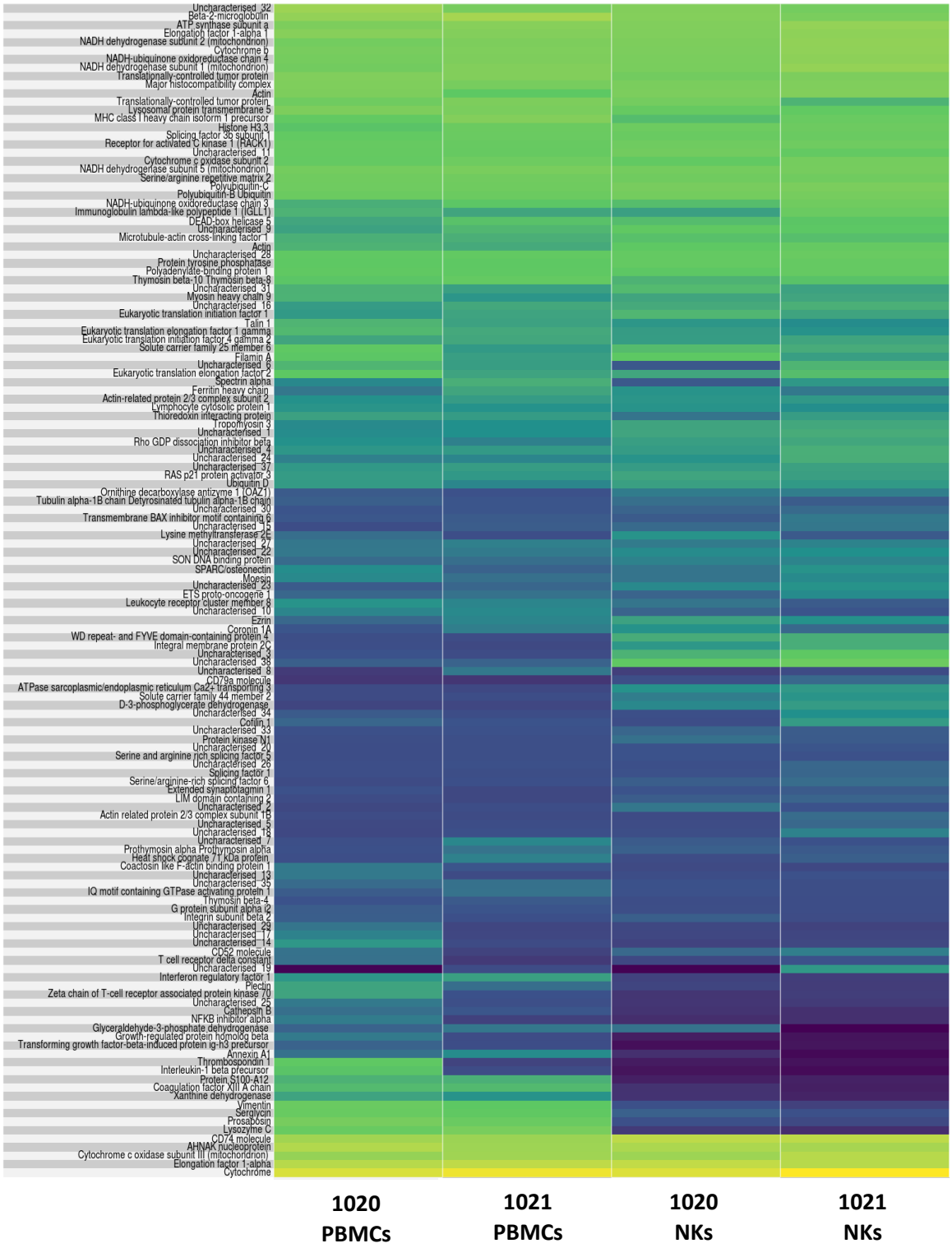


Figure 3-9. Comparison of the top 100 transcriptionally active genes in each sample. The 100 non-ribosomal related genes with the highest read count generated by GEMTools were identified for each sample and collated to produce a list of 138 unique genes IDs. Gene IDs were imported into Biomart (Kinsella et al. 2011) to produce gene names and descriptions for each ID. Read counts for each of the 138 genes were extracted for each sample and normalised to RPM and transformed to their square root values to reduce the scale of the range of read counts. A heatmap was generated using the superheat package in R. Hierarchical clustering was applied during heatmap generation to order the rows.

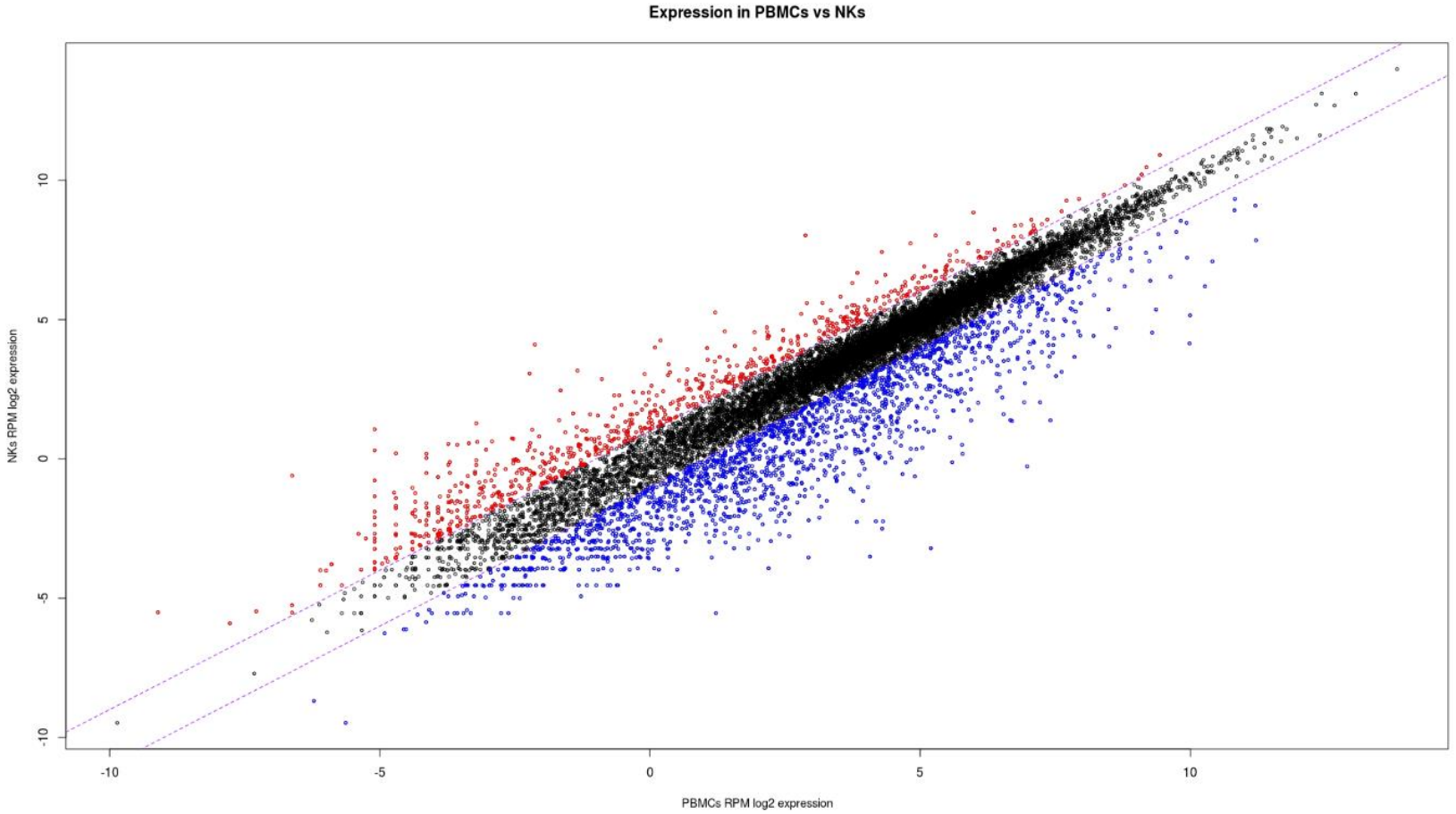


Figure 3-10. Comparison of global transcription between PBMCs and NKs. Read counts generated by GEMTools were normalised to reads per million (RPM) and averaged for each cell type. Normalised read counts were transformed to log2 values and plotted using the plot function in R. Red dots indicate genes for which transcription is at least 1 fold higher in the NK population, blue dots indicate the same for the PBMC population and black dots indicate the remainder of the genes.

3.3.5 Whole transcriptome variation exists between cattle PBMC and NK populations

As general gene transcription in cattle NK cells is poorly understood, we compared global transcription between our PBMC and NK cell populations to identify genes transcribed more highly in the NK cell population.

After normalising the read counts for each sample to reads per million and averaging the normalised read counts of both animals for each cell population, they were converted to $\log_2$ values to determine the fold change between the populations. Of the 13,533 genes detected in both populations, 1872 (13.83%) were at least one fold higher in the PBMC population than the NK population (figure 3-10). Comparatively, 750 (5.54%) of the total genes were transcribed at a level at least one fold higher in the NK population than the PBMC population. A search of genes transcribed only in the NK population reveals 417 genes. Of these 417 genes, 98.08% have a read count of 20 or lower and are potentially NK specific transcripts too lowly transcribed in the PBMC population to be detected. The top five transcribed genes only detected in the NK population are a gastric inhibitory polypeptide receptor (GIPR-201), interleukin 34 (IL-34), interleukin 7 (IL-7), connector enhancer of kinase suppressor of ras 1 (CNKSR1) and an unknown pseudogene located on chromosome 26.

As individual analysis of the 750 genes transcribed at least one fold higher in the NK population is not practical, pathway analysis was carried out using Reactome (Fabregat et al. 2018) to determine which pathways the genes belonged to (figure 3-11). Reactome was able to find matches for 267/750 identifiers imported and from those, identified 77401 interactions. Signal Transduction, Immune System and Metabolism were the largest pathways, encompassing 64.2% of the interactions. Innate Immune System, Adaptive Immune System and Cytokine Signaling in Immune system pathways were ranked 8th, 15th and 16th respectively. The highest ranked pathway within Signal Transduction was Signaling by GPCR and the highest within metabolism was Metabolism of lipids.

3.3.6 Genes of the LRC and NKC do not rank highly in global transcription in resting PBMCs or NK cells

To understand how transcription of LRC and NKC genes compares with the rest of the transcriptome, their positions were calculated using the read counts obtained from UniMMap. The highest ranked of the NKC in PBMCs is *KLRK*, ranked at 9530 and 9497 in animal 1020 and 1021 respectively (figure 3-12). The LRC gene with the highest ranking in PBMCs is *KIR2DS1*, ranked at 11013 and 11597 respectively. Of the NKC and LRC genes with recorded read counts, *KLRH5* (15927) and *3DXL3* (15806) in animal 1020 PBMCs and *KLRH1* (15257) and *3DXL2* (15255) in animal 1021 PBMCs are ranked lowest for each of the respective gene complexes. There is a difference of 6397 positions for the highest and lowest transcribed gene of the NKC in animal 1020 PBMCs and a difference of 5760 positions in animal 1021 PBMCs. The difference for the LRC is smaller, 4793 positions in animal 1020 PBMCs and 3658 positions in animal 1021 PBMCs.

Within the NK population, *KLRI2* (5083) and *KLRE* (7414) are the highest ranked NKC genes and *2DS3* is the highest ranked LRC gene (8128 and 9710) in animal 1020 and 1021 respectively. The lowest ranked NKC and LRC gene in animal 1020 is *KLRC1-6* (14379) and *3DXL3* (13343) respectively. For animal 1021 the lowest ranked NKC and LRC gene is *KLRH1* (13912) and *3DXL2* (12903). The NKC genes of animal 1020 span a range of 5215 positions and the LRC genes 9296 positions. In animal 1021 the highest and lowest NKC genes are 6498 positions apart and the LRC genes are 3193 positions apart.

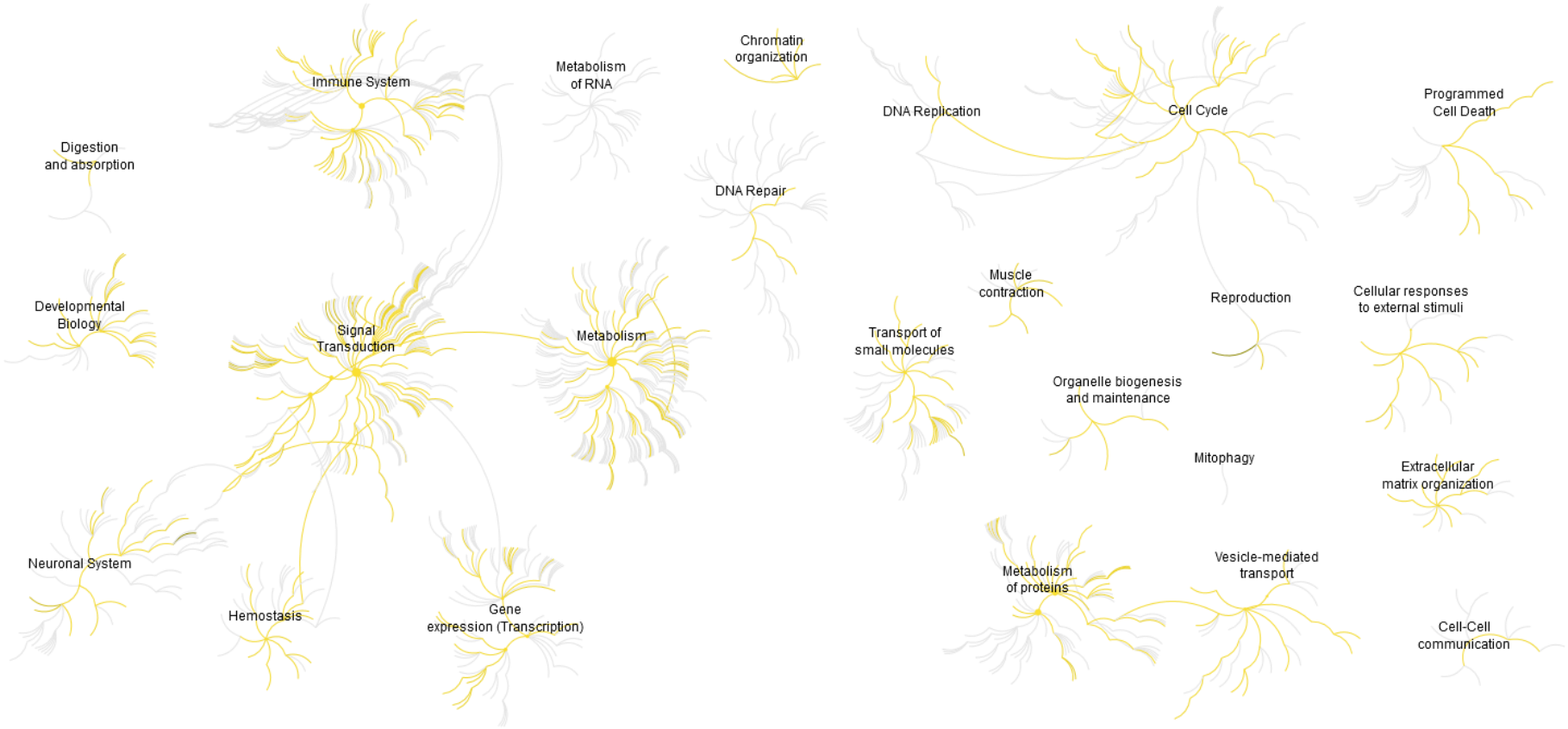


Figure 3-11. Pathway analysis of genes transcribed at least 1 fold higher in the NK population compared to the PBMC population. Reads for each sample were aligned to UMD 3.1 using GEMTools rna-pipeline. Weighted read counts were generated by GEMTools rna-pipeline using the UMD3.1.92 annotation. Read counts were normalised to reads per million, averaged across the two samples for each cell population and transformed to their $\log_2$ values. Gene IDs of the genes transcribed at least 1 fold higher in the NK population were imported into reactome for pathway analysis of known *Bos taurus* pathways.

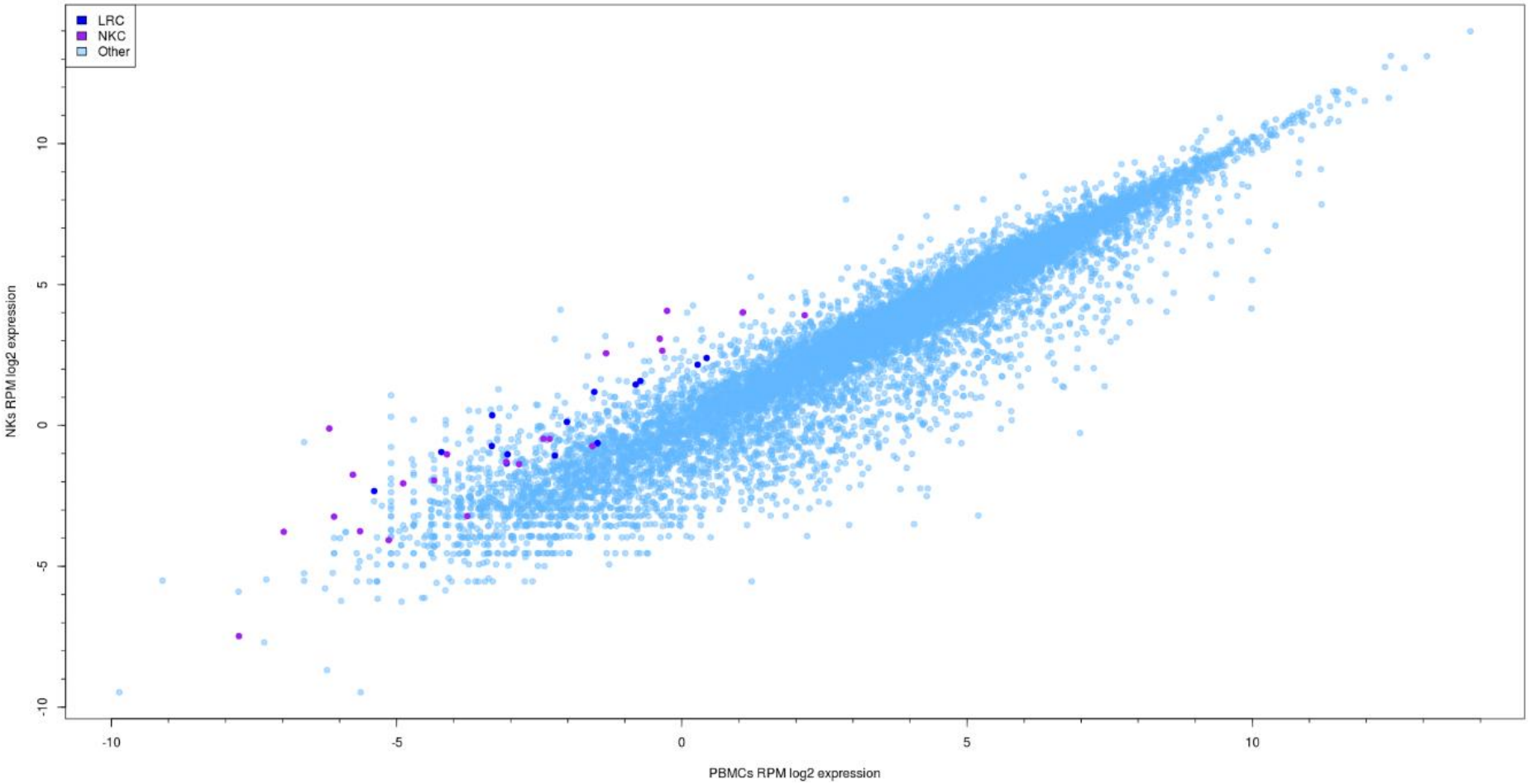


Figure 3-12. Comparison of global transcription between PBMCs and NK cells with incorporated LRC and NKC read counts from UniMMap. Global transcription read counts generated by GEMTools and LRC/NKC read counts from UniMMap were normalised to RPM and averaged for each cell type. Normalised read counts were transformed to $\log_2$ values and plotted using the plot function in R. Dark blue dots indicate genes of the LRC, purple dots indicate genes of the NKC and light blue dots indicate the remainder of the genes.

3.4 Discussion

Illumina RNA-Seq was carried out on PBMCs and NK cells isolated from two MHC homozygous animals. Our custom pipeline UniMMap (discussed in chapter 2) was used to accurately quantify and compare transcription of the LRC and NKC in these cell populations. We carried out an aligner comparison using our data to highlight the issue of inconsistent read counts between analysis methods. Genes that were transcribed at a higher level in NK cells were analysed using Reactome to determine to which functional pathways they belonged. Transcription of these loci was also compared to the global transcriptome of each of the populations to determine to what extent they were transcribed in relation to the rest of the transcriptome.

3.4.1 Transcription of the NKC and LRC genes is most likely not exclusive to NK cells

Isolation of NCR1+ cells resulted in a 16.5x increase in observed *NCR1* transcription and FACS of both populations showed a ~30x increase in NCR1+ cells after enrichment. Despite this, the maximum increase in transcription in the NK population compared to PBMCs in the LRC and NKC was 12.9x and 10.1x respectively. This suggests that these genes are transcribed in other cell types within the PBMC population. If expression were limited to NK cells, a higher difference in transcription between the two cell populations would have been observed. In humans, expression of KIR can be detected in up to 5% of peripheral blood T cells, including CD4+, CD8+ and gamma delta T cells (Phillips 1995). KLR expression has also been observed in CD8+ T cells (Coles 2000). As CD3+ T cells make up 45-70% of PBMCs, the presence of KIR and KLR on a subpopulation of T cells would explain the observed difference in transcription between PBMCs and NK cells. Higher transcription of *KLRC1-6*, *KLC1-5* and *KLRH1* was observed in the PBMC population, suggesting that these receptors may be preferentially expressed on a different immune cell subtype.

In humans, both inhibitory KIR and KLR are able to inhibit T cell function, even in the presence of a stimulatory signal from the T cell receptor (D'Andrea et al. 1996; Nakajima, Tomiyama, and Takiguchi 1995). Expression of KIR by T cells also mediates T cell tolerance to self-antigens (Huard and Karlsson 2000). The transcription here of *KIR* and *KLR* in immune cell types other than NK cells,

suggests they have similar functionality in cattle. Further analysis on pure immune cell populations (chapter 4), coupled with the disproportionally low *NCR1* transcription in the PBMC populations, shows that this transcription is not simply due to the presence of NK cells in the PBMC populations.

3.4.2 Non-functional status in the reference genome does not correlate with an absence of transcription

Most of the transcription observed in the NKC occurs between *KLRI2* and *KLRE*. Located within this region of comparatively high transcription is *KLRI2*, which encodes an activating receptor, which is a null allele in the cattle reference genome. The gene is however not heavily disrupted, a single nucleotide substitution in the initiation codon is predicted to prevent transcription, and a potential alternative start site is present. As high *KLRI2* transcription is observed in our samples it suggests that it is either functional, or that a functional allele exists. The only genes predicted to be non-functional for which transcription was not observed were *KLRH7* and *KLRH3*. However, some of the *KLR* are lowly transcribed with read counts for some less than 0.1 RPM. It is therefore possible that these two genes are transcribed, but not at a detectable level. These genes could be tightly regulated and only expressed under certain conditions, such as cytokine stimulation.

Transcription of all the genes predicted to encode null alleles in haplotype 1 was observed in both our samples, with the exception of *1DP2* and *1DLP1*, which cannot be accurately quantified by UniMMap as they lack any unique regions. Both of these genes however are highly disrupted and are considered gene fragments, rather than full-length genes, and are highly unlikely to produce a functional receptor even if transcribed. The gene *3DXL6*, which is inactivated in haplotype 1 by a small number of mutations, was observed in both the samples. *3DXL6* is considered the most divergent *KIR* in cattle; both intact inhibitory and activating, as well as non-functional alleles have been discovered (Sanderson et al. 2014). Transcription of the *3DL* lineage genes, *2DS3*, *2DS2* and *2DS1*, the known alleles of which are all non-functional, was also observed and *2DS3* was the highest transcribed *KIR* in either animal. A stop codon in domain 0 and two stop codons in domain 2 in all of the *2DS2/3* alleles, as well as an additional stop codon in the transmembrane domain of some of the alleles cause the genes to be predicted as non-functional. Of the two known alleles of *2DS1*, one is

rendered non-functional by two stop codons in both domain 0 and 2, as well as a stop codon in the transmembrane domain. The other *2DS1* allele is non-functional due to a stop codon in the transmembrane domain; it also has a missense mutation in domain 0. Due to the multiple mutations present in each allele, it is unlikely these null alleles produce a functional protein. As mentioned previously for *KLRI2* in the NKC, it is possible that functional alleles of these genes exist. For human *KIR*, multiple mechanisms of transcriptional silencing, such as DNA-methylation (Santourlidis 2002) and dsRNA-mediated silencing of the proximal promoter (Pascal 2006) are employed. If these genes do not have functional alleles, that they remain transcriptionally active may suggest an alternate function of the mRNA, such as transcriptional regulation of other *KIR*. The truncated human *KIR3DP1* is predicted to encode for a protein that is secreted into the extracellular medium, highlighting the existence of *KIR* that do not encode the prototypical cell surface receptor.

3.4.3 Some NKC/LRC genes vary in amount of transcription between animals but not in presence/absence

Within the LRC, a portion of the genes including *3DXL7*, *3DXL4*, *3DXL2* and *3DXS1* vary very little with regards to the level of transcription occurring in either animal. The same is seen in the NKC with the genes between *KLRA* and *KLRJ*. Whereas others are transcribed at a level at least twice as high in one animal. That the most variable genes between animals also tend to be the most transcriptionally active could suggest some receptors are always present in the repertoire at a constant level whereas others are conditional or variable. The observed differences could also be due to the nature of RNA-Seq, it only captures a 'snapshot' of the transcription occurring in the population. Whilst the multiple day sampling strategy employed here would likely reduce this effect, it may not be eliminated.

Comparisons to the transcriptional analysis by Allan et al (2015) is made difficult by the different methods used (qPCR of groups of genes vs. RNA-Seq of individual genes) and experimental design (individual NK clones vs. large population). However, it is clear that the percentage of clones transcribing a gene does not always correlate well with transcription levels of that gene in a large population of cells. Despite their observation of transcription of *KLRC2* on over 80% of NK clones, *KLRC2* genes are transcribed at a relatively low level. Depending on the

animal, between ~40-50% of NK clones transcribe *3DXL6* and between ~50-100% transcribe *3DXL1*. We observe *3DXL6* transcription at an average 8.88x higher than *3DXL1* transcription. This suggests that genes vary in transcriptional level, not just in the proportion of cells transcribing them. However, for some genes such as *KLRD1* and *KLRD2*, there does appear to be a correlation between percentage of clones transcribing each and their relative transcription levels. This could potentially mean that within the cells that transcribe them, their transcription is regulated in a similar manner.

Isolation was carried out using an antibody that binds the activating receptor NCR1; it is possible the differences observed here are due to differential responses to this activation. Alternatively, the genes that appear to be transcribed more highly in one animal than the other are may be due to the allele of that gene having higher sequence identity in that animal to the reference allele. This higher sequence identity would allow a greater number of reads to be mapped.

It is also possible that if animals from another herd or of a different breed were added to the analysis, the variation between the two animals would appear to be relatively minor. Two individuals from the same herd are likely to have similar exposures to pathogen, which has been shown to be a significant factor in the NK cell receptor repertoire (Lappalainen et al. 2013). The herd from which the animals were selected is also highly inbred. Consequently, the variation of *KIR* haplotypes is most likely limited. To build a robust picture of individual variation, future work would need to be done to increase sample size and the variety of individuals. Future work would also benefit from genotyping of NKC, LRC and MHC genes as this would enable determination of which differences were due to haplotypic variation, copy number variation, and also potentially linking transcription levels of receptor and ligand. Utilising animals from a pedigree herd with a known breeding structure would also allow more accurate comparisons of transcriptional variation.

3.4.4 The analysis method used influences the result

Three commonly used methods of generating RNA-Seq read counts, Tophat2, STAR and Salmon, were compared with each other and UniMMap. Tophat2 was selected as it is an extremely popular tool for the generation of RNA-Seq

alignments. According to Web of Science, the Tophat2 paper has been cited 3454 times as of 01/07/2018. Another popular RNA-Seq aligner is STAR, which was compared to Tophat2 by Teng (2016) and was shown to marginally better. Also included in the comparison was Salmon. Salmon differs from Tophat2 and STAR in that it uses the reference transcriptome rather than the genome and that it uses quasi-mapping instead of full alignments. Comparing the rankings of the LRC and NKC genes between the animals shows the scale of the disagreement between the analysis methods, 6/32 and 1/54 rankings are the same between methods respectively. Also problematic in the NKC was the variability on whether or not a gene generated read counts between the methods. Depending on which analysis method is chosen, the resulting read counts and determination of whether a gene is transcribed can vary greatly. The analysis tools compared here with UniMMap were all run using their default parameters. It is highly probable that by tweaking these parameters, more accurate results could be obtained. However, none of these tools would provide the same confidence in the determination of transcriptional status of the NKC/LRC genes as UniMMap. It is also likely that optimisation of these tools would have to be redone if analysing a different species, genetic region or read length/library type.

3.4.5 There are a large number of genes upregulated in NK cells but little difference within the top 100 genes

A total of 750 genes were found to be transcribed at a level at least 1 fold higher in NK cells than in PBMCs. Functional pathway analysis of these genes is problematic as cattle are a much less extensively studied species compared to humans. Approximately 66 % of the genes analysed were of unknown function. Unsurprisingly, the majority of the genes that did have a known function were within the immune system and signal transduction pathways. Within the signal transduction pathway, the pathways with the most hits was signalling by G protein-coupled receptors (GCPR). It has been shown GCPRs control the migration of NK cells from the bone marrow to activated lymph nodes, their extravasation into sites of inflammation as well as movement within lymph nodes and tumours (Walzer and Vivier 2011). A small number of the genes upregulated in NK cells compared to PBMCs are present in the top 100 NK genes. Within these upregulated genes, the majority are predominantly general leukocyte specific genes (leukocytes make up 70-90% of PBMCs in humans). Also

upregulated and present in the top 100 NK genes is lysozyme, which is expressed at a high level in NK cells and neutrophils.

3.4.6 Transcription of genes of the LRC and NKC is relatively low in *ex vivo* NK cells

Comparing the transcriptional ranking of the LRC/NKC genes in the context of global transcription reveals that they are transcribed relatively lowly compared to the rest of the transcriptome. This suggests than in immediately *ex vivo* NK cells; these genes are not particularly active. In humans, the NK receptor repertoire is very stable *in vivo*, unchanging over a period of six months, but is also capable of rapidly changing in response to external factors (Strauss-Albee 2015). The receptor repertoire also varies between individual NK cells, meaning that not all receptors are expressed on each NK cell. This would also result in lower overall transcription within the population. Stimulation with cytokines post-isolation may have the effect of increasing LRC/NKC genes transcription.

Chapter 4. Gene expression atlas for LRC and NKC genes in cattle

4.1 Introduction

Cattle have undergone a larger gene expansion within both the LRC and NKC than any other species studied to date (Sanderson, 2014; Schwartz, 2017). In contrast to the human and other simian primate LRC, in which the *3DL* lineage of *KIR* has expanded, cattle have also expanded the *3DX* lineage (Guethlein, 2007). The divergence of the *3DL* and *3DX* lineages is estimated to have occurred ~135 million years ago (mya) (Guethlein et al. 2007). The divergence of the *KIR* lineages from a common ancestor predates the radiation of placental mammals, which occurred 95.3-113 mya (Benton, 2007). The function of the *3DX KIR* is unknown, although they are predicted to encode receptors that bind MHC class I. Within the NKC, cattle and goats have evolved a second *KLRC* locus independently from other species, located between *KLRA* and *KLRJ*, and also possess a novel *KLRH*-like gene. Duplication of this novel gene has occurred several times in cattle (Schwartz, 2017). In summary, cattle possess a variety of genes within the LRC/NKC either not present, or not expanded, to the same extent in other species. The function of these genes in cattle has yet to be elucidated and very little is known about in which cell or tissue types they are transcribed/expressed.

In humans, expression of a majority of the genes within the LRC and NKC was first observed on NK cells. Expression has also been observed on CD8+ T cells, gamma delta T cells and to a lesser extent, CD4+ T cells (Ferrini, 1994; Nakajima, 1995; Phillips, 1995). KIR play a similar role in T cells as they do in NK cells, although they do not govern tolerance in T cells. Activating KIR in T cells also have a co-stimulatory effect with the T-cell receptor, rather than an independent effect (van Bergen and Koning 2010). Understanding where transcription of these receptors occurs in cattle is therefore an important step towards determining their function. UniMMap analysis of the LRC/NKC in multiple immune cell types provides an indication of the distribution of these receptors in the immune system. RNA-Seq datasets from L1 Dominette 01449, the animal from which DNA was used as part of the cattle UMD3.1 reference assembly, are available on the sequence read archive (SRA). The L1 Dominette

01449 datasets were used to create a gene atlas of LRC/NKC transcription across 24 tissue types.

NK cells have been shown to respond to infection by *Mycobacterium bovis* in humans and cattle (Esin, 2004; Denis, 2007). RNA-Seq datasets from 16 cattle in a study investigating gene expression in peripheral blood leukocytes during *M. bovis* infection were also downloaded from the SRA. Analysis of these datasets with our custom pipeline UniMMap allows the impact of *M. bovis* infection on the transcriptional profile on the LRC/NKC to be accurately assessed.

The influence of age on KIR expression in humans is still unclear (Lutz, 2005; Garff-Tarvernier, 2010; Almeida-Oliveira, 2011). NK cells from newborns have poor KIR expression (Garff-Tarvernier, 2010). A significant decrease in expression of KLRD, a molecule which associates with several receptors encoded by the NKC, has been observed in the elderly (Almeida-Oliveira, 2011). Although equivalent studies have yet to be carried out in cattle, expression of the NK markers NCR1 and CD2 have been shown to decline by day 1 post-birth, subsequently increasing into adulthood albeit never reaching the level observed at day 0 (Graham, 2009). Analysis of PBMC RNA-Seq datasets from three calves at multiple time points, up to 28 days post-birth, enabled a study of LRC/NKC transcription over this time period. Comparing transcription between the calves and their respective dams provides the first insight into the effect of age on LRC/NKC gene transcription in cattle.

4.2 Methods

4.2.1 UniMMap analysis of samples

The UniMMap analysis pipeline has been described and examined in detail in chapter 2. The pipeline was run on various RNA-Seq datasets to generate read counts for both the LRC and NKC. For each different RNA-Seq experiment, mappability calculations were carried out with GEM-mappability on the custom UMD3.1 transcriptome described in chapter 2. When calculating mappability, a kmer length matching the read length of the RNA-Seq reads used to ensure accuracy.

4.2.2 RNA-Seq data acquisition

RNA-Seq datasets from fluorescence-activated cell sorting (FACS) enriched PBMC subsets were kindly provided by Jessica Powell and Dr Liam Morrison at the Roslin Institute. Antibody usage for magnetic-activating cell sorting (MACS)/FACS is detailed in Table 4-1. B cells, monocytes and CD4+ T cells were isolated using FACS. A negative MACS sort to remove monocytes, CD4+ T cells and B cells was carried out before FACS sorting to isolate NK cells, gamma delta T cells or CD8+ T cells. Gamma delta T cells were sorted from this population on CD8-/NCR1-, CD8+ T cells were sorted on CD8+/NCR1- and NK cells on CD8-/NCR1+. Multiple NK cell RNA samples were combined per animal to generate enough material for sequencing.

Target cell	Antibody	Antigen
B cell	ILA58	Antibody light chain
CD4+ T cells	ILA12	CD4
CD8+ T cells	ILA51	CD8
Gamma delta T cells	GB21A	TCR1-N24 (δ chain)
Monocytes	ILA24	MyD-1
NK cells	AKS8	NCR1

Table 4-1. Antibody usage for FACS sorting of PBMCs to select for individual immune cell types.

RNA-Seq data from cattle infected with *M. bovis* (McLoughlin, 2014) was downloaded from the SRA – accession number: PRJNA257841. The study used 16 age-matched Holstein-Friesians - eight control and eight *M. bovis* infected. RNA was extracted from peripheral blood leukocytes. Sequencing was carried out using an Illumina Genome Analyzer IIx to generate single-end reads of length 84bp.

Azad K. Kaushik at the University of Guelph provided us with RNA-Seq data from six individuals (three dams and three calves – Table 4-2). PBMCs were isolated from blood collected from three calves at day 0, 7, 14 and 28 post-birth and from the three dams at day 0 and day 7 post-parturition. Samples were sequenced using an Illumina HiSeq 2500, generating 151bp paired-end reads.

Dam ID no.	Date of birth	Calf ID no.	Date of birth
3761	18/11/2007	4324	12/06/2013
3731	18/08/2007	4327	11/07/2013
4045	09/11/2010	4355	10/10/2013

Table 4-2. Date of birth and ID no. of dams and their respective calves (adapted from Pasman, 2017)

SRA project PRJNA379574 was downloaded to create a LRC/NKC expression atlas. RNA-Seq was carried out on 23 tissues from L1 Dominette 01449 and the testis from SuperBull 99375 (sire of Dominette). Sequencing was carried out using an Illumina NextSeq500 and paired-end reads of length 75bp generated.

4.2.3 Correspondence analysis

UniMMap read counts were used in conjunction with the ca R library (Nenadic, 2007) and ggplot2 (Wickham 2009) to produce correspondence analysis plots.

4.3 Results

4.3.1 LRC/NKC gene transcription between animals

Three RNA-Seq datasets from an enriched NK cell population isolated from three different Holstein Friesian (HF) animals (HF3457, HF3458 and HF3471) (Jessica Powell and Liam Morrison; Roslin Institute) were compared with the NK RNA-Seq data discussed in chapter 2. The method of isolation used by the Roslin Institute involved a negative FACS sort to remove gamma delta T cells, CD4+ T cells, monocytes and B cells and then a positive FACS sort for NCR1+ cells. In comparison, we positively selected for NCR1+ cells using antibody tagged MACS beads.

As was done for animals 1020 and 1021, Uni*MM*ap was run on samples HF3457, HF3458 and HF3471; read counts were normalised based on the number of reads mapped. NCR1 transcription was compared between the samples as a basic comparison of purity of the populations. The average NCR1 read counts of the three Roslin samples was 219.2; the average for 1020/1021 was 17.2 (supplementary figure 4-1). This suggests that a round of negative selection before positive selection of NCR1+ cells results in a purer population of NK cells.

Comparing the transcription of the LRC genes between the samples reveals a large difference between the Roslin samples and 1020/1021 from *3DXL2* at the 3' end to *2DL1* at the 5', all of which are transcribed much more highly in the Roslin samples (figure 4-1). The *KIR* located between *3DXL6* and *3DXL5* are however much more comparable between all of the samples. Transcription in either 1020 or 1021 is higher for *3DXL6*, *2DS3*, *3DXS3*, *2DS2* and *3DXS2* than at least one of the Roslin samples. Notably all of these genes are activating receptors in haplotype 1 with the exception of *3DXL6*, which while inhibitory in haplotype 1, has activating alleles on other haplotypes. The other activating receptors present in haplotype 1, *3DXS1* and *2DS1*, are transcribed at a much higher rate in the Roslin samples. Transcription of all of the detectable genes in haplotype 1 was observed in all five samples.

The pattern of transcription when comparing transcription of the NKC genes in the Roslin samples to 1020/1021 is similar to the LRC (figure 4-2). Many of the genes are transcribed at a rate much higher in the Roslin samples. However, *KLRC1-7, KLRC1*-5, *KLRH2* and *KLRHJ*, are transcribed more highly in at least one

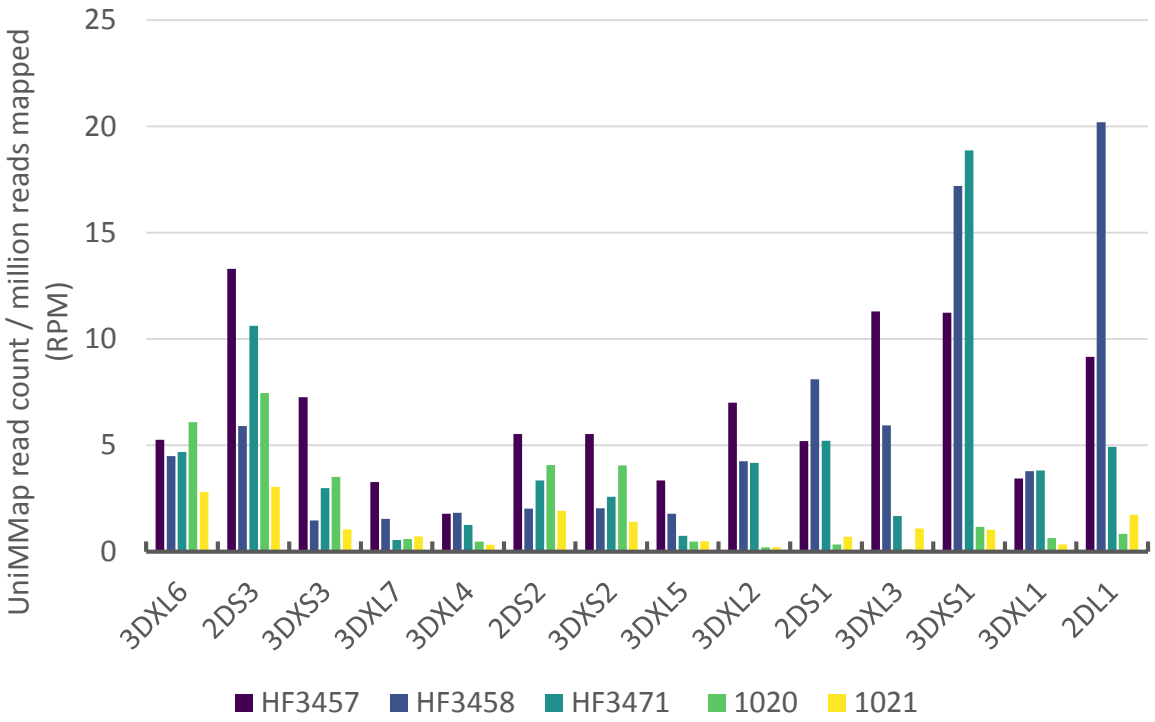


Figure 4-1. Comparison of LRC gene transcription in NK cells from five animals. UniMMap was run on RNA-Seq data from NK cells isolated from five cattle. NK cells from HF347, HF3458 and HF3471 were isolated using a combination of negative and positive sorting by FACS. MACS beads were used to positively select for NK cells from animals 1020 and 1021. Read counts from UniMMap were normalised based on the number of reads mapped.

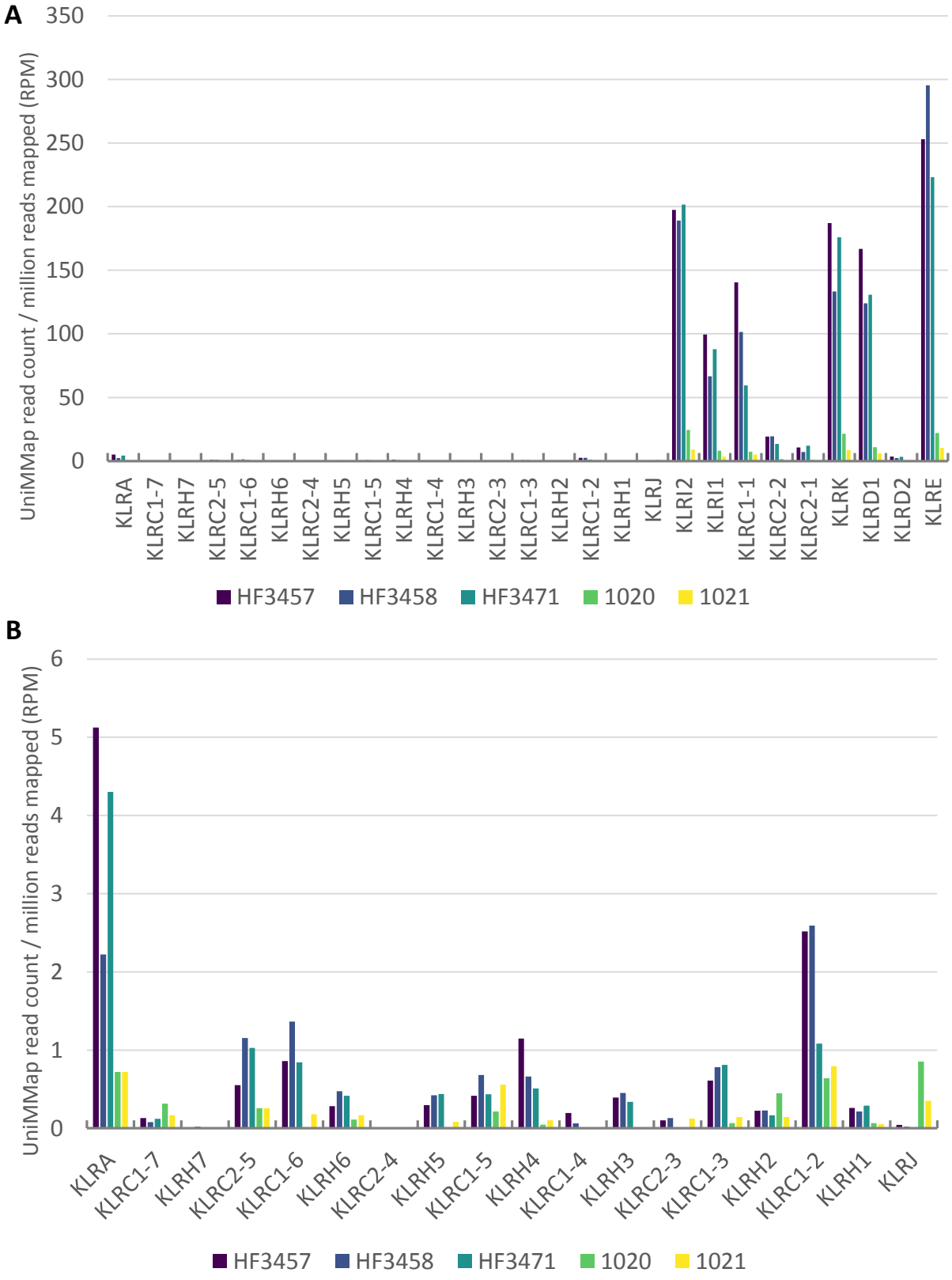


Figure 4-2. Comparison of NKC gene transcription (A) and the region of the NKC between *KLRA* and *KLRJ* (B) in NK cells from five animals. UniMMap was run on RNA-Seq data from NK cells isolated from five cattle. NK cells from HF347, HF3458 and HF3471 were isolated using a combination of negative and positive sorting by FACS. MACS beads were used to positively select for NK cells from animals 1020 and 1021. Read counts from UniMMap were normalised based on the number of reads mapped.

of 1020/1021. The region of comparatively high transcription between *KLRI2* and *KLRE* observed in 1020/1021 is also distinctly more active in the Roslin samples. Despite an average of 36.5 RPM mapped to the NKC, compared to the 2.7 RPM observed for 1020/1021, transcription of the region between *KLRA* and *KLRJ* remains similarly low. Some of the genes in this region of low transcription are not detected in all of the samples, including *KLR1-6*, *KLRH5*, *KLRC1-4*, *KLRC2-3* and *KLRJ*. No transcription was observed in any of the samples for either *KLRH7* or *KLRC2-4*.

We show that the method of NK cell isolation affects the amount of transcription observed for the genes of both the LRC and NKC. The difference is not limited simply to higher transcription in one method over another and varies on a gene-by-gene basis.

4.3.2 Transcription of the LRC and NKC genes is not exclusive to NK cells

In addition to the three NK cell datasets, the Roslin Institute also provided us with B cell, CD4+ T cell, CD8+ T cell, gamma delta T cell and monocyte datasets from the same animals. These samples were analysed with UniMMap to understand the extent to which LRC/NKC transcription occurred in these cell types.

Analysis of these immune cell types with UniMMap reveals that LRC gene transcription occurs in CD8+ T cells at an equivalent level to NK cells (figure 4-3). Transcription of *3DXS1* (5.86x), *3DXL1* (5.27x) and *2DL1* (3.62x) occurs at a much higher level in NK cells than CD8+ T cells. The only *KIR* transcribed highest in a cell type other than NK cells is *3DXL6*, occurring at a level 2.27x higher in CD8+ T cells. Gamma Delta T cells also appear to transcribe LRC genes, although to a lower extent than either NK cells or CD8+ T cells. All of the detectable *KIR* appear to be transcribed in NK cells, CD8+ T cells and Gamma Delta T cells. Transcription of LRC genes was observed at a very low level in the other cell types. The total RPM observed for the LRC genes in B cells was 0.24 RPM, in CD4+ T cells it was 1.21 RPM and for monocytes the total observed read counts was 0.20 RPM. These total read counts are all much lower than NK cells (79.30 RPM), CD8+ T cells (43.44 RPM) or Gamma Delta T cells (11.25 RPM). Transcription of LRC genes is not limited solely to NK cells, transcription is observed in all cell types, albeit at varying levels.

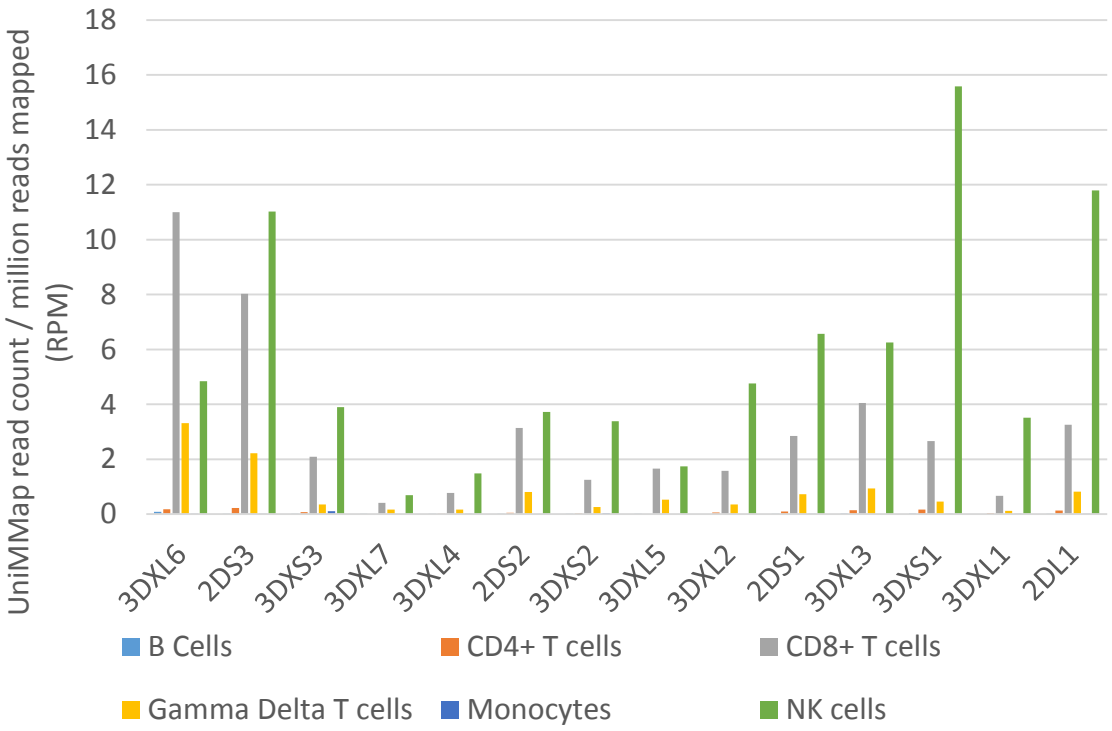


Figure 4-3. Comparison of average LRC gene transcription from six immune cell types from three animals. UniMMap was run on RNA-Seq data from each cell type from the three cattle. Read counts from UniMMap were normalised based on the number of reads mapped and then averaged for each cell type across the three animals.

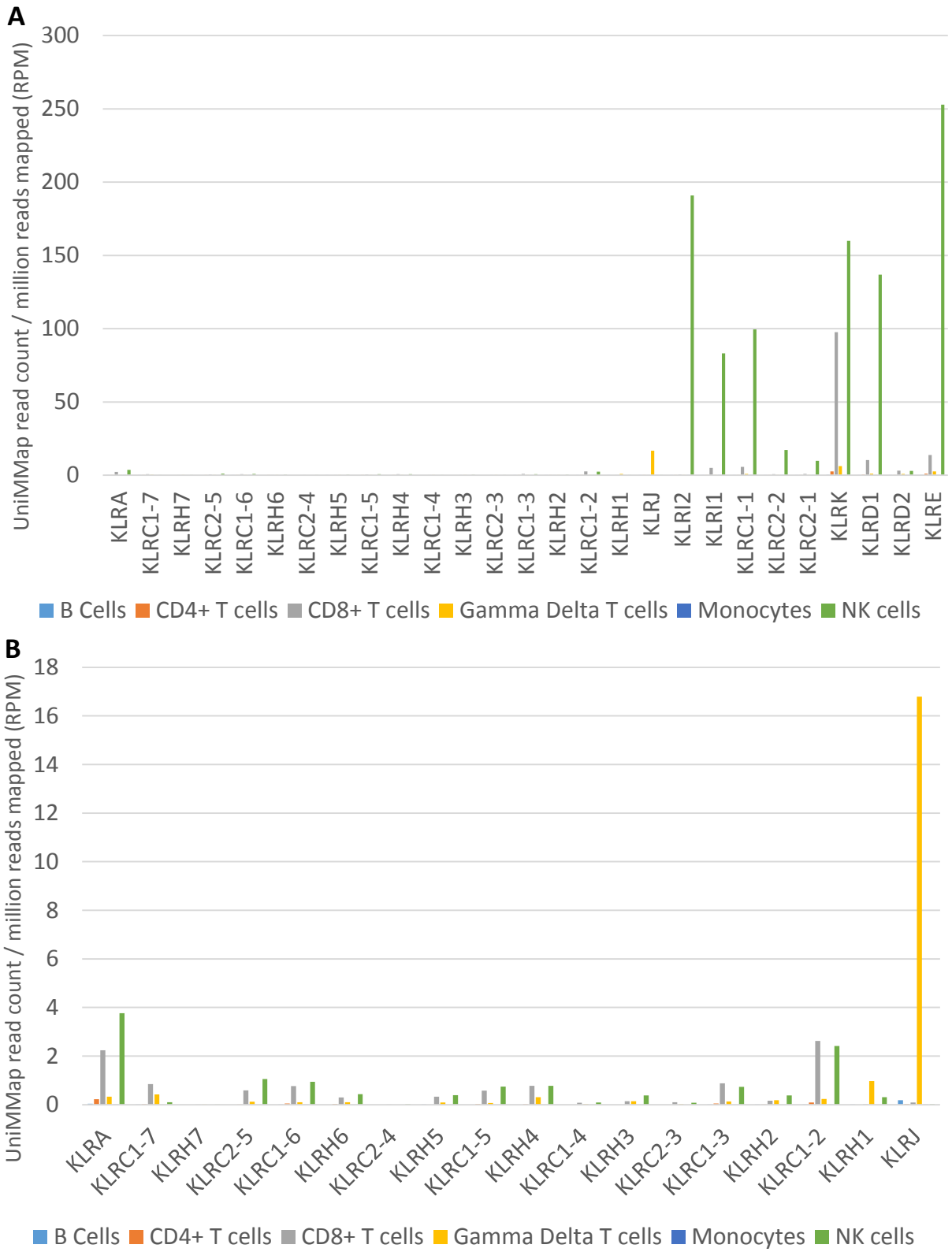


Figure 4-4. Comparison of average NKC gene transcription (A) and the region of the NKC between *KLRA* and *KLRJ* (B) from six immune cell types from three animals. UniMMap was run on RNA-Seq data from each cell type from the three cattle. Read counts from UniMMap were normalised based on the number of reads mapped and then averaged for each cell type across the three animals.

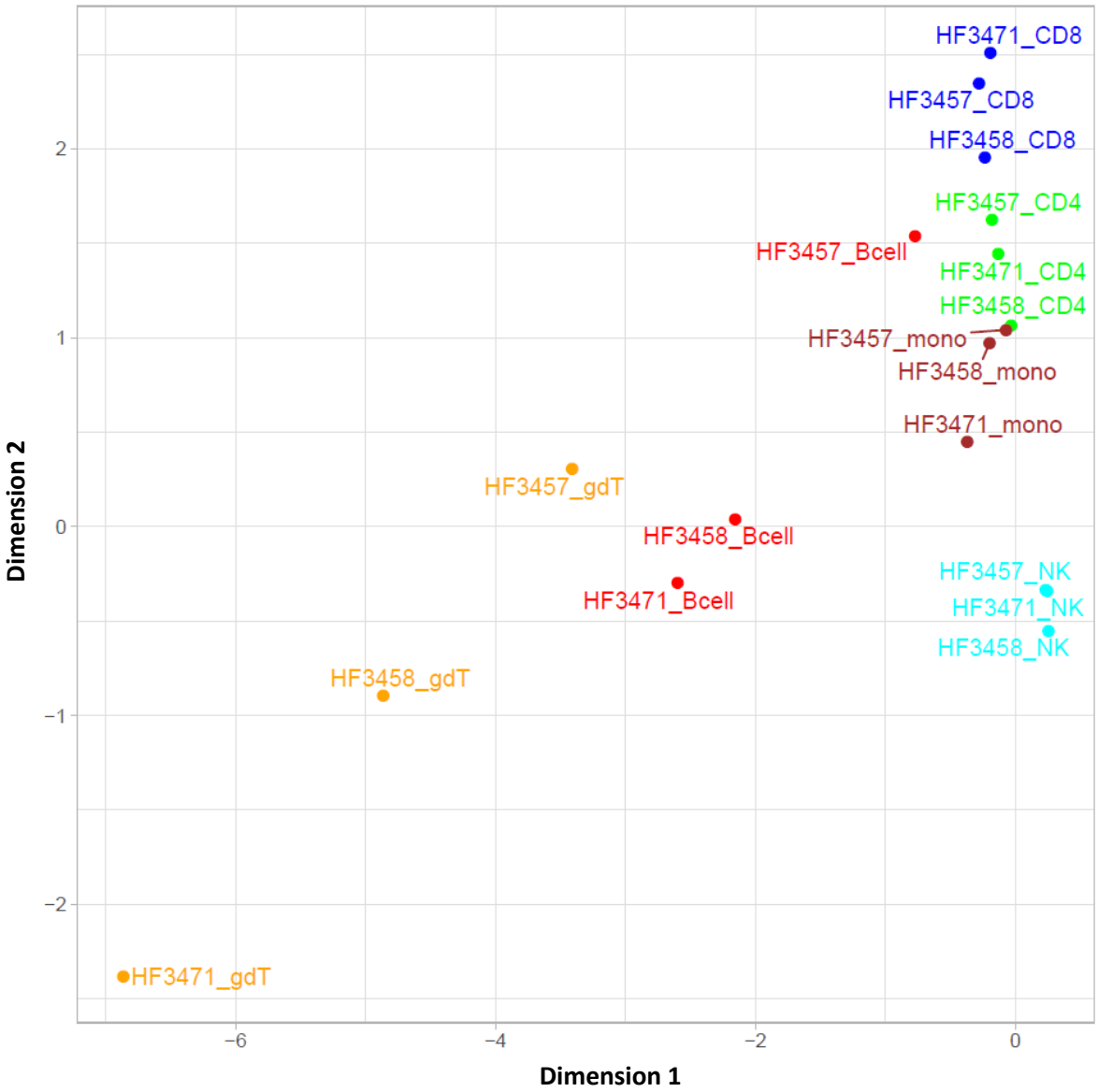


Figure 4-5. Correspondence analysis of LRC/NKC gene transcription from six immune cell types from three animals. UniMMap was run on RNA-Seq data from each cell type from the three cattle. Read counts from UniMMap were normalised based on the number of reads mapped. Correspondence analysis was carried out in R using the ca library and plotted with ggplot2.

The region of high transcription observed between *KLRI2* and *KLRE* in all analysed NK datasets is not transcribed to the same extent in any of the other examined cell types (figure 4-4A). The exception to this is *KLRK*, which is transcribed at a similar level in CD8+ T cells compared to NK cells. There is little transcriptional activity between *KLRA* and *KLRJ* in all of the examined cell types, with the exception of Gamma Delta T cells (figure 4-4B). Relatively high levels of *KLRJ* transcription is observed in Gamma Delta T cells. Transcription of *KLRJ* is 94.51x higher than in B cells (the cell type with 2nd highest *KLRJ* transcription) and 945.91x higher than in NK cells. Only 11/27 of the NKC genes are transcribed in all the cell types and transcription of the entire collection of NKC genes was only observed in NK cells and CD8+ T cells.

Correspondence analysis of the individual datasets shows the relationship between the cell populations as well as individual animals (figure 4-5). Cell types group together on both axis, indicating that there is a different pattern of transcription specific to each cell type. Both B cells and gamma delta T cells are separated to a higher extent across both axis than the other cell types. This suggests that for these cell types there is more variance in LRC/NKC gene transcription between animals.

Described here is the first known investigation of LRC/NKC transcription in multiple cattle immune cell populations. We show that transcription of these two loci is not exclusive to NK cells, particularly the LRC genes which is transcribed at a similar level in CD8+ T cells. We also show that transcription of NKC genes while much more prevalent in NK cells, does occur at a lower level in other cell types.

4.3.3 Variation exists in LRC/NKC transcription between non-infected and *M. bovis* infected animals

Natural killer cells have been show to play a role in restricting the replication of *Mycobacterium bovis* in infected macrophages (Denis et al 2007). We used data from a study by McLoughlin et al (2014) in which they carried out Illumina RNA-Seq on peripheral blood leukocytes from 16 Holstein-Friesian cattle. Eight of the cattle were animals naturally infected with *M. bovis* and the other eight were age and sex matched control animals. We used these datasets in combination

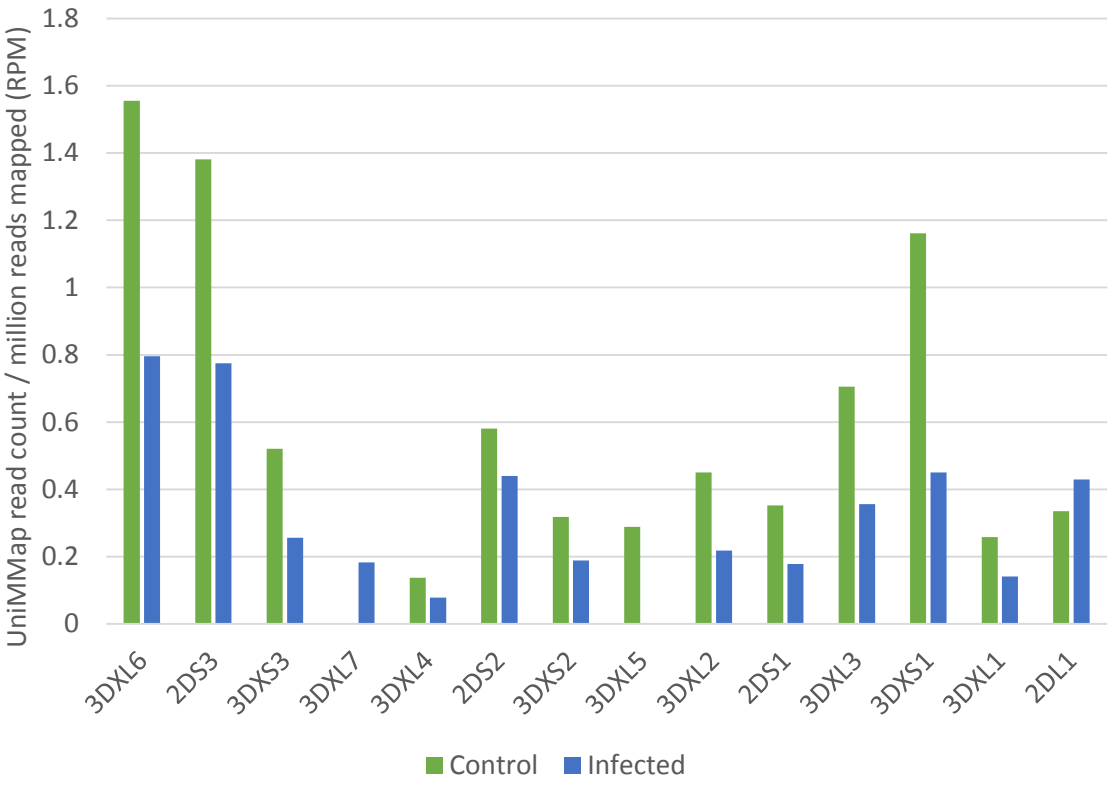


Figure 4-6. Comparison of average LRC gene transcription between control cattle and cattle infected with *M. bovis*. RNA-Seq data was obtained from the sequence read archive (Accession: PRJNA257841). The eight control and eight infected animal datasets were combined to produce two datasets, one control and one infected. UniMMap was run on the merged datasets. Read counts from UniMMap were normalised based on the number of reads mapped.

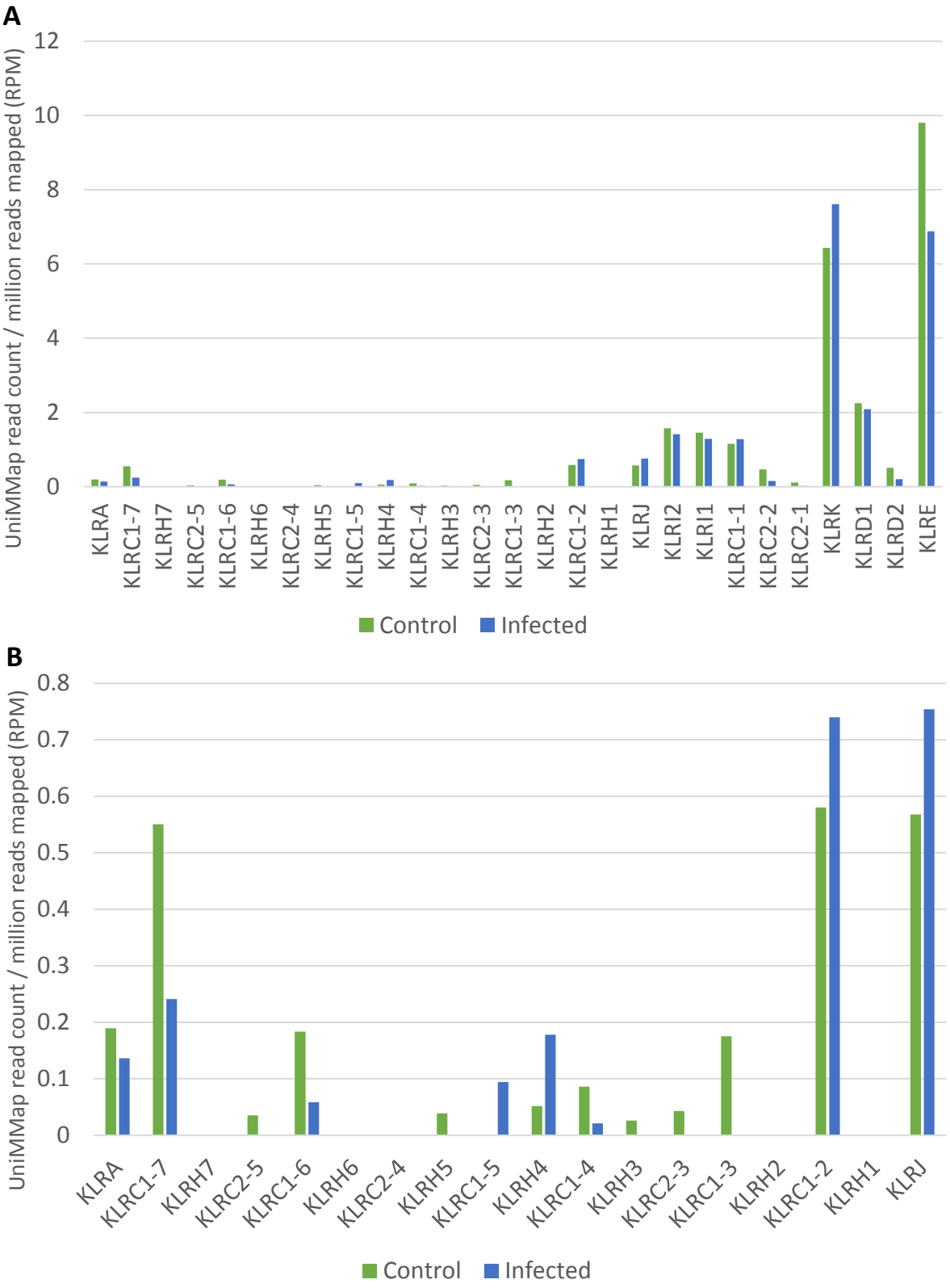


Figure 4-7. Comparison of average NKC gene transcription (A) and the region of the NKC between *KLRA* and *KLRJ* (B) between control cattle and cattle infected with *M. bovis*. RNA-Seq data was obtained from the sequence read archive (Accession: PRJNA257841). The eight control and eight infected animal datasets were combined to produce two datasets, one control and one infected. UniMMap was run on the merged datasets. Read counts from UniMMap were normalised based on the number of reads mapped.

with our pipeline UniMMap to attempt to understand if the genes of the LRC/NKC may play a role in infection by *M. bovis*.

As these datasets contained a comparatively low number of reads compared to the PBMC RNA-Seq we previously generated and analysed, combining the datasets was necessary. Subsequently we compared the transcription of the combined control animals to that of the combined infected animals. Merging of the datasets resulted in read counts similar to those observed for the PBMC dataset we generated and discussed in chapter 3.

Transcription of almost all of the LRC genes is higher in the control animals (figure 4-6). The exceptions to this are *3DXL7* and *2DL1*. Transcription of *3DXL7* was only observed in the infected animals and *3DXL5* only in the control animals. Transcription of *2DL1* was 1.28x higher in the infected animals than the control animals. The largest difference in read count for a gene transcribed in both datasets was *3DXS1*, which was 2.58x higher in the control animals. The total read count of the LRC in the control animals was 8.05 RPM, compared to 4.49 RPM for the infected animals.

Comparing the total number of reads mapped to the NKC shows little difference between control (26.28 RPM) and infected animals (23.15 RPM) (figure 4-7A). Transcription of several genes however was observed only in the control animals, albeit at a very low level (figure 4-7B). Transcription of *KLRC2-1* (5.34x), *KLRC1-4* (4.1x), *KLRC1-6* (3.13x) and *KLRC2-2* (3.06x) was considerably higher in the control population. The only gene transcribed higher in the infected population at an equivalent difference is *KLRH4* (3.45x).

Despite the low coverage of the available datasets, we show there is a difference in transcription levels between the non-infected and infected populations. Transcription of several genes was identified in only one of the populations and multiple genes have substantial differences in read count.

4.3.4 Analysis of multiple dams and calves reveals genotypic variation

RNA-Seq data from PBMCs isolated from three calves and their respective dam was provided to us by the University of Guelph (Pasman and Kaushik). PBMCs were isolated from the calves on their day of birth as well as 7, 14 and 28 days post birth. The datasets from the dams were generated from PBMCs isolated on

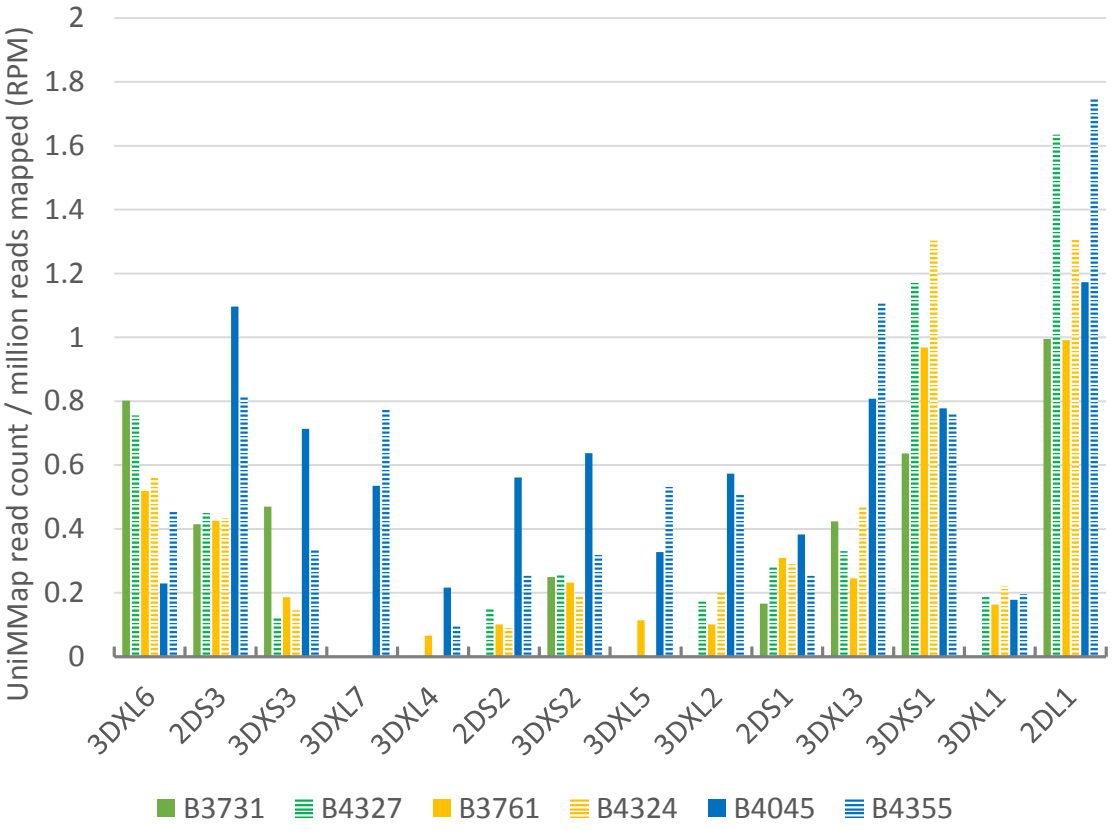


Figure 4-8. Comparison of LRC gene transcription of PBMCs from six cattle. PBMCs were isolated and sequenced from three dams and three calves, birthed by each of the dams, at multiple time points by the University of Guelph. Time points for each animal were merged to increase coverage and UniMMap was run on the merged datasets. Read counts from UniMMap were normalised based on the number of reads mapped. Bars of the same colour indicate dam (filled bar) and respective calf (striped bar).

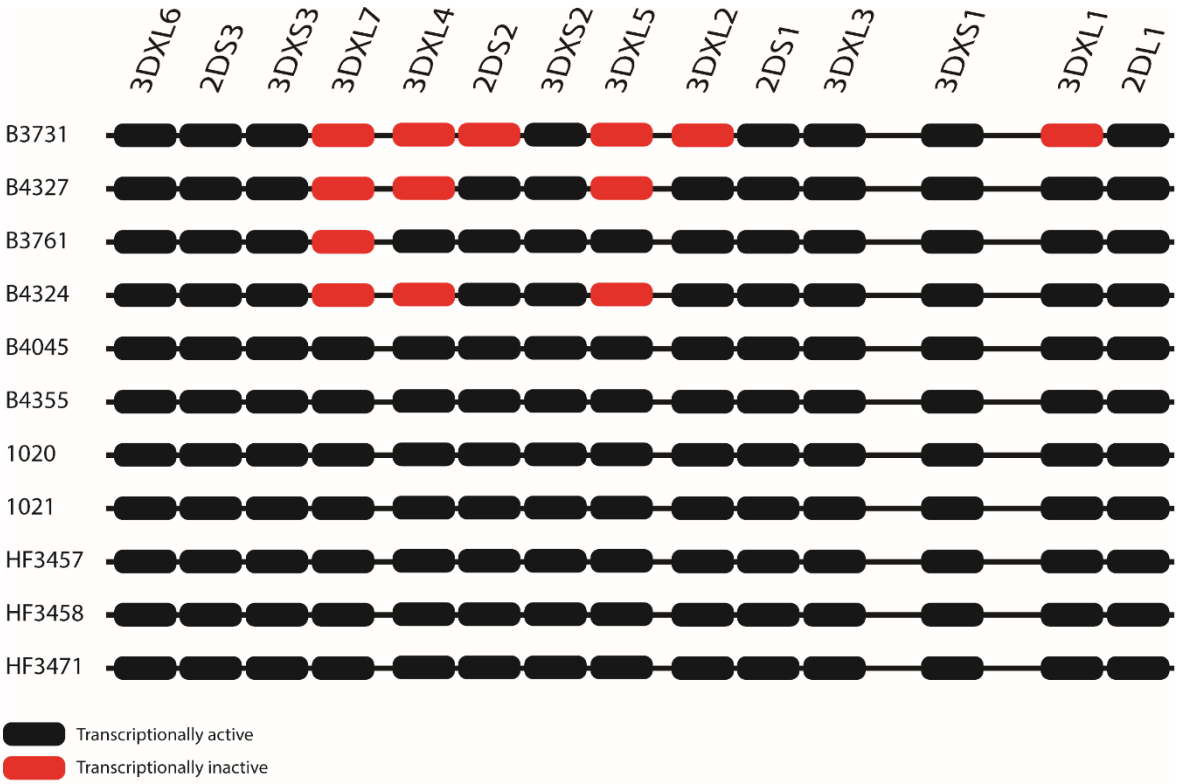


Figure 4-9. Transcriptional status of the LRC genes in the eleven cattle currently characterised. A red bar indicates that transcription of that gene was not detected in the animal, a black bar indicates transcription was observed.

the day of parturition and 7 days post parturition. Reads from the multiple time points were combined for each animal in an effort to increase coverage of the LRC/NKC. UniMMap was run on the combined datasets to obtain accurate read counts.

Comparing read counts between these animals reveals for the first time, variation in transcriptional status of the cattle LRC genes in an RNA-Seq experiment (figure 4-8, 4-9). Absence of transcription of one or more LRC gene occurs in four of the six animals (B3731, B4327, B3761 and B4324). Between the four animals, six *KIR* that have been consistently observed in all PBMC and NK datasets are undetected. Transcription of *3DXL7* was not observed in any of the four, three did not transcribe *3DXL4* or *3DXL5*, and one also did not transcribe either *3DXL2* or *3DXL1*. The two animals that are identical in *KIR* transcriptional status to the previously studied datasets are a dam (B4045) and its respective calf (B4355).

The number of *KIR* genes correlates well with total transcription of the LRC. Animals B4045 and B4355 transcribe the full repertoire of known *KIR* and have the highest total read count to the LRC (8.21 and 8.16 RPM respectively). The two animals (B4327 and B4324) which both do not appear to transcribe *3DXL7*, *3DXL4* and *3DXL5*, have similar read counts – 5.52 and 5.22 RPM respectively. B3731, which does not appear to transcribe six of the *KIR*, is the least transcriptionally active (4.16 RPM). The only outlier is B3761, despite only not transcribing *3DXL7*, its total read count is the second lowest (4.42 RPM).

As with all animals analysed so far, there is a region of relatively high transcription between *KLRI2* and *KLRE* compared to the rest of the NKC observed in the six animals (figure 4-10). The genes within this region are transcribed in all of the six animals. The remainder of the loci however is highly variable in transcriptional status, as in all previously studied animals (figure 4-11). Just two genes located between *KLRA* and *KLRJ* in the 11 individual cattle examined so far are identical in transcriptional status in all animals. *KLRC1-3* was observed in all animals and *KLRC2-4* was absent in every animal. Both *KLRA* and *KLRJ* were absent only once, in two separate animals.

Correspondence analysis of LRC/NKC transcription of the six animals shows a relationship between calf and dam (figure 4-12. Dams and calves also appear to

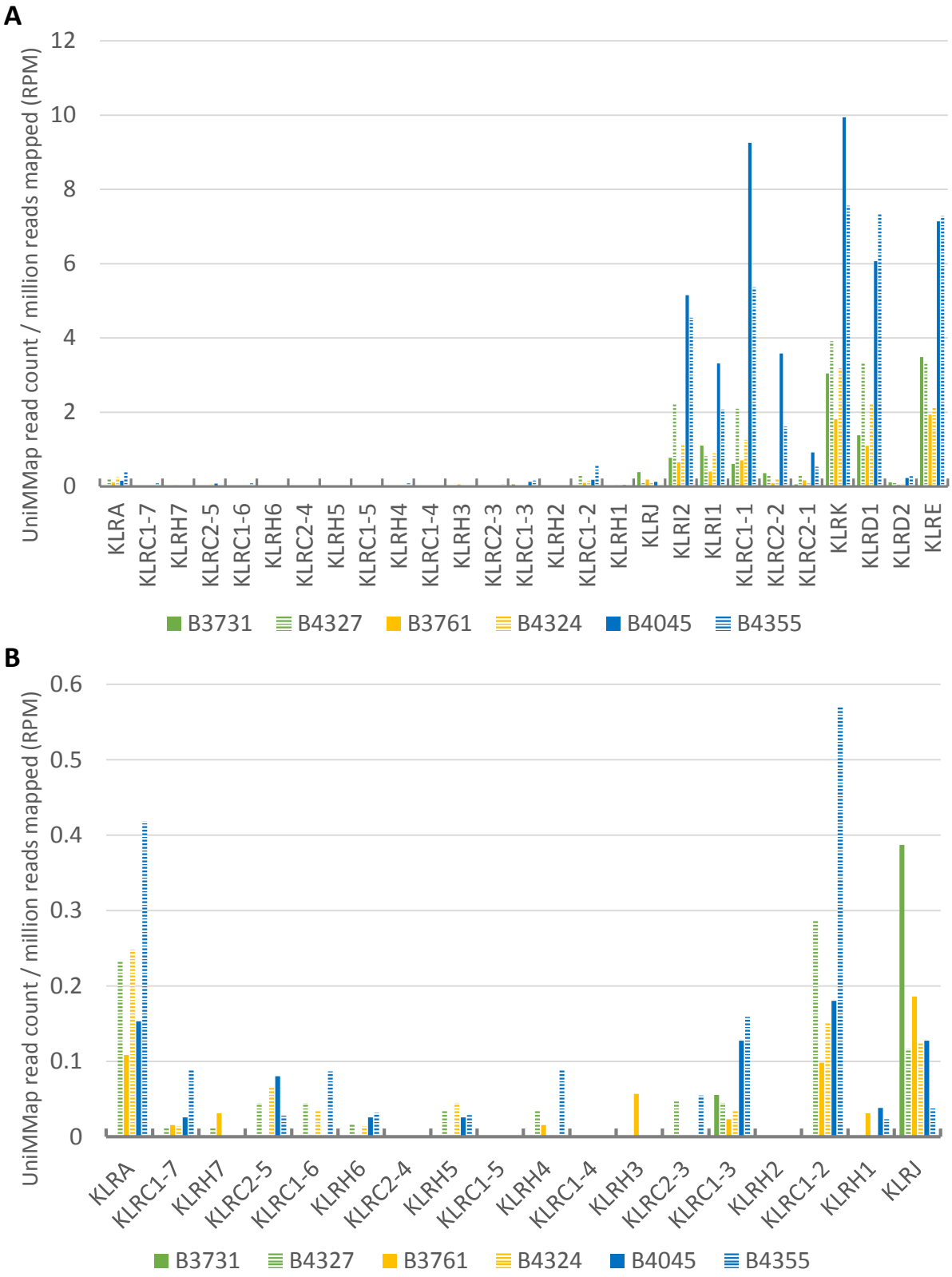


Figure 4-10. Comparison of NKC gene transcription (A) and the region of the NKC between *KLRA* and *KLRJ* (B) of PBMCs from six cattle. PBMCs were isolated and sequenced from three dams and three calves, birthed by each of the dams, at multiple time points by the University of Guelph. Time points for each animal were merged to increase coverage and UniMMap was run on the merged datasets. Read counts from UniMMap were normalised based on the number of reads mapped. Bars of the same colour indicate dam (filled bar) and respective calf (striped bar).

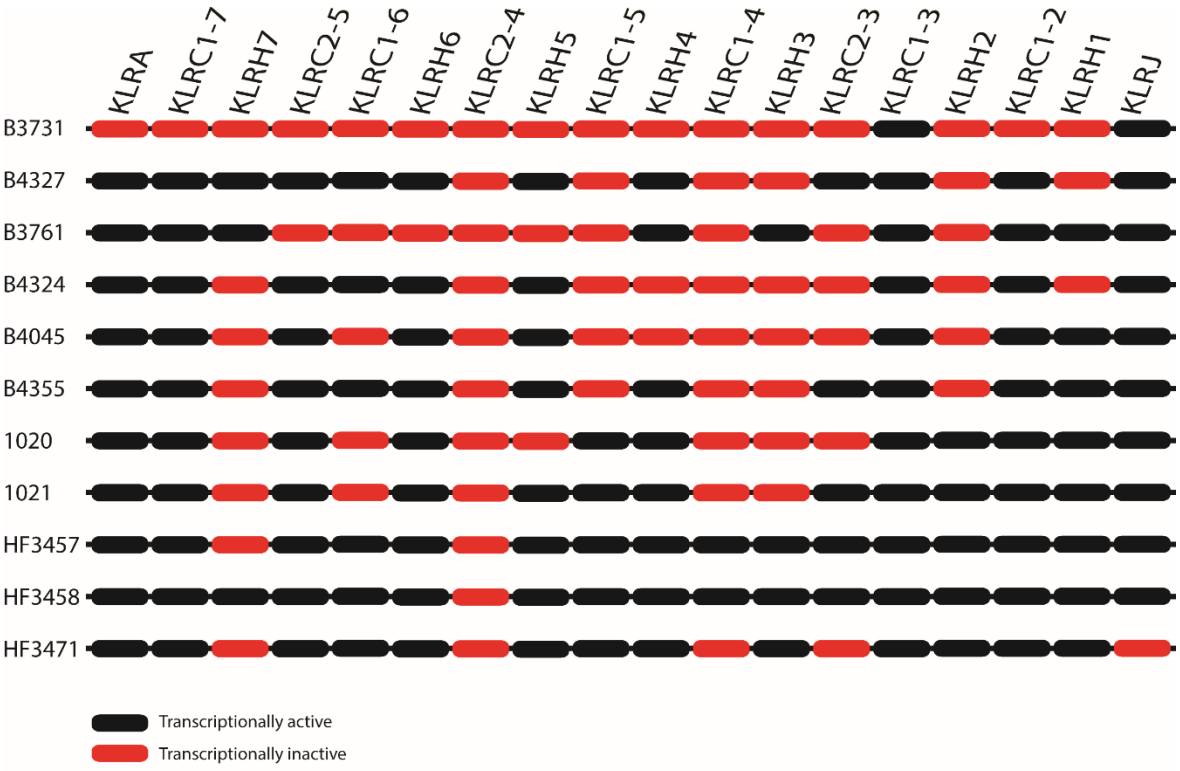


Figure 4-11. Transcriptional status of region of the NKC between *KLRA* and *KLRJ* in the eleven cattle currently characterised. A red bar indicates that transcription of that gene was not detected in the animal, a black bar indicates transcription was observed.

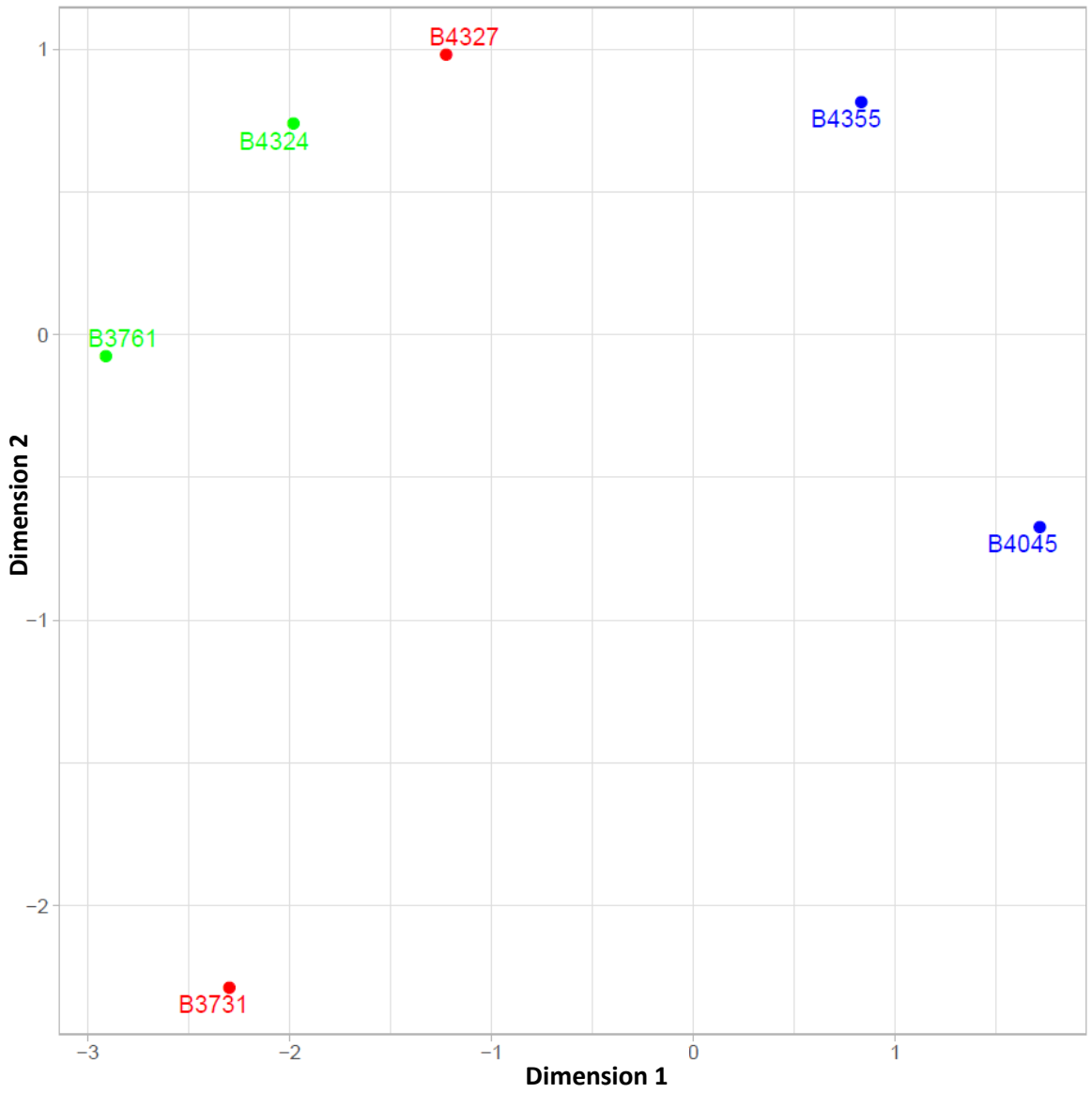


Figure 4-12. Correspondence analysis of LRC/NKC gene transcription from six cattle. PBMCs were isolated and sequenced from three dams and three calves, birthed by each of the dams, at multiple time points by the University of Guelph. Time points for each animal were merged to increase coverage and UniMMap was run on the merged datasets. Read counts from UniMMap were normalised based on the number of reads mapped. Correspondence analysis was carried out in R using the ca library and plotted with ggplot2. Points of the same colour indicate the dam/calf relationship.

separate on the Y axis suggesting a difference in the transcriptional pattern of the LRC/NKC between young and older animals.

There appears to be large and consistent variation in transcriptional status of genes of both the LRC and NKC. Analysis of the LRC across multiple animals reveals that there may be a number of *KIR* that are framework genes, present and functional in all cattle. Within the NKC there is a region of genes that are relatively highly transcribed in all animals. Also present in the NKC is a region of much lower relative transcription that appears to be highly variable in transcriptional status between animals.

4.3.5 Level of NKC and LRC transcription varies over 28 days post birth

The RNA-Seq data from the three calves was combined at each of the four time points (Day 0, Day 7, Day 14 and Day 28 post-parturition) and analysed with UniMMap to understand what changes the NK receptor repertoire undergoes after birth.

Comparing the total transcription of the LRC/NKC over the four time points shows that transcription decreases between day 0 and day 14 post birth (figure 4-13). This decrease in transcription is followed by an increase to a total larger than the day 0 value by day 28. Transcription of the NKC is 4.23x higher at day 28 than day 0. The total read count observed to the LRC is 2.55x higher at day 28 compared to day 0. Analysing the transcription of both the LRC and NKC across the time points shows that the proportional transcription of each genes remains relatively similar, despite the variable total read count (figure 4-14, 4-15). Minor variation can however be observed across the time points. Transcription of *3DXL7* and *2DS2* appears is not detected at day 7 but is observed at all other time points. The proportion of *3DXS1* transcription increases from 15.6% of the total at day 0 to 26.7% at day 7, before decreasing to 13.4% at day 14 and 16.1% at day 28. Despite the 4.23x increase in NKC transcription, as with the LRC there is little variation of the proportion each gene contributes to the total. Minor variation also exists in the NK, transcription of *KLRC1-4* and *KLRC2-4* is only observed at day 28. Transcription of *KLRH2*, *KLRH6* and *KLRC2-3* was observed at all time points except day 14. The amount of transcription observed for the currently used cattle NK marker *NCR1* follows the same pattern of

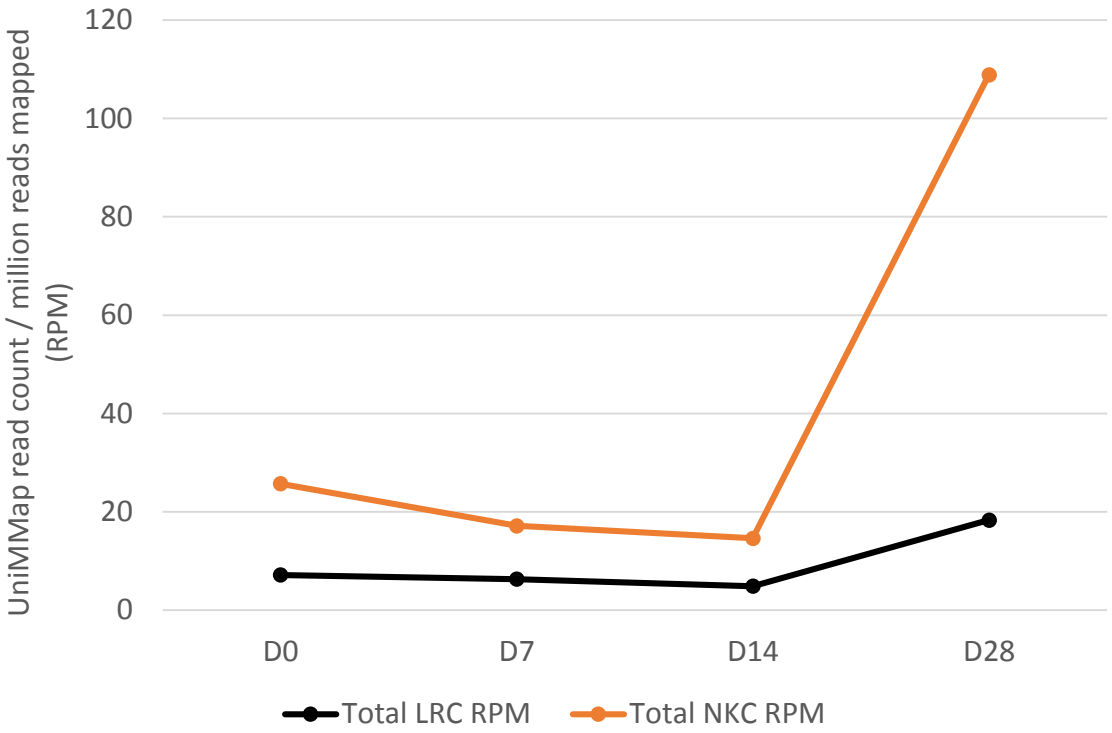


Figure 4-13. Average read count across the LRC and NKC at multiple time points post-birth. PBMCs were isolated and sequenced from three calves at day 0, day 7, day 14 and day 28 post-birth by the University of Guelph. Sequencing data from the three calves was merged at each time points to increase coverage. Read counts from UniMMap were normalised based on the number of reads mapped.

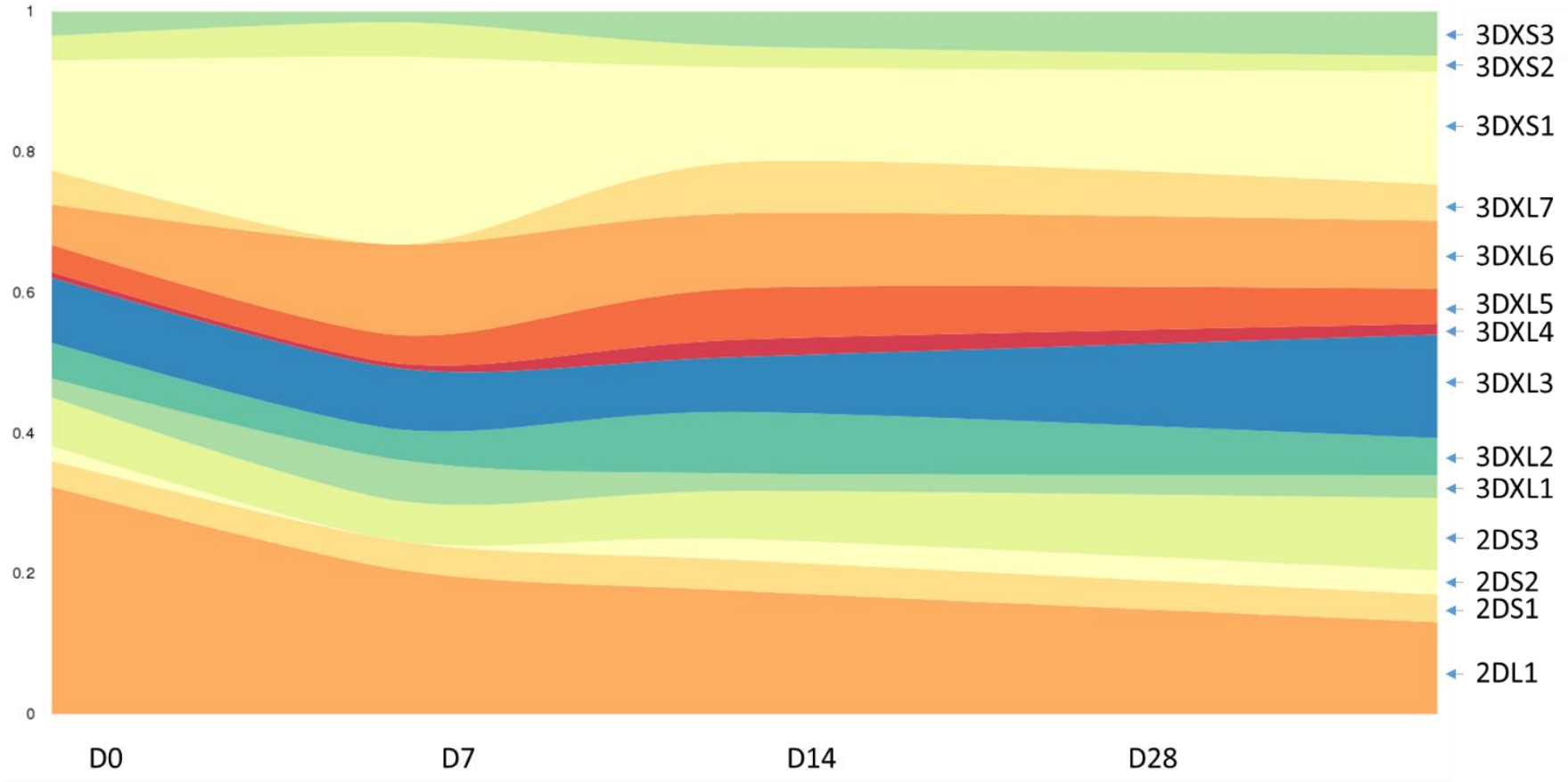


Figure 4-14. Streamgraph of proportional LRC gene transcription at multiple time points post-birth. PBMCs were isolated and sequenced from three calves at day 0, day 7, day 14 and day 28 post-birth by the University of Guelph. Sequencing data from the three calves was merged at each time points to increase coverage. Read counts from UniMMap were normalised based on the number of reads mapped. Read counts for each gene were calculated as a percentage of total transcription at each time point.

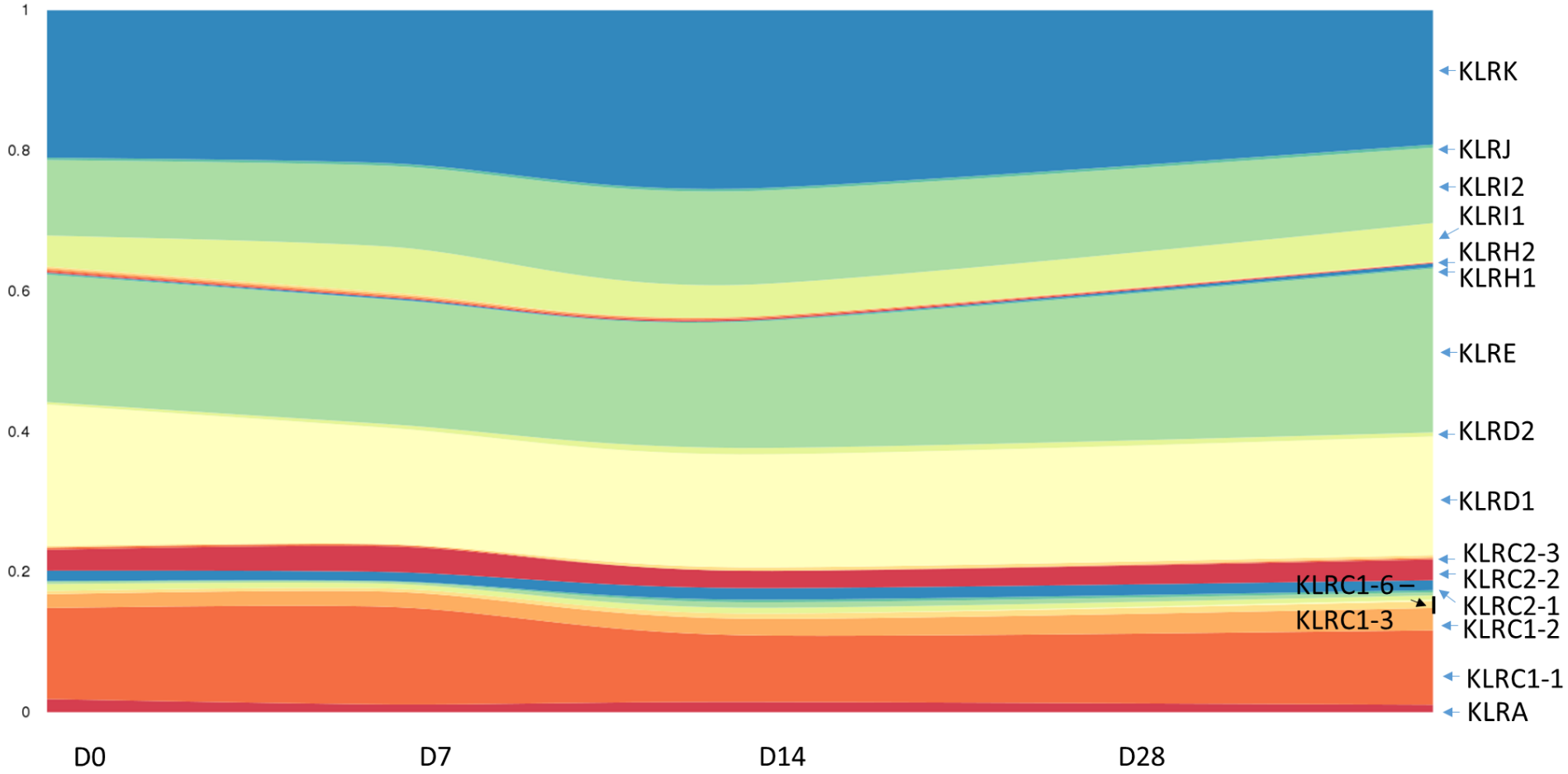


Figure 4-15. Streamgraph of proportional NKC gene transcription at multiple time points post-birth. PBMCs were isolated and sequenced from three calves at day 0, day 7, day 14 and day 28 post-birth by the University of Guelph. Sequencing data from the three calves was merged at each time points to increase coverage. Read counts from UniMMap were normalised based on the number of reads mapped. Read counts for each gene were calculated as a percentage of total transcription at each time point.

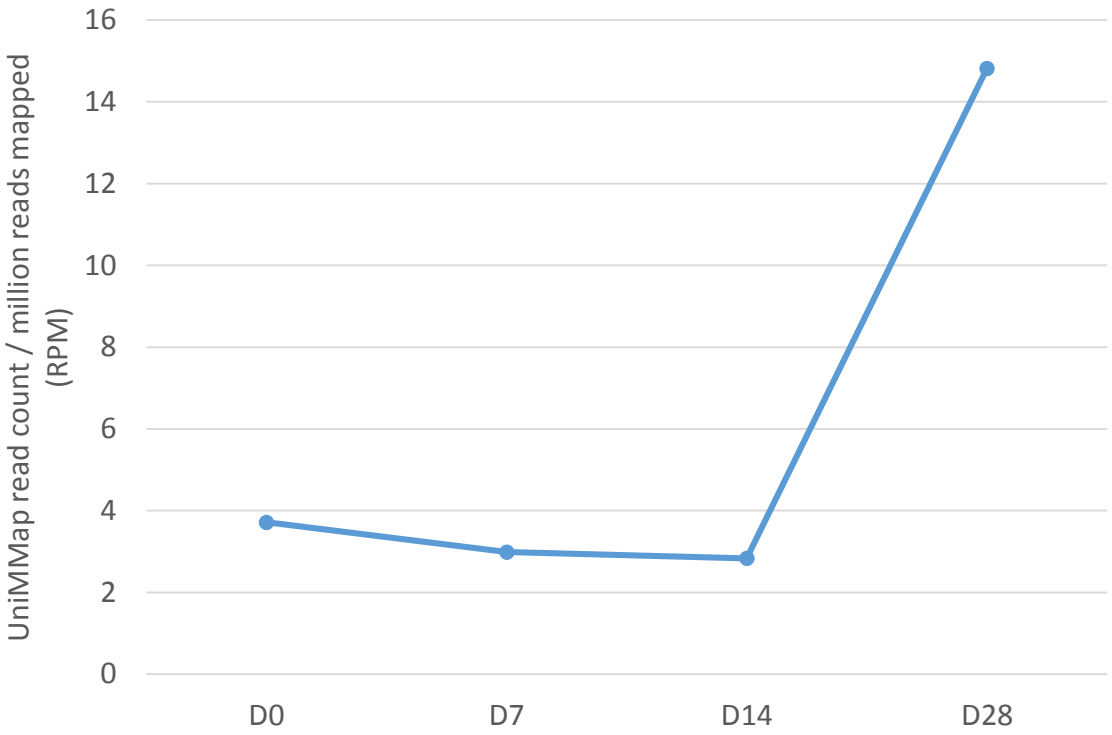


Figure 4-16. Average NCR1 read count at multiple time points post-birth. PBMCs were isolated and sequenced from three calves at day 0, day 7, day 14 and day 28 post-birth by the University of Guelph. Sequencing data from the three calves was merged at each time points to increase coverage. Read counts from UniMMap were normalised based on the number of reads mapped.

variability as the NKC (figure 4-16). The total read count of *NCR1* at day 28 is 3.99x higher than day 0.

The results of this analysis suggest there is an expansion of transcription of NK receptors between day 14 and 28 post birth. This expansion however is evenly distributed across the entire receptor repertoire, with some minor variation observed.

4.3.6 LRC/NKC gene transcription varies between tissue types

RNA-Seq data generated for the generation of a cattle gene atlas was accessed and analysed to compare LRC/NKC gene transcription in multiple tissue types. A total of 24 tissue types were sequenced as part of the gene atlas and all 24 were assessed for LRC/NKC gene transcription. With the exception of the testis sample which originated from SuperBull 99375, tissue samples originate from the animal L1 Dominette 01449, which was used in the UMD3.1 reference assembly.

There is large variation of LRC/NKC gene transcription between the examined tissue types (figure 4-17). No gene appears to be ubiquitously transcribed, despite all tissues transcribing at least one *KIR* and one *KLR* gene. Transcription of *KLRE* is observed in all of the tissues with the exception of the hypothalamus. Within the *KIR*, *2DL1* is the most ubiquitously transcribed, absent in six tissues. Transcription of *2DS3* and *3DXL6* is absent in seven tissues, four of which are the same tissue type. Neither *3DXL7*, *3DXL4*, *KLRC2*-4, *KLRH6* or *KLRH3* were observed in any sequenced tissue.

Comparing total read counts for each gene complex across the multiple tissues shows that NKC and LRC gene transcription follows a similar pattern across the tissue types (figure 4-18). LRC gene transcription is higher than NKC transcription in 14 out of the 24 tested tissue types. However total NKC gene transcription of all tissues combined is 1.73x higher than the LRC. The three most transcriptionally active tissues for both the NKC and LRC genes were lymph node, mammary gland, and lung. The duodenum is the fourth most transcriptionally active tissue when examining NKC transcription, but the 7th most active for LRC gene transcription. LRC gene transcription was higher in bone marrow, atrium and ventricle than duodenum tissue. Regression analysis of total NKC and LRC gene transcription in each tissue type suggests that

transcription of the genes of the two loci is correlated, with an R^2 value of 0.9219 (figure 4-19).

Transcription of genes of the LRC and NKC appears to be variable between tissue types. Transcription of at least one *KIR* and *KLR* gene could be detected in every tissue type examined. Despite higher transcription of the NKC than LRC in all examined PBMC and NK datasets, 14/24 of the tested tissue types had higher total LRC transcription. Transcription of *NCR1* is comparatively high in bone marrow, lung and lymph node (supplementary figure 4-2).

Figure 4-17. Heatmap of LRC/NKC gene transcription in 24 tissue types. RNA-Seq data was downloaded from the sequence read archive (Accession: PRJNA379574). The BioProject contains RNA-Seq from 23 Dominette tissues as well as the testis of SuperBull 99375. Read counts from UniMMap were normalised based on the number of reads mapped.

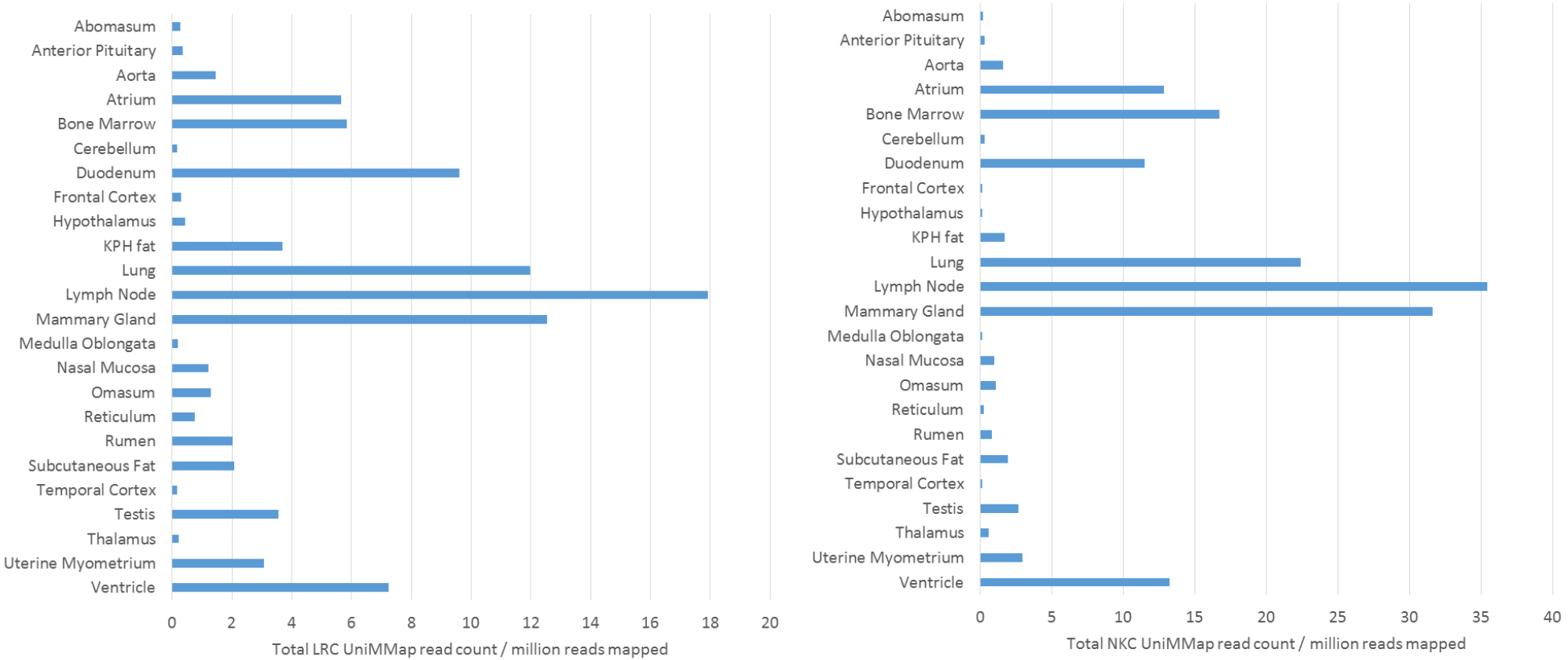


Figure 4-18. Comparison of total LRC/NKC gene transcription in 24 tissue types. RNA-Seq data was downloaded from the sequence read archive (Accession: PRJNA379574). The BioProject contains RNA-Seq from 23 Dominette tissues as well as the testis of SuperBull 99375. Read counts from UniMMap were normalised based on the number of reads mapped.

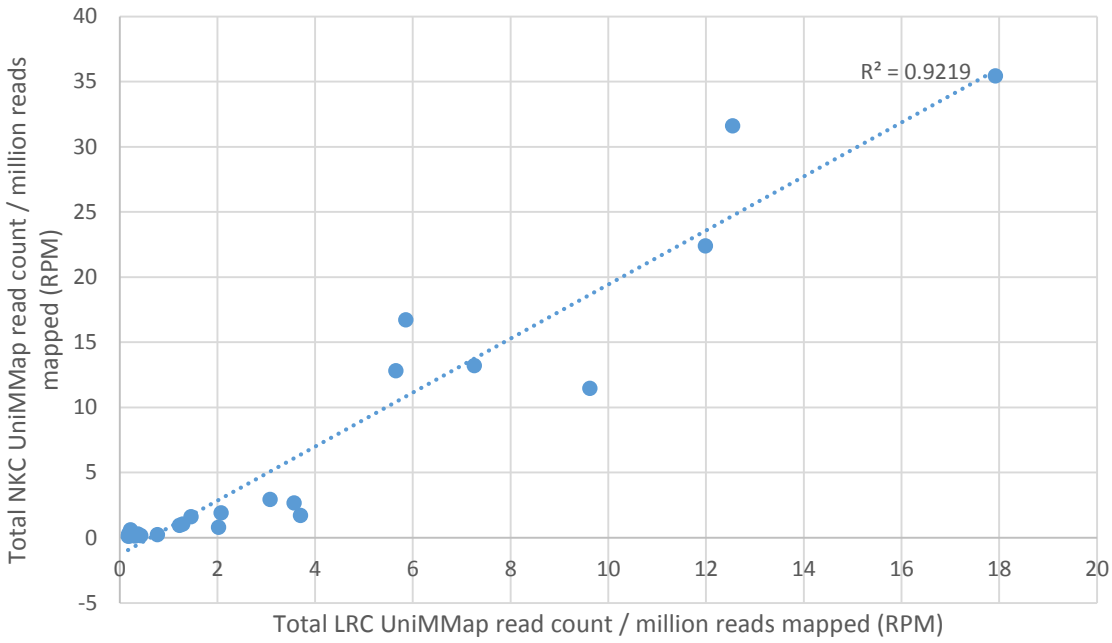


Figure 4-19. Scatterplot of total LRC/NKC gene transcription in 24 tissue types. RNA-Seq data was downloaded from the SRA (Accession: PRJNA379574). The BioProject contains RNA-Seq from 23 Dominette tissues as well as the testis of SuperBull 99375. Read counts from UniMMap were normalised based on the number of reads mapped.

4.4 Discussion

Additional NK RNA-Seq datasets generated using a different NK isolation method were analysed and compared with the two NK populations previously examined. The extent to which the LRC/NKC genes are transcribed on multiple different immune cell population was determined using UniMMap. Publically available datasets were utilised to attempt to understand the role of LRC/NKC genes in *M. bovis* infection, as well as for the generation of an expression atlas across multiple tissue types. RNA-Seq data from PBMCs isolated from dam and calf was used to assess the relationship of LRC/NKC transcription and age. The development of the repertoire over the month post-birth was also characterised. These additional datasets enabled identification of genes that are consistently transcribed in all individuals.

4.4.1 Variation in LRC/NKC gene transcription correlates with the method of isolation

The method used to isolate NK cells appears to correlate with the pattern of transcription of both the NKC and LRC genes. The first method compared was a positive selection of NCR1+ cells using MACS beads. The second method was a FACS protocol involving negative sorting of CD4+ T cells, monocytes and B cells, followed by a positive sort for NCR1+ cells. The FACS method would likely have yielded a purer final population due to the removal of multiple potentially contaminating cell types. The downside to the FACS method is that the cells will likely have spent longer with an antibody bound to the cattle NK cell marker NCR1. As NCR1 is an activating receptor, the longer period of time taken for isolation when using a combination of MACS/FACS will most likely have resulted in a larger shift in transcription from *in vivo* cells.

Within the LRC genes, the longer period of activation appears to result in an increase in transcription of all of the inhibitory *KIR*. The majority of the activating *KIR* are transcribed at a comparable level in all animals, with the exception of *2DS1* and *3DXS1*. This increase in inhibitory *KIR* transcription could be part of a mechanism controlling NK cell activation. Although a different activation method, exposure of humans NK cells to IL-2 has been shown to reduce transcription of the activating receptor *2DS4* and increase transcription of the inhibitory *2DL4* (Küçük et al. 2016).

Higher transcription of the entire region of the NKC between *KLRI2* and *KLRE* is observed in the FACS isolated samples. This is most likely caused by the purer population of NK cells obtained through FACS or activation via NCR1. Analysis of transcription of this region on other immune cell types supports this, with the exception of *KLRK,* these genes are transcribed at a much higher rate on NK cells than any other cell type. Direct comparisons between the two isolation methods are complicated by variation between animals. A number of environmental factors differ between the animals from either herd, all of which could contribute to variation in transcriptional level, such as housing conditions, diet and level of activity. Experimental conditions also differed, potentially further contributing to the observed variation, Different methodologies for RNA-Seq isolation, library preparation and sequencing will have introduced unique biases to the data from either set of animals. However, comparisons between the two sets of animals provides information towards designing future experiments, as well as on the frequency of presence/absence of these genes. A comparison of NK cells from the same animal, isolated by the two methods, is required to fully understand the impact of isolation method.

4.4.2 Transcription of the LRC/NKC genes is not limited to NK cells

Transcription of *KIR* occurs at a similar level in CD8+ T cells to NK cells. Lower level transcription of all *KIR* was also observed in gamma delta T cells. CD4+ T cells appear to transcribe a limited repertoire of *KIR* at a level lower than observed in gamma delta T cells. Work in humans has shown that there is a subset of CD8+ T cells that also express KIR. However the pattern of KIR expression in CD8+ T cells is different to that of NK cells (Bjorkstrom et al. 2012). Almost 90% of CD8+ T cells that express KIR, express just one inhibitory KIR. KIR expression is also dominated by a single KIR in human CD8+ T cells, on average 68% of total expression originated from a single KIR in each individual, higher than observed in NK cells. They also found no correlation between expression of KIR on NK cells and expression on CD8+ T cells in the same individual. When analysing expression of activating KIR on CD8+ T cells, they found that over 50% of KIR2DS4+ cells did no co-express an inhibitory KIR. No correlation is seen between transcription of the *KIR* in NK cells and CD8+ T cells in our cattle datasets (R^2 value 0.2127).

With a few exceptions, transcription of the NKC genes is much higher in NK cells than the other immune cell types examined. However, a small amount of transcription of the majority of the NKC was observed in all cell types. The region between *KLRI2* and *KLRE* for which high transcription is consistently observed, is largely only highly active in NK cells. The exception to this is *KLRK* which is transcribed at a similar level in CD8+ T cells. Expression of KLRK has been observed cattle CD8+ T cells (Guzman, Birch, and Ellis 2010). Transcription of *KLRJ* in gamma delta T cells occurs at a level higher than any other tested cell type. *KLRJ* is typically regarded to be an NK receptor, however a study by Boysen, et al. (2006) found that transcription was low in NK cells, requiring 30 PCR cycles to detect. We show *KLRJ* to be transcribed at a similar level in NK cells to all other cell types with the exception of gamma delta T cells, where transcription is much higher. This suggests KLRJ should be considered a gamma delta T cell receptor rather than a NK receptor. Lacking either an activating or inhibitory component, *KLRJ* potentially forms a heterodimer with an unknown partner, as is observed with *KLRD* and *KLRE* (Schwartz et al. 2017). The only other *KLR* gene to be transcribed higher in gamma delta T cells than CD4+ T cells, B cells or monocytes is *KLRK*. This allows us to hypothesis that if *KLRJ* indeed does form a heterodimer, then *KLRK* is the partner.

4.4.3 Transcription of activating *KIR* is lower in *M. bovis* infected animals

Natural killer cells have been shown to respond to infection with *Mycobacterium bovis* in humans and cattle (Vankayalapati et al. 2002; Denis et al. 2007). The *KIR* haplotype of an individual has been shown to play an important role in the level of response to mycobacteria in humans (Portevin et al. 2012). They show that the presence of the activating *KIR*, *2DS3* and *2DS5*, was significantly over-represented in the high responder group. Transcription of *KIR* was much lower in the infected population compared to the control. However, two inhibitory receptors, *3DXL7* and *2DL1*, were transcribed more highly in the infected group. The activating *KIR* are all transcribed at a lower rate in the infected population. In humans, the inhibitory genes *KIR3DL1* and *KIR2DL3*, as well as the activating genes *KIR2DS1* and *KIR2DS5* confer susceptibility towards *Mycobacterium tuberculosis* (Pydi et al. 2013). This highlights the role of KIR in *mycobacterium* infections. As the control and infected animals are not related and the *KIR* genotypes of the animals are unknown, it is not possible to determine if

infection with *M. bovis* causes a downregulation of *KIR*. However the level of activating *KIR* transcription may correlate with infection status. As the data is from PBMCs and multiple animals, it is not possible to infer information on clonal expansion of NK cells. However a future experiment with pure NK cell populations would allow investigation of *KIR* transcription during *M. bovis* infection in greater resolution.

4.4.4 LRC genes are variable in transcriptional status

RNA-Seq analysis of PBMCs from six cattle shows variation in transcriptional status of multiple *KIR*. Variability of *KIR* transcription in cattle has been previously demonstrated (Allan et al. 2015). They observed absence of transcription of *3DXL6* as well as the *3DXL3/5/7* gene family. In comparison, we observed the absence of *3DXL5* and *3DXL7*, as well as *3DXL1*, *3DXL2* and *3DXL4*. Transcription of *3DXL6* was consistently seen in our analysed data. Notably, they showed transcription of cattle *KIR* can be rescued with cytokine stimulation. This suggests that the absence of *KIR* in an *ex vivo* dataset is not an accurate indicator of the functionality of a *KIR* in an individual. Accurate determination of transcriptional status is further confused by the low number of *KIR* reads present in PBMC RNA-Seq datasets. With the very low read counts observed in these six animals (most often less than one read per million), it becomes difficult to distinguish between absence of transcription and a read count too low to detect. Cytokine stimulation of PBMCs may provide a method of increasing transcription and subsequently read counts in RNA-Seq data. Although with the downside of shifting the transcriptional profile of the cells further away from their *in vivo* status. However, haplotypic variation is expected and absence of transcription may be caused by variability in gene content between haplotypes. The repeated detection of transcription of the genes encoding null-alleles in the reference haplotype is particularly interesting. Only one of these genes (*2DS2*) is absent in one of the six cattle. This finding reinforces the possibility that these genes either have functional alleles or produce an alternate protein.

4.4.5 LRC/NKC genes undergo a large increase in transcription post-birth

Transcription of the LRC, and to a larger extent the NKC, is much higher at 28 days post-birth than any other previous time point. An approximately 5x higher amount of transcription observed in the NKC at day 28 compared to day 14.

Despite this, the proportion of transcription of each gene remains highly similar. The frequency of NCR1+ NK cells has been shown to be at its highest at day 0 post-birth before dropping drastically at day 1. The frequency then rises until day 14 where it remains relatively constant into adulthood, but does not return to the day 0 level (Graham et al. 2009). A reason for this phenomenon has not been discovered. Graham et al. (2009) propose that it may be due to distribution of NK cells to tissues and organs. Massively increased transcription of NK receptors, including NCR1, at day 28 suggests that NK cells may undergo a period of altered phenotype for a period of time post-birth. This is then followed by a return to a more typical NK receptor phenotype.

4.4.6 Transcription of *KLR*/*KIR* is ubiquitous but varies in amount

Unsurprisingly, transcription of both the NKC and LRC is relatively high in the bone marrow, the site of NK development, and the lymph node, a major immune tissue. Relatively high transcription was also observed in lung tissue. The lungs are a major site for pathogens to enter the body and subsequently, NK cells make up 10% of the lymphocyte population in humans (Grégoire et al. 2007). Particularly high transcription of the LRC/NKC was observed in the mammary gland tissue. NK cells have recently shown to be present in the mammary gland of cattle and have been implicated in mastitis pathogenesis (Sipka et al. 2016).

The LRC/NKC genes were also heavily transcribed in various heart tissue, including the atrium, ventricle and to a lesser extent the aorta. NK cells have been implicated in a number of protective roles in the heart of both humans and mice. They are required for protection against both *Coxsackie B virus* (CBV) and *Mouse cytomegalovirus* (MCMV) induced myocarditis (Kanda et al. 2000) (Barin et al. 2013). NK cells have also been shown to restrict the accumulation of eosinophils, limiting the pathogenesis of myocarditis (Ong et al. 2015).

Cattle major histocompatibility complex class I chain-related (MIC) is constitutively expressed on cattle intestinal epithelium and is the ligand for KLRK (NKG2D) (Guzman, Birch, and Ellis 2010). Transcription of *KLRK* was higher than any other gene in either the NKC or LRC in the duodenum. KLRK has been shown to be expressed on NK cells, CD8+ T cells and to a lesser extent gamma delta T cells. This highlights its importance as an activating receptor.

4.4.7 Conclusions

Through UniMMap analysis of various cattle RNA-Seq datasets, a number of important questions have been addressed. Transcription of LRC and NKC genes is not confined to NK cells, an equivalent level of transcription is observed in CD8+ T cells. Transcription of *KLRI2* is limited almost entirely to NK cells and if expressed on the cell surface, represents a potential additional NK marker. UniMMap analysis of publically available datasets from a study of *M. bovis* infection was used to provide insight into the impact of infection on transcription. The transcription of *KIR* appear to be influenced by *M. bovis* infection, particularly *3DXL7* and *2DL1* which are transcribed at a higher level in the infected group. Analysis of three dams reveals that the transcription of LRC/NKC genes undergoes a large increase in the 28 days post-birth. This includes the entire repertoire of genes and is not limited to a subset. Transcription was also assessed in a number of tissue types, revealing that at least one LRC and NKC gene is transcribed in all examined cell types. This provides evidence as to the distribution of NK cells throughout cattle tissue.

Chapter 5. Gene expression atlas for LRC and NKC genes in goats

5.1 Introduction

Goats (*Capra hircus*) are an important livestock species farmed for their meat, milk and skin. Able to subsist on a diet of poor quality vegetation, they are often famed on land unsuitable for other livestock species. Subsequently, goat farming is particularly prevalent in Africa, Asia and the Middle East and they are the most important livestock species in many areas of these regions (Gutierrez-A 2018). The recent release of a high quality reference assembly of the goat genome, coupled with the characterisation of the goat LRC and NKC (Bickhart et al. 2017; Schwartz et al. 2017; Schwartz, Sanderson, and Hammond, unpublished data) and our analysis pipeline UniMMap, allowed us to analyse transcription of these immune-related regions.

The NKC of the goat has expanded in a similar manner to cattle (figure 5-1B). Both cattle and goats have expanded a novel region of *KLRC*-like and *KLRH*-like genes between *KLRA* and *KLRJ*. This expansion has occurred to a lesser extent in goats compared to cattle, cattle possess 16 of these genes whilst goats possess 6. Within this region in both cattle and goats, there is at least one *KLRH*-like gene possessing a *KLRC2*-like activating tail. The presence of this expansion and *KLRH*/*KLRC2* recombination in both species suggests that it occurred before the two species diverged ~30 million years ago (mya) (Hiendleder et al. 1998; Schwartz et al. 2017). Also like cattle, goats have expanded the *3DX*-lineage of *KIR* (figure 5-1A). This expansion differs from cattle however as many of the goat genes do not clade with the cattle *KIR* groups, rather forming their own distinct groups, and they not possess a functional *3DL* lineage gene. Of the 15 *KIR* located within the goat LRC, 7 are putatively functional (Schwartz, Sanderson, and Hammond, unpublished data). Analysis of goat gene expression data deposited in the sequence read archive (SRA) with UniMMap allowed the first insight into transcription of genes within these regions.

Peste des petits ruminants virus (PPRV) causes severe disease in goats and sheep. The mortality rate of the virus can reach 70% or higher dependent on both the virus strain and health of the infected animals. It is distributed across

much of Africa, Asia and the Middle East where it consequently has a significant economic impact (Diallo et al. 2007). PPRV has two cellular receptors, the first of which is signalling lymphocyte activation molecule (SLAM). Expression of SLAM is exclusive to immune cells and is present on leukocytes, macrophages and dendritic cells. The second is Nectin-4, expression of which occurs on epithelial cells but not lymphocytes or dendritic cells (Kumar et al. 2014). PPRV has been shown to inhibit proliferation of peripheral blood leukocytes in both goats and cattle (Heaney, Barrett, and Cosby 2002). PBMCs infected with PPRV produce higher levels of interferon gamma (IFNγ) than non-infected PBMCs. Along with CD8 T cells and gamma delta T cells, NK cells are a major producer of IFNγ, suggesting they may respond to PPRV infection (Dhanasekaran et al. 2014). Although the role of NK cells in PPRV infection is poorly understood, these three cell types are now known to transcribe NK cell receptors in cattle (chapter 3). RNA-Seq data generated from PBMCs isolated from goats involved in a PPRV vaccination study was used to characterise LRC and NKC genes in response to PPRV as part of the wider immune response. It also provided the opportunity to demonstrate the utility of UniMMap in a different ruminant species.

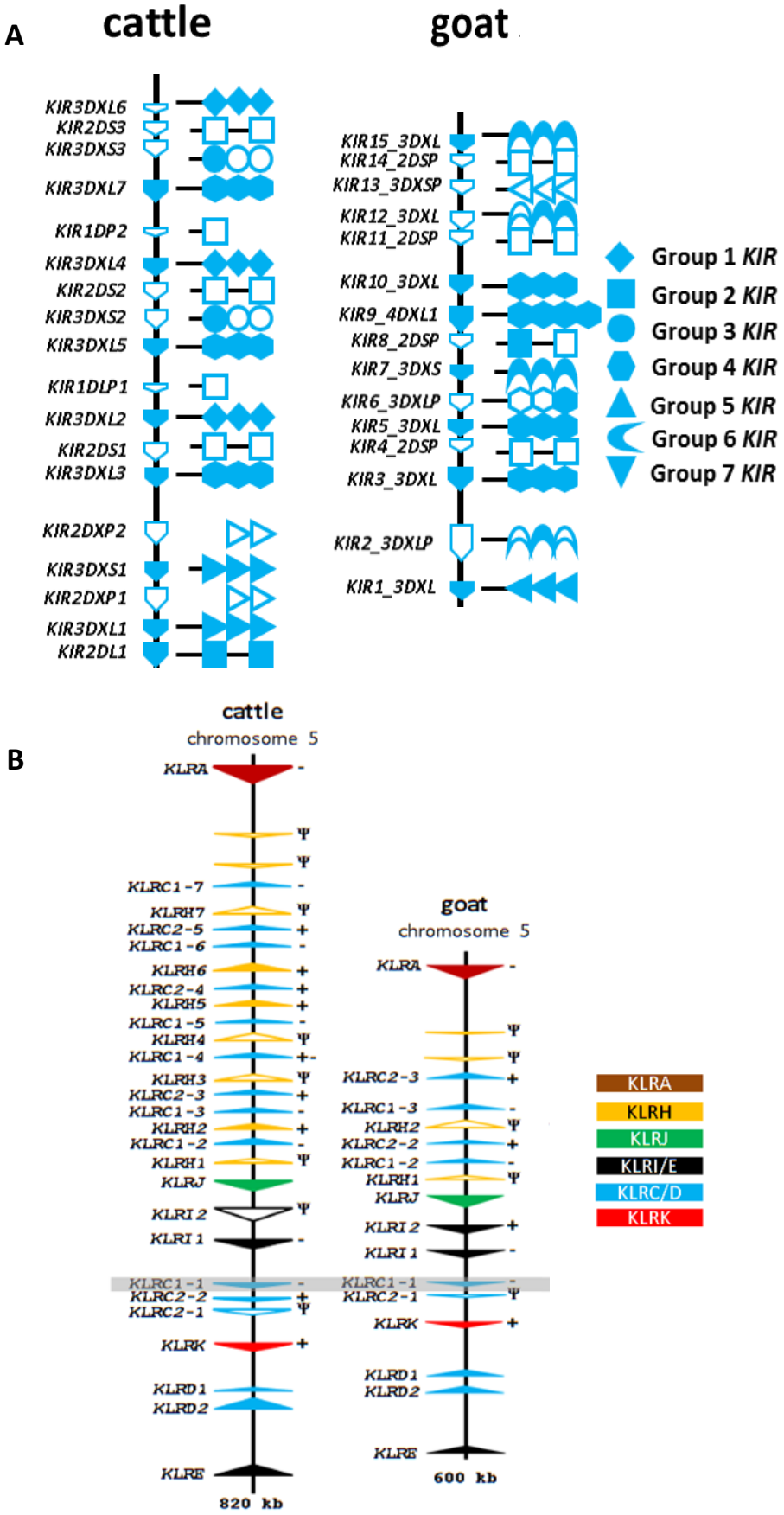


Figure 5-1. The gene structure of the cattle and goat LRC (A) and NKC (B). Arrows indicate individual genes. Open arrows represent predicted non-functional and filled represent predicted functional genes. The orientation of the arrow indicates the direction of transcription. Figure A is adapted from Schwartz et al. (manuscript in preparation) and figure B is taken and modified from (Schwartz et al. 2017).

5.2 Methods

5.2.1 Goat LRC/NKC gene atlas data acquisition

RNA-Seq datasets from BioProject accession number PRJEB23196 (Bush et al. 2019) were downloaded from the sequence read archive (SRA). Sequencing was carried out on an Illumina HiSeq 4000, generating 75bp paired-end reads. The datasets consist of 17 different goat tissues and 3 goat immune cell populations. The tissues and cell types were obtained from varying numbers of individuals, resulting in 54 total datasets.

5.2.2 Goat LRC/NKC gene atlas analysis

The coding sequences of all of the NKC and LRC genes were extracted from the goat genome assembly ARS1 (Bickhart et al. 2017). The 54 RNA-Seq datasets were individually mapped to the coding sequences using the UniMMap pipeline (as described in chapter 2). UniMMap read counts were normalised based on the number of reads mapped for each sample. Normalised reads counts were averaged to generate a single count for each gene of the NKC and LRC in each of the 17 tissue types and 3 cell types. A heatmap was then generated from the $\log_{10}$ values of the counts using the superheat package in R (Barter and Yu 2017). Kendall rank correlation was carried out using the cor function from the stats package in R.

5.2.3 PPRV vaccination study sample preparation

Three animals were taken from a PPRV vaccination study being carried out at the Pirbright Institute (Karin Darpel, Orbivirus Research). All animal experiments were approved by The Pirbright Institute Ethics Committee and carried out in accordance with the U.K. Animal (Scientific Procedures) Act 1986. This work was carried out under home office license number PPL70/8833 ‘PPRV pathogenesis and immune response to PPR vaccines’. With the exception of the control (unvaccinated) animals, goats were inoculated with an experimental PPRV vaccine. After 28 days post-vaccination, all goats (including unvaccinated controls) were challenged with PPRV. Whole blood was taken from three of the vaccinated animals and two of the unvaccinated animals 14 days post-vaccination, 2 days post-challenge (28 days post-vaccination) and 8 days post-challenge. Due to the severity of infection that developed in one of the control

animals, it had to be euthanised before the final time point. PBMCs were isolated from whole blood using a histopaque-1083 gradient and suspended in Trizol. Total RNA was extracted using an Invitrogen PureLink RNA Mini Kit and quantified using a Qubit Fluorometer. RNA integrity was checked with an Agilent Bioanalyser 2100 and samples with a RIN score equal to or greater than 9.0 taken forward for sequencing. Sequencing was carried out on an Illumina NextSeq 550, producing ~50 million 150bp single-end reads per sample.

5.2.4 PPRV vaccination study analysis

LRC/NKC cDNA sequences were extracted from the *Capra hircus* ARS1 genome assembly based on an in-house annotation. Read counts for the LRC/NKC genes were generated using the UniMMap pipeline, which is extensively described in chapter 2. Read counts were normalised based on the number of reads mapped for each sample. Streamgraphs of gene proportion at each time point in each animal were generated using the streamgraph R package (Rudis 2015). Read counts for whole transcriptome analysis were generated by mapping the RNA-Seq datasets to the goat ARS1 genome and ARS1.93 annotation using the Gemtools RNA-pipeline. Differential expression between unvaccinated and vaccinated animals was carried out using edgeR (Robinson, McCarthy, and Smyth 2010). For analysis with edgeR, genes expressed in 2 or more samples were selected. Differentially expressed genes with an FDR-adjusted (Benjamini-Hochberg) p-value equal to or less than 0.05 were carried forward for further analysis. GO term analysis was carried out by importing selected gene identifiers into Ensembl BioMart (Kinsella et al. 2011) and filtered using awk.

5.3 Results

5.3.1 Large variance of transcription within the LRC/NKC genes occurs between goat tissue types

RNA-Seq datasets used to generate a goat gene expression atlas were accessed to enable a comparison of transcription of LRC/NKC genes between different tissue types. Within the gene atlas, 17 tissue types and 3 immune cell types were sequenced and all were subsequently analysed with UniMMap.

The transcriptional patterns of the LRC/NKC genes are highly variable between tissue/cell types (figure 5-2). Within the LRC, *1_KIR3DXL* and *3_KIR3DXL* are transcribed in all of the analysed datasets. Of the NKC genes, only *KLRE1* is transcribed ubiquitously. Transcription of all examined LRC/NKC genes occurs in both the ileum and the spleen. Almost all of the genes are transcribed in the testis, with the exception of *KLRJ1*. Transcription of *KLRJ1* is observed elsewhere however, present in 5 of the 20 total tissue/cell types. Transcription of *1_KIR3DXL* is higher in alveolar macrophages than any other LRC/NKC gene in any of the tissue/cell types. Bone marrow (BM) macrophages exhibited the second-highest level of transcription of *1_KIR3DXL*, both when unstimulated and stimulated with lipopolysaccharide (LPS) for 7 days. There is a small amount of variation between stimulated and unstimulated bone marrow macrophages. Transcription of *3_KIR3DXL* is only observed in the stimulated population and *KLRC1-1*, *KLRK1* and *KLRD2* are transcribed only in the unstimulated population.

Despite transcribing only 8 of the 15 *KIR*, alveolar macrophages have the highest total *KIR* transcription (figure 5-3A). The spleen, in which the second highest level of *KIR* transcription occurs and which transcribes all 15 *KIR* genes, has a total *KIR* read count almost half that of alveolar macrophages. Total *KIR* transcription is also noticeably elevated in both BM macrophage populations as well as the testis. Comparatively, the spleen (53.75 RPM) is the highest total transcriber of NKC genes and the testis (19.84 RPM) in the second highest (figure 5-3B). Transcription of the NK marker *NCR1* is highest in the spleen (3.22 RPM) (figure 5-3C), which is 11.41x higher than the average across all tissue/cell types (0.28 RPM). Comparing the ratio of *KIR* transcription to *NCR1* transcription

Figure 5-2. Heatmap of LRC/NKC gene transcription in various goat tissues. Read counts for each gene in each tissue were generated by UniMMap. Read counts were normalised per million reads mapped. Normalised read counts were transformed to their $\log_{10}$ values and a heatmap generated using the superheat R package.

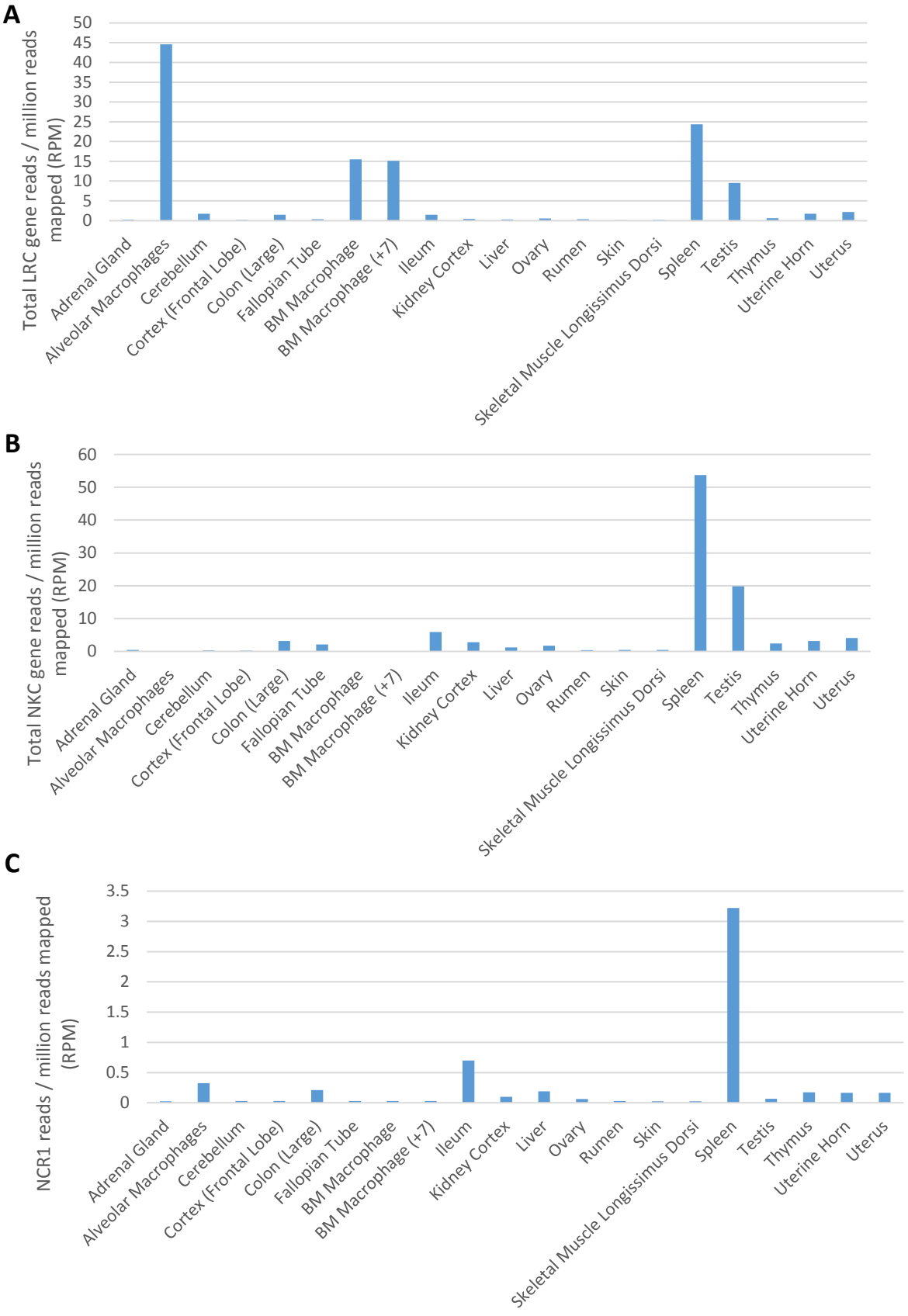


Figure 5-3. Total LRC gene (A) NKC gene (B) NCR1 (C) transcription in each tissue. Read counts for each gene in each tissue were generated by UniMMap. Read counts were normalised per million reads mapped and total transcription calculated.

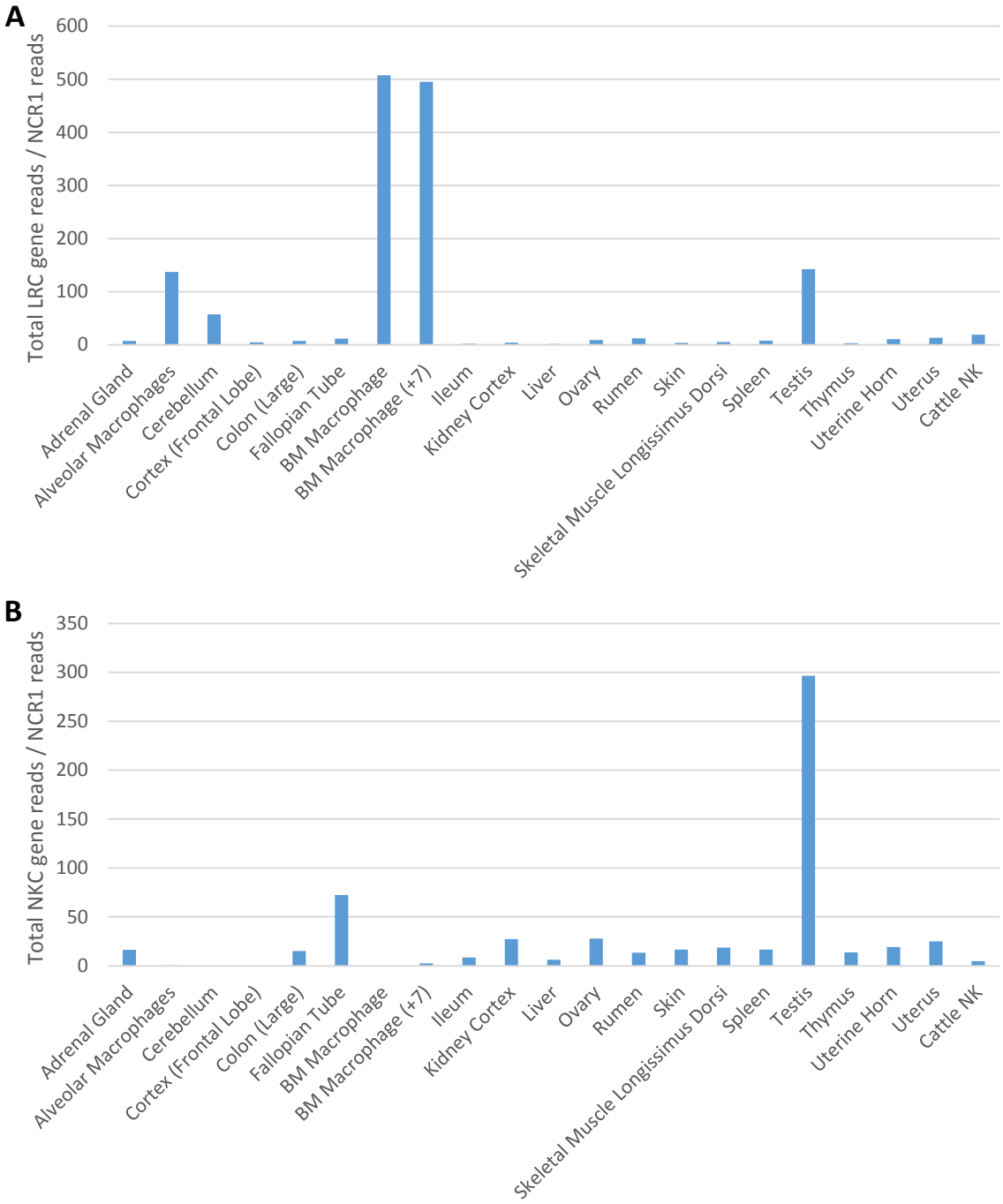


Figure 5-4. Total LRC gene (A) NKC gene (B) transcription relative to NCR1 transcription in each tissue. Read counts for each gene in each tissue were generated by UniMMap. Read counts were normalised per million reads mapped and total transcription calculated. Total read counts for each tissue were divided by the NCR1 read count of that tissue.

provides an estimate of the contribution of NK cells to the total *KIR* transcription (figure 5-4A). The higher the value, the less likely it is that the *KIR* transcription is occurring in NK cells. As a reference point, the ratio observed in cattle NK cells is 18.83 *KIR* reads/*NCR1* read. Both BM macrophage populations have the highest ratio (~500 *KIR* reads/*NCR1* read) and alveolar macrophages are also elevated (136.92 *KIR* reads/*NCR1* read). High values are also observed for the testis (142.55 RPM) and cerebellum (52.26 *KIR* reads/*NCR1* read). The same comparison for total NKC gene transcription again shows an elevated ratio for testis tissue (296.34 NKC gene reads/*NCR1* read) (figure 5-4B) which is much higher than the value for cattle NK cells (4.88 NKC gene reads/*NCR1* read). The value for fallopian tube tissue was also elevated (72.50 NKC reads/*NCR1* read). The Kendall rank correlation is used to measure the association between two quantities. The Kendall rank correlation coefficient for total NKC and *NCR1* transcription is 0.478 (*p-value* = 7.231×10^{-7}) and for total *KIR* gene and *NCR1* is 0.347 (*p-value* = 3.264×10^{-4}). This supports the hypothesis that NK cells account for more NKC gene transcription than *KIR* transcription. This supports that these genes are transcribed in cell types other than NK cells, as observed in cattle (chapter 4).

Transcription of all the of the LRC/NKC genes is observed in at least one tissue/cell type. However, the pattern of transcription is highly variable between the tissue/cell types, transcription of just three of the 31 genes was consistently observed. The high level of transcription of LRC/NKC genes relative to *NCR1* transcription in testis tissue is particularly striking. Combined with the comparatively high transcription of the genes of both the LRC and NKC, hints at a previously unseen *NCR1*/LRC/NKC cell phenotype.

5.3.2 Variation in LRC/NKC gene transcription occurs primarily between individuals rather than vaccination status

To investigate changes in transcription of the LRC/NKC genes in response to viral infection, samples were collected from a PPRV vaccination study carried out in goats. RNA was extracted and sequenced from PBMCs isolated from five individuals (three PPRV vaccinated and two unvaccinated controls) at three separate time points. The first time point (-14) is 14 days post-vaccination and 14 days pre-PPRV challenge, the second (+2) is 2 days post-challenge with PPRV, and the third (+8) is 8 days post-challenge with PPRV.

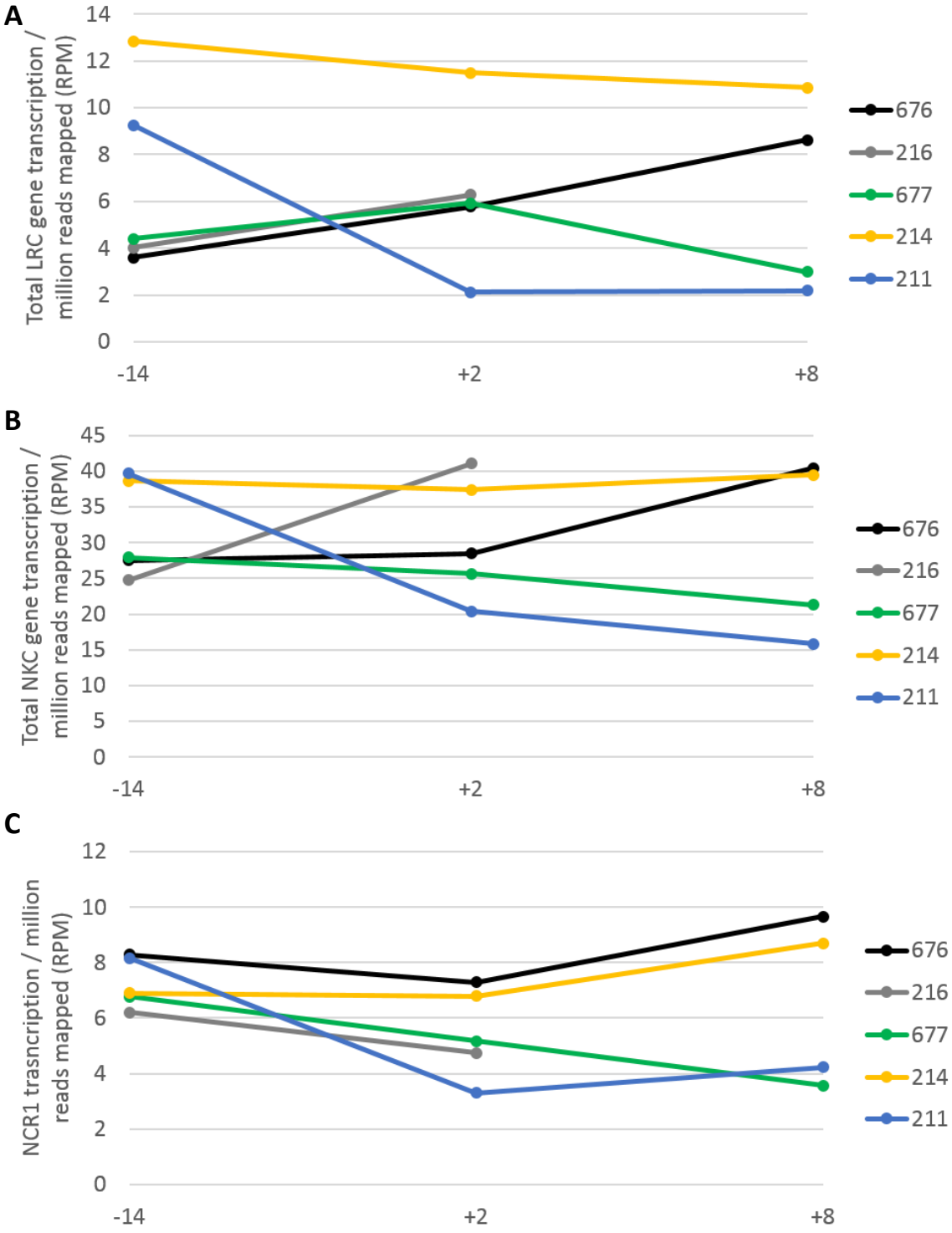


Figure 5-5. Total LRC gene (A) NKC gene (B) *NCR1* (C) transcription over the course of a PPRV vaccination study. Read counts were generated by UniMMap, normalised per million reads mapped and total transcription calculated. Animals 676 and 216 are unvaccinated controls and 677, 214 and 211 were inoculated with an experimental PPRV vaccine. The first time point (-14) is 14 days post-vaccination and 14 days pre-PPRV challenge, the second time point (+2) is 2 days post-PPRV challenge, and the third time point (+8) is 8 days post-PPRV challenge.

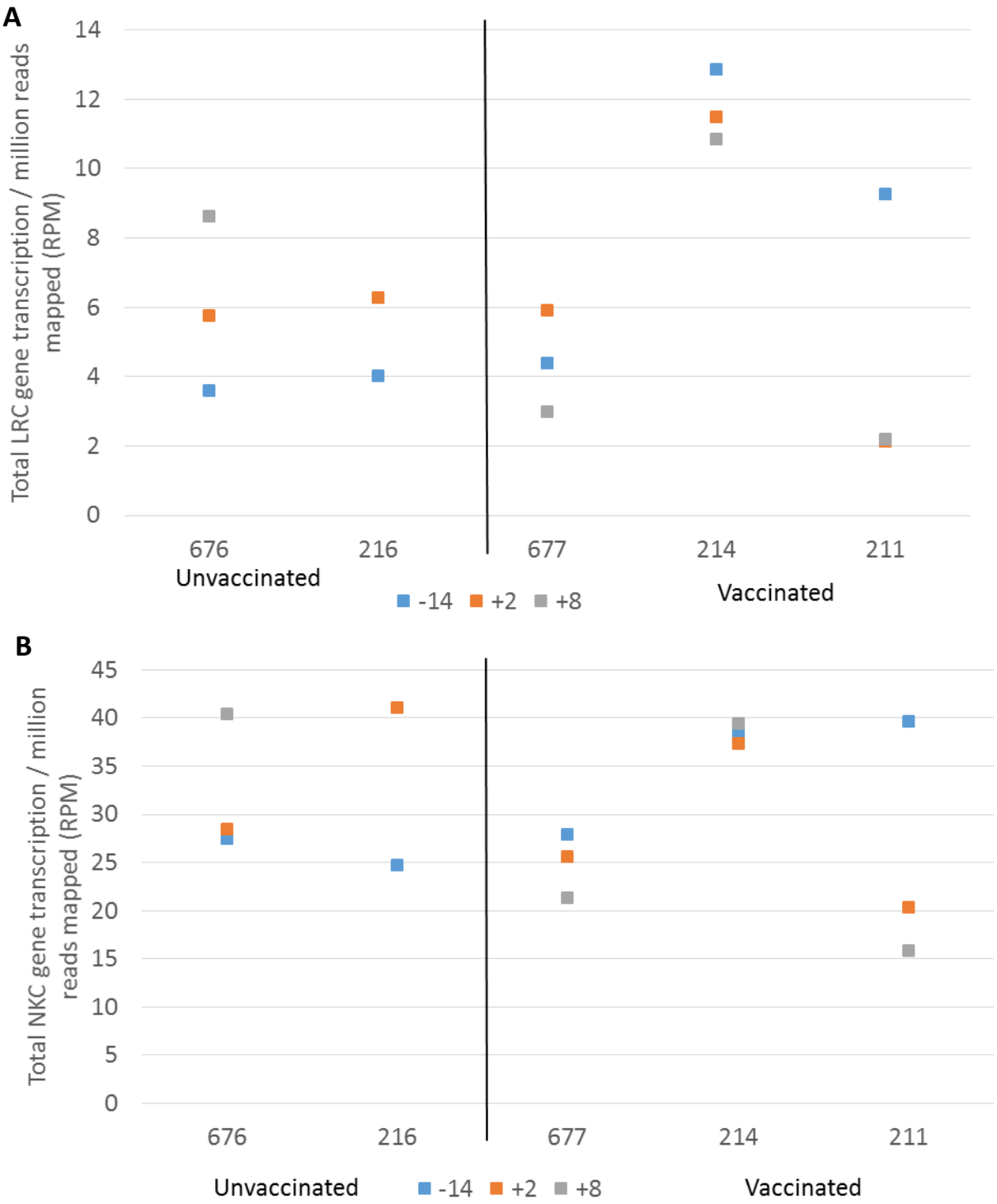


Figure 5-6. Total LRC gene (A) NKC gene (B) transcription over the course of a PPRV vaccination study. Read counts were generated by UniMMap, normalised per million reads mapped and total transcription calculated. Animals 676 and 216 are unvaccinated controls and 677, 214 and 211 were inoculated with an experimental PPRV vaccine. The first time point (-14) is 14 days post-vaccination and 14 days pre-PPRV challenge, the second time point (+2) is 2 days post-PPRV challenge, and the third time point (+8) is 8 days post-PPRV challenge.

Total transcription of LRC/NKC genes was calculated to compare changes across the three time points (figures 5-5, 5-6). Although comparisons between the two unvaccinated animals (676 and 216) are not possible at the final time point due to the euthanization of 216 prior to the time point, comparisons can be made between the first two time points. The two unvaccinated animals have an identical change in LRC gene transcription between -14 and +2 (figure 5-5A, 5-6A). Both animals again have similar transcription at -14 (3.60 and 4.02 RPM for 676 and 216 respectively). Transcription of LRC gene increases in both animals by the +2 time point (increasing by 2.17 and 2.27 RPM in 676 and 216 respectively). Transcription of LRC genes increases an additional 2.38 RPM by +8 days post-challenge in 676. Both animals have similar total transcription of NKC genes (27.47 RPM and 24.71 RPM for 676 and 216 respectively) at the initial time point, pre-PPRV challenge (figure 5-5B, 5-6B). At the first time point post-challenge, total transcription for 676 remains almost identical (27.47 to 28.43 RPM), whereas 216 increases from 24.71 to 41.05 RPM. A highly similar increase occurs at the next time point for 676, from 28.43 RPM at +2 to 40.40 RPM at +8.

There is much more variability in the response of LRC/NKC gene transcription within the three vaccinated animals (677, 214 and 211). Animals 214 and 211 have almost identical total NKC gene transcription at the first time point (38.62 and 39.66 RPM respectively) (figure 5-5B, 5-6B). At the final time point, the total NKC gene count for 214 has increased slightly (38.62 to 39.45 RPM) and has dropped considerably for animal 211 (39.66 to 15.82 RPM). The starting point for animal 677 is the lowest of the vaccinated animals, and while it has decreased by the final time point, it is still higher than for animal 211 (21.30 compared to 15.82 RPM). The overall trend for total LRC gene transcription within the vaccinated animals is similar to that of the NKC genes (figure 5-5A, 5-6A). At the initial time point, 214 and 211 have higher total LRC gene transcription (12.85 and 9.26 RPM respectively) than the other three animals. The vaccinated animal 677 (4.40 RPM) and the two unvaccinated animals, 676 (3.60 RPM) and 216 (4.02 RPM) group closely together. The response to challenge with PPRV again differs considerably between 214 and 211. Total LRC gene transcription for 214 drops slightly the three time points, decreasing by 1.36 RPM by the second time point and by 0.64 RPM by the third. The decrease is much higher for 211, dropping by 7.14 RPM between the first and second time points, and increasing by 0.065 RPM

by the third time point. The trend for animal 677 closely matches the unvaccinated animals across the first two time points, before decreasing below the starting value by the final time point.

Transcription of the NK marker NCR1 was also compared between the five animals (figure 5-5C). The changes to *NCR1* transcription within each animal are similar to those observed for total NKC gene transcription. The major exception to this is animal 216 in which *NCR1* transcription decreases between the first two time points, total NKC gene transcription undergoes a large increase over this same time period.

More variability is observed between the individual animals, particularly within the vaccinated group, rather than based on vaccination status. The overall trends for total LRC/NKC gene transcription as well as *NCR1* transcription are relatively consistent within each individual.

5.3.3 Average counts reveal variability between unvaccinated and vaccinated animals

To reduce the effects of individual variability when trying to compare the unvaccinated and vaccinated groups, read counts within the groups were averaged (figure 5-7). Two distinct trends can be seen between the groups. Total *KIR* transcription is at its lowest at the first time point in the unvaccinated animals and increases up to the final time point (figure 5-7A). In the vaccinated animals, the opposite occurs and total *KIR* transcription decreases over the three time points. Grouping the *KIR* based on their activating or inhibitory status highlights the inverse response between unvaccinated and vaccinated animals (figure 5-7C). The same trend is observed when comparing total transcription of the NKC genes between unvaccinated and vaccinated animals (figure 5-7B).

5.3.4 Changes in total transcription are driven by the entire repertoire of genes

To understand what changes in transcription were occurring at the individual gene level, the proportion of the total transcription that each gene represents was plotted over the three time points for each individual. Despite the large changes in total *KIR* transcription observed in multiple animals over time, there is only minor variation in the proportion of each *KIR* (figure 5-8). Transcription of

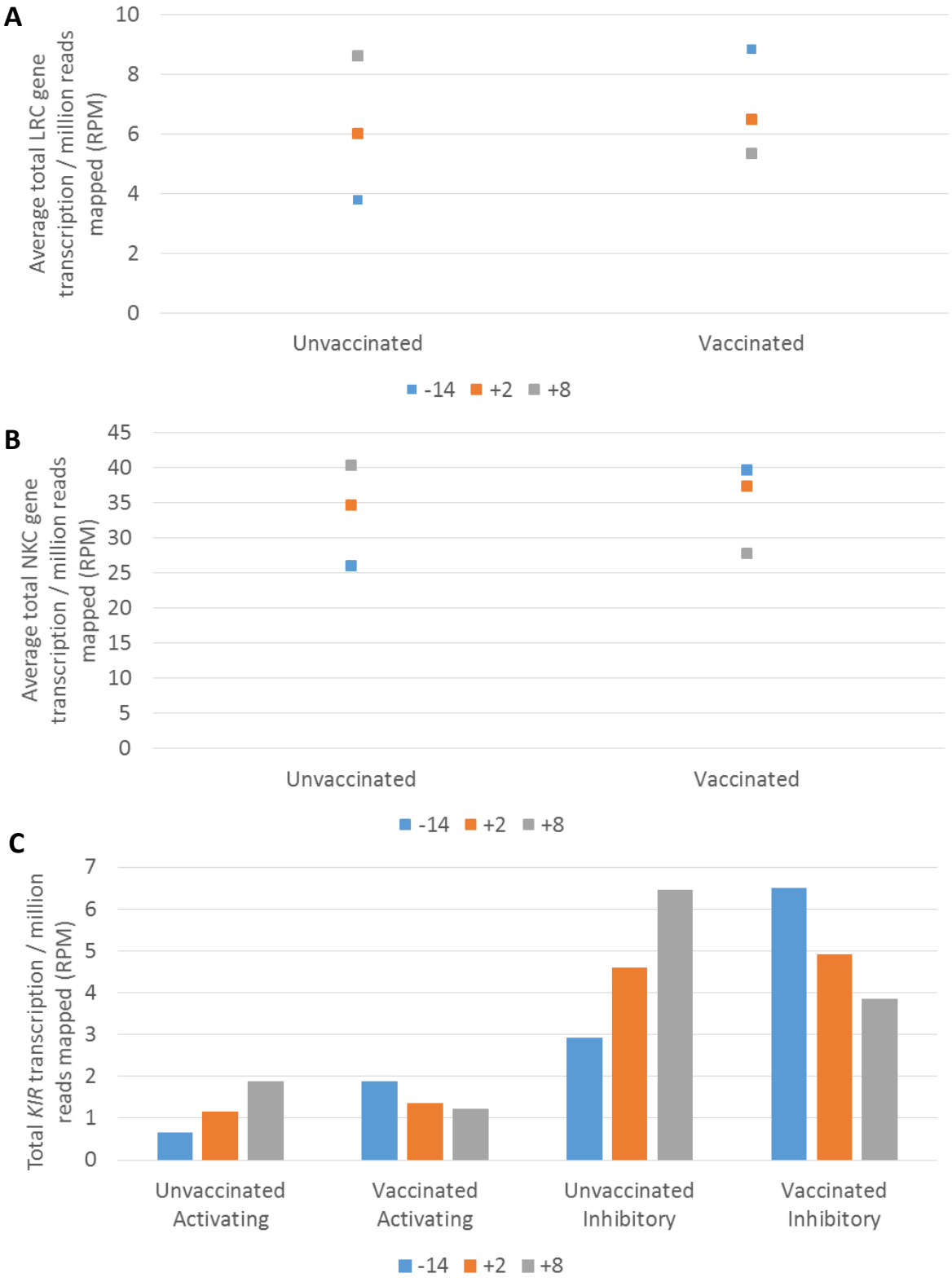


Figure 5-7. Average total LRC gene (A) NKC gene (B) activating and inhibitory *KIR* (C) transcription in unvaccinated and vaccinated animals. Read counts were generated by UniMMap, normalised per million reads mapped and total transcription calculated. Animals 676 and 216 are unvaccinated controls and 677, 214 and 211 were inoculated with an experimental PPRV vaccine. The first time point (-14) is 14 days post-vaccination and 14 days pre-PPRV challenge, the second time point (+2) is 2 days post-PPRV challenge, and the third time point (+8) is 8 days post-PPRV challenge.

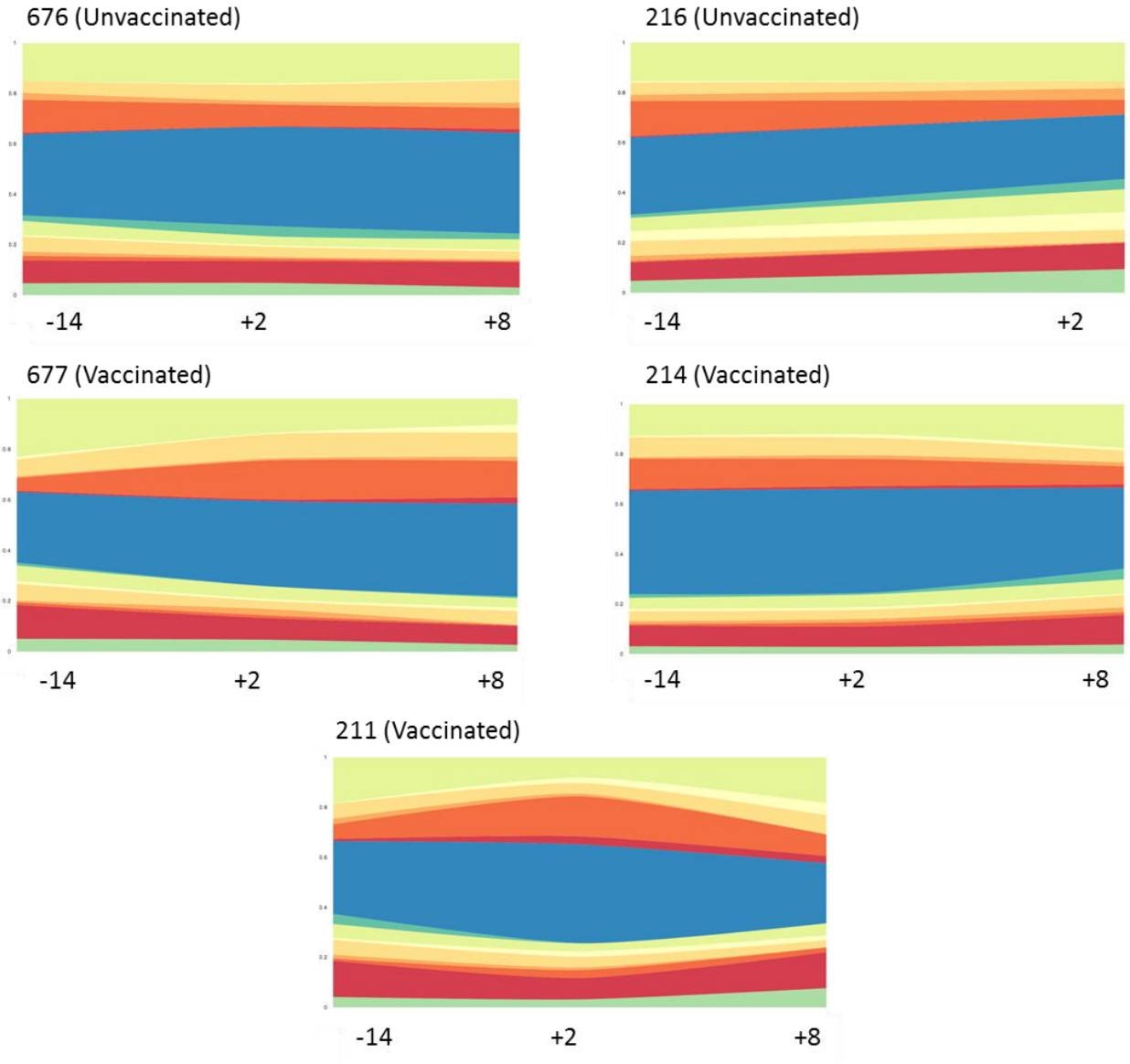


Figure 5-8. Proportional *KIR* transcription in each individual across the three time points. Read counts were generated by UniMMap and normalised per million reads. Animals 676 and 216 are unvaccinated controls and 677, 214 and 211 were inoculated with an experimental PPRV vaccine. The first time point (-14) is 14 days post-vaccination and 14 days pre-PPRV challenge, the second time point (+2) is 2 days post-PPRV challenge, and the third time point (+8) is 8 days post-PPRV challenge. Streamgraphs were generated using the streamgraph package in R.

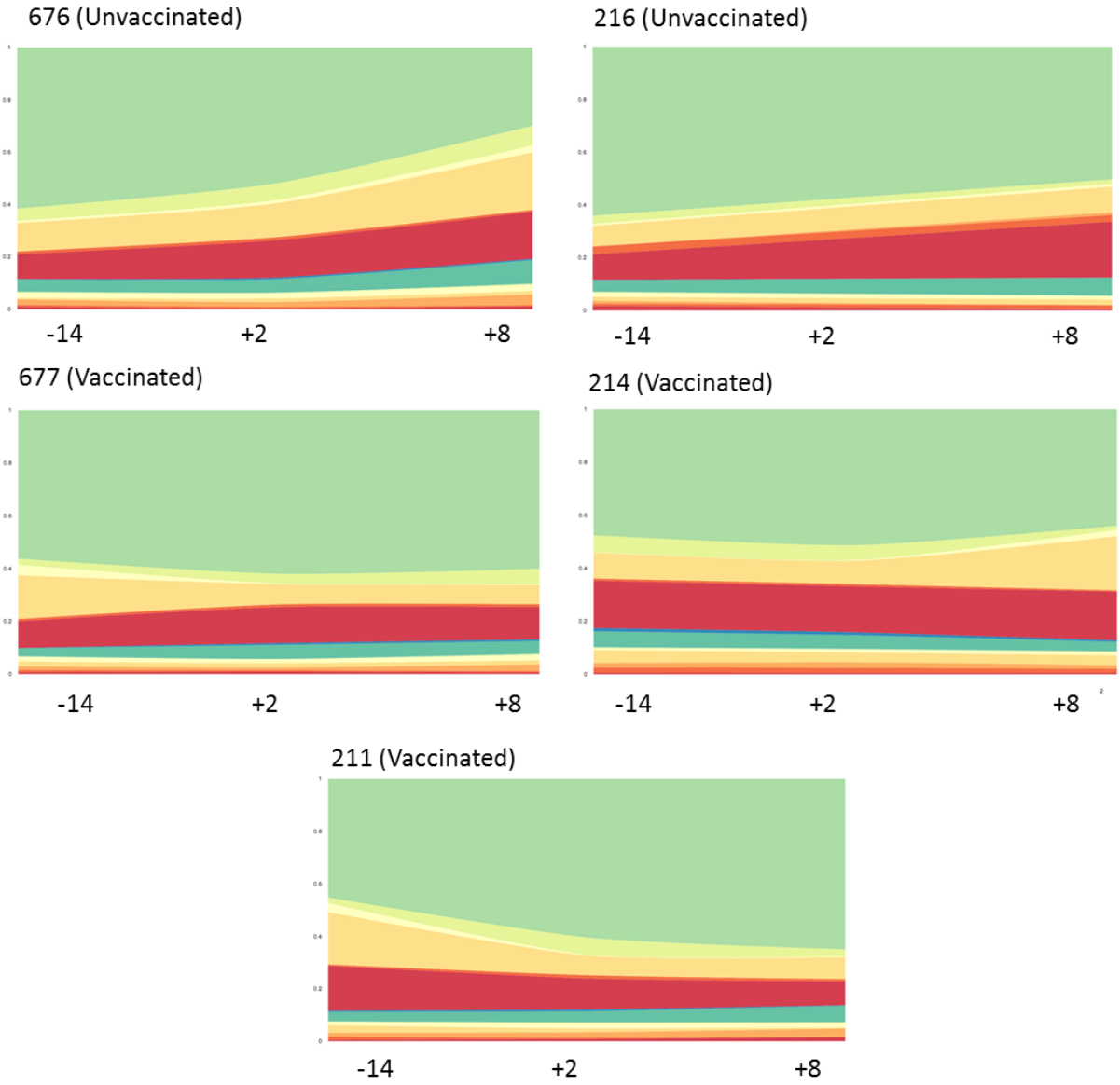


Figure 5-9. Proportional *KLR* transcription in each individual across the three time points. Read counts were generated by UniMMap and normalised per million reads. Animals 676 and 216 are unvaccinated controls and 677, 214 and 211 were inoculated with an experimental PPRV vaccine. The first time point (-14) is 14 days post-vaccination and 14 days pre-PPRV challenge, the second time point (+2) is 2 days post-PPRV challenge, and the third time point (+8) is 8 days post-PPRV challenge. Streamgraphs were generated using the streamgraph package in R.

3_KIR3DXL is the highest in all animals at all time points. The lowest transcribed gene at each time point is consistently one of the six predicted non-functional genes. Of the NKC genes, *KLRK1* is the highest transcribed gene within each animal at each time point (figure 5-9). As was observed in the cattle datasets (chapter 3 and 4), the novel *KLRC*-like and *KLRH*-like genes located in the expanded region between *KLRA* and *KLRJ* are consistently the lowest transcribed in goats.

5.3.5 Variability in transcription of individual LRC/NKC genes occurs between unvaccinated and vaccinated animals

The difference in read count between the first time point and the second/third time point was used to compare the transcriptional response of each of the individual LRC/NKC genes between vaccination status. Transcription of many LRC genes is similar between unvaccinated and vaccinated animals (figure 5-10). However there are clearly differences in the response between multiple genes. There is a large divergence in *3_KIR3DXL* transcription between the two groups, transcription increases in the unvaccinated group and decreases by a similar amount in the vaccinated group. Similar but less extreme divergences also occur for *5_KIR3DXL* and *7_KIR3DXS*. The difference in *NCR1* transcription between the two groups is also particularly noticeable. Transcription of *NCR1* decreases in both groups between the first and second time point, the decrease in the vaccinated group is twice that of the unvaccinated group. Within the unvaccinated group, *NCR1* transcription increases above the level at the initial time point whilst in the vaccinated group it increases only slightly and remains below the initial level.

A very similar pattern to *NCR1* transcription is seen with *KLRI1* in the NKC, although the magnitude of the change is greater (figure 5-10). The same diverging pattern observed in *3_KIR3DXL*, *5_KIR3DXL* and *7_KIR3DXS*, is also seen with *KLRJ1*, *KLRD1* and *KLRE1* in the NKC. Transcription of *KLRK1* increases in the unvaccinated group and decreases within the unvaccinated group at the second time point. By the third time point, transcription of *KLRK1* has dropped to almost the same level in both groups.

Transcriptional levels of only a subset of LRC and NKC genes change over the course of the study. The only gene to show a large response over time and have

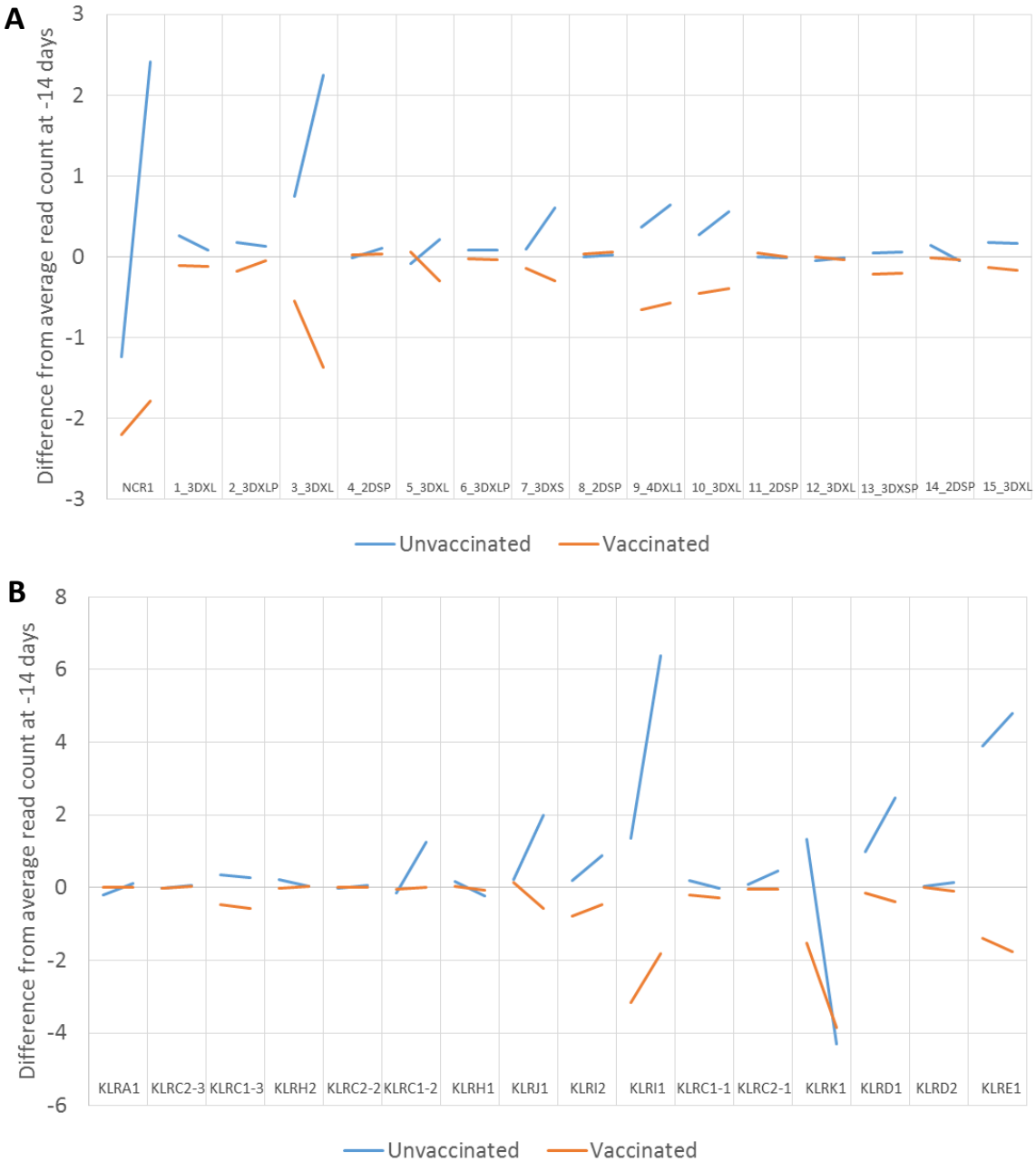


Figure 5-10. Average deviation in read count from the initial time point for individual *KIR* (A) and *KLR* (B) at the second and third time points. Read counts were generated by UniMMap and normalised per million reads. The first time point (-14) is 14 days post-vaccination and 14 days pre-PPRV challenge, the second time point (+2) is 2 days post-PPRV challenge, and the third time point (+8) is 8 days post-PPRV challenge.

a similar level in both groups is *KLRK1*. The majority of the changes in total LRC gene transcription appear to be driven by *3_3DXL*.

5.3.6 LRC/NKC genes are not well represented in the global PBMC transcriptome

To compare the transcription of the LRC/NKC genes with the rest if the transcriptome, a comparison was carried out between values at the first time point in the unvaccinated group. PBMCs from this group and time point most closely resemble *in vivo* PBMCs. The highest transcribed gene in the unvaccinated population at the first time point is *3_KIR3DXL*, which is ranked 10597 out of 19283 detected genes (figure 5-11A). The lowest transcribed (*8_KIR2DSP*) is ranked 16888. Of the 19283 goat genes for which transcription was detected, 54% are transcribed at a higher level than the most transcribed LRC gene. When comparing the average position of the LRC genes (13505), 70% of the total transcriptome is transcribed at a higher level.

Within the NKC genes, *KLRK1* is the highest transcribed and is ranked 4618/19283 and *KLRH2* is the lowest at 16515/19283 (figure 5-11B). Transcription of *KLRK1* is higher than 76% of the total detected genes. The average position of the NKC genes is 12079, meaning that 63% of the total transcriptome is transcribed at a higher level.

5.3.7 A large number of genes are differentially expressed between unvaccinated and vaccinated animals

Differential expression analysis was carried out to understand the extent to which the global transcriptome (25144 genes) changes in response to vaccination as well as subsequent challenge with PPRV. The total number of genes exhibiting increased or decreased expression in the vaccine group compared to the unvaccinated group was determined at each of the three time points (figure 5-12). At the first time point (14 days post-vaccination and 14 days pre-PPRV challenge), a total of 984 genes are significantly differentially expressed. Of the 984 differentially expressed genes, 430 and 554 are down and upregulated in the vaccine group respectively. There is little change in differential expression between the first two time points. The total number increases from 984 to 1021, 404 of the 1021 are downregulated and 617 are upregulated. Of the 404

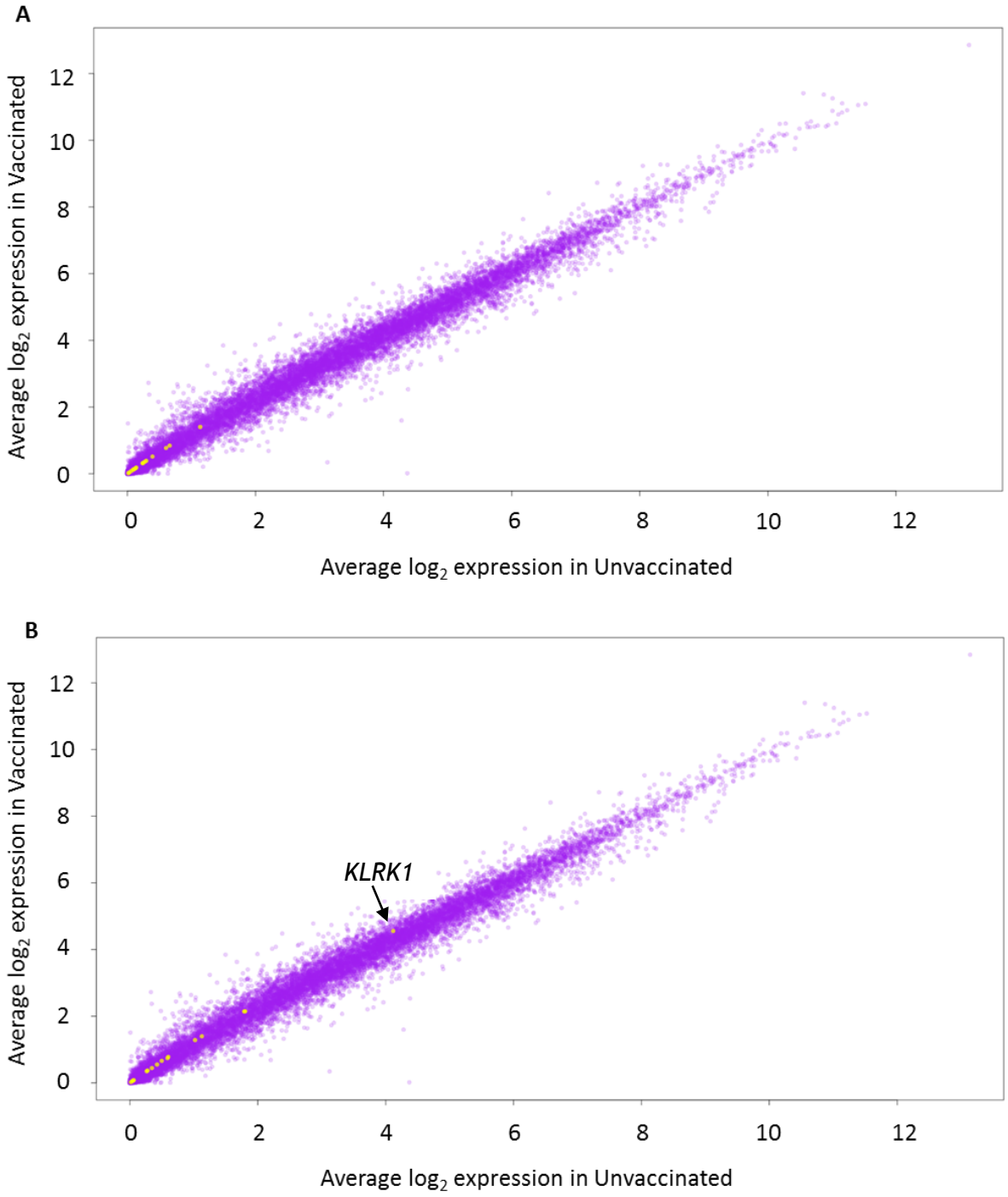


Figure 5-11. Transcription of LRC genes (A) and NKC genes (B) in comparison with the global transcriptome of unvaccinated and vaccinated animals. Average read counts from unvaccinated and unvaccinated at the initial time point (-14) were transformed to $\log_2$ values and plotted against each other using the plot function in R. Purple dots represent the global transcriptome and yellow dots represent either LRC (A) or NKC (B) genes.

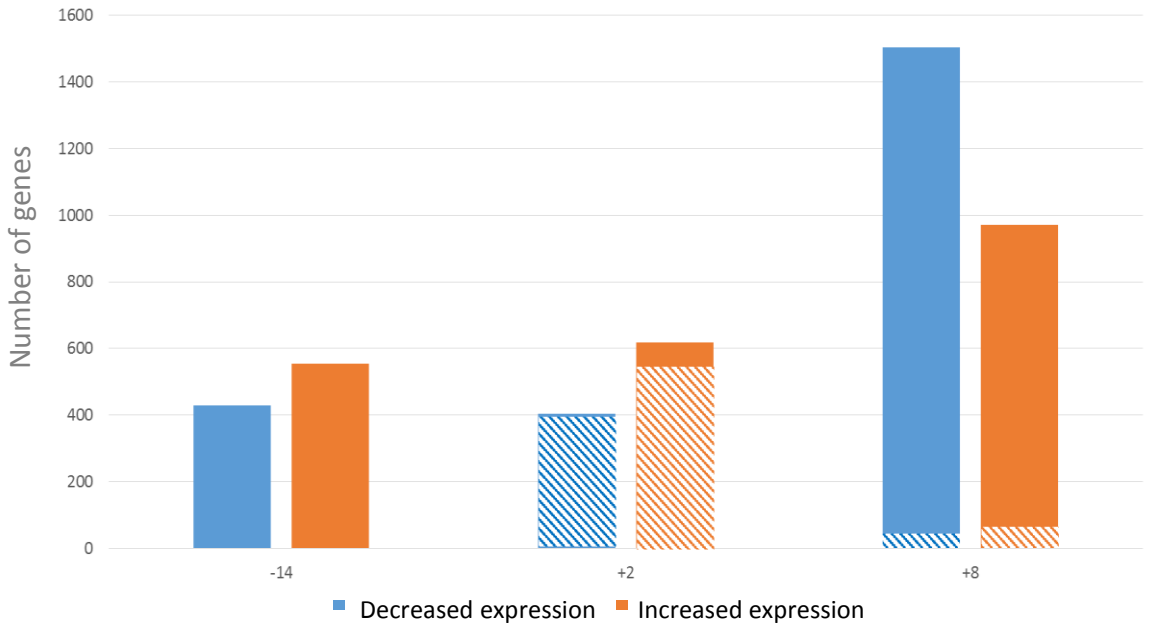


Figure 5-12. Number of statistically significant genes differentially expressed at each time point. Read counts were imported into the edgeR R package to calculate the number of genes exhibiting decreased and increased expression in the vaccinated group at each time point. Differentially expressed genes with a *p-value* of <0.05 were carried forward. Blue and orange bars indicate the number of downregulated and upregulated genes in the vaccine group at each time point. The diagonal lines of the same colour indicate how many of those genes were also differentially expressed at the first time point (-14).

downregulated genes at the second time point, 400 were also downregulated at the first time point. Within the 617 upregulated genes at the second time point, 550 were also upregulated at the first time point. The number of differentially expressed genes is much higher at the third and final time point compared to the previous time points. A total of 2473 genes are significantly differentially expressed, 970 of those are upregulated and 1503 are downregulated in the vaccine group. Despite the much higher number of differentially expressed genes, there is little overlap in the genes between the third time point and the first and second time points. Only 37 of the 1503 downregulated genes and 54 of the 970 upregulated genes were also differentially expressed at the first time point.

5.3.8 The top GO terms are largely identical between downregulated and upregulated genes

GO term analysis was carried out on the differentially expressed genes at each time points to understand which biological pathways were influenced. Across the three time points, the differentially expressed genes corresponded with 136 GO terms. The top 25 represented GO terms were compared between time points and vaccination status to identify differences (figure 5-13). Present in the top 25 upregulated terms but not downregulated are cell cycle, cell proliferation, chromosome and cytoskeleton. Contained in the top 25 downregulated terms but not upregulated are catabolic process, endoplasmic reticulum, ion binding and locomotion. Minimal variation occurs between the order of GO terms between time points. The immune system process is present in the top 25 downregulated (13/25) and upregulated (18/25) GO terms.

5.3.9 T cell related GO terms represent the majority of the total immune system process related genes

The GO terms that are contained within the immune system process GO term were compared between time points. This provided insight into differences in the immune system response between unvaccinated and vaccinated animals (figure 5-14a). Differential expression of a high number of T cell related genes occurs at all three time points. This number is highest at the third time point where 156 T cell related genes are differentially expressed, 5 and 3 of which are CD8 T cell and gamma delta T cell specific respectively. B cell related terms are

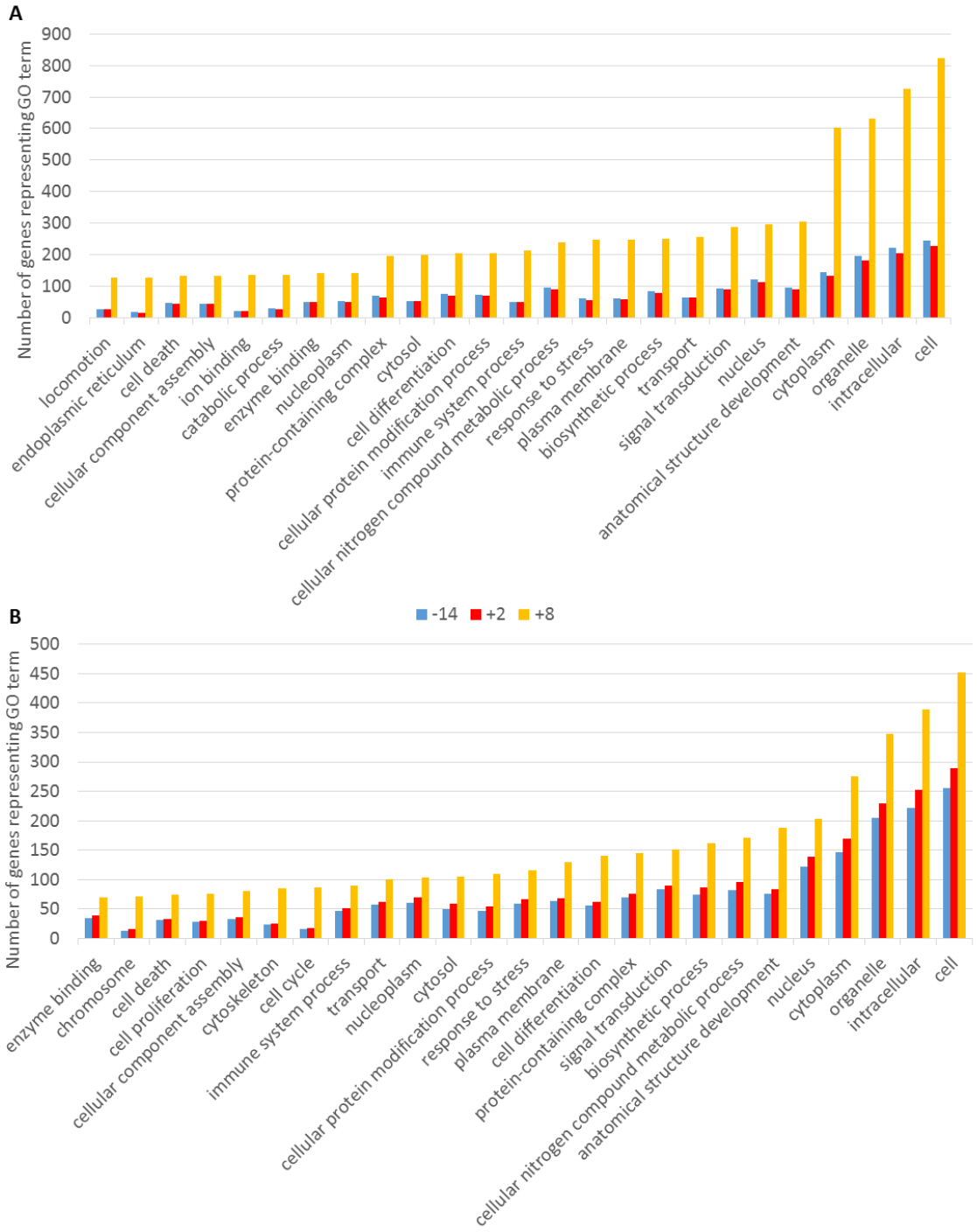


Figure 5-13. Top 25 GO terms downregulated (A) and upregulated (B) in the vaccinated group at all time points. Differentially expressed genes were imported into Ensembl BioMart to assign GO terms to each gene. The number of genes matching each GO term was calculated using awk.

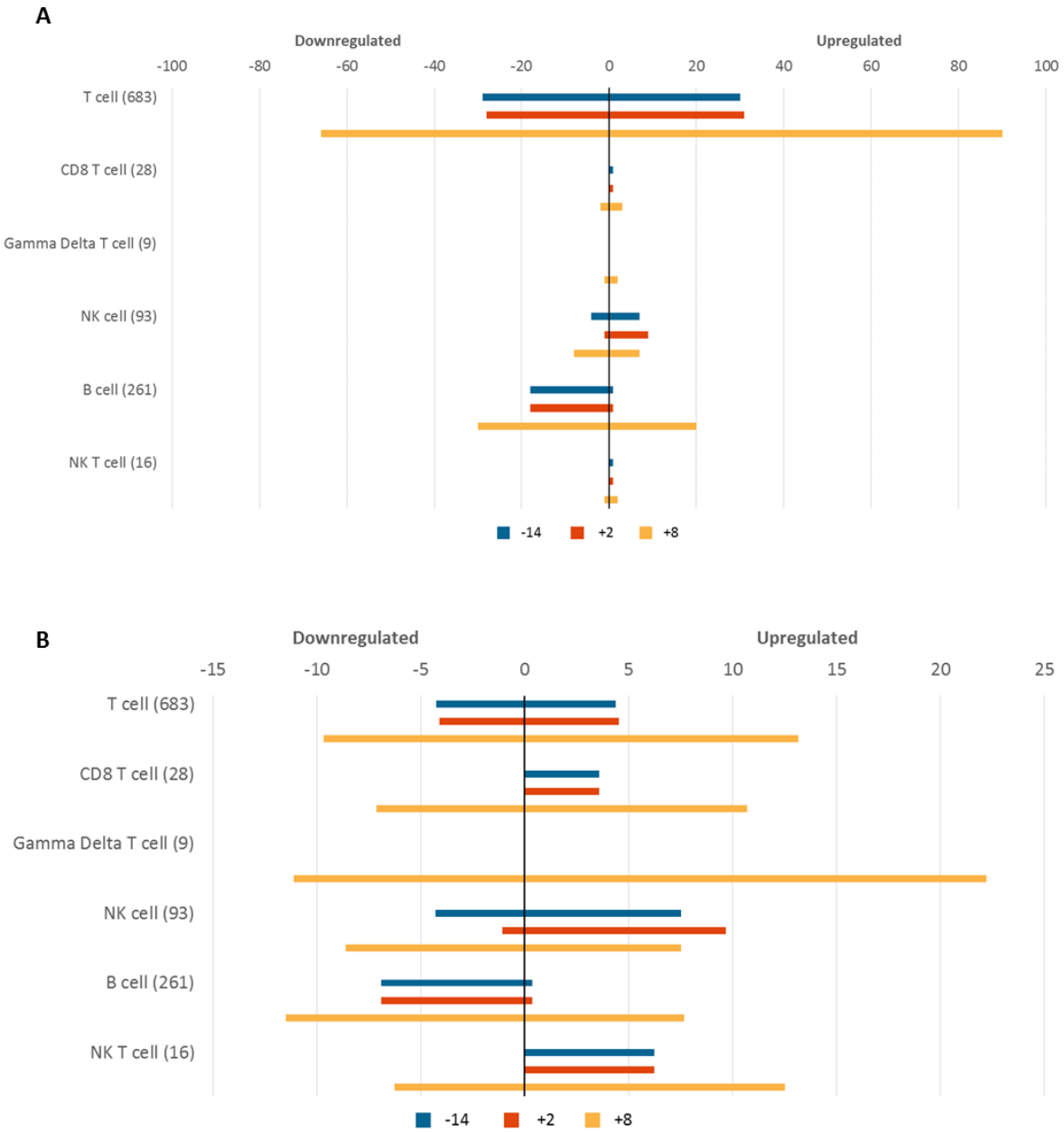


Figure 5-14. Number (A) and percentage (B) of occurrences of GO terms related to various immune cell types. GO terms was obtained by importing differentially expressed genes into Ensembl BioMart. GO terms related to each cell type were identified using awk. Upregulated or downregulated indicates genes that were upregulated or downregulated in the vaccine group compared to the unvaccinated group. The number in brackets next to each cell type indicates how many GO terms occur when analysing the entire genome.

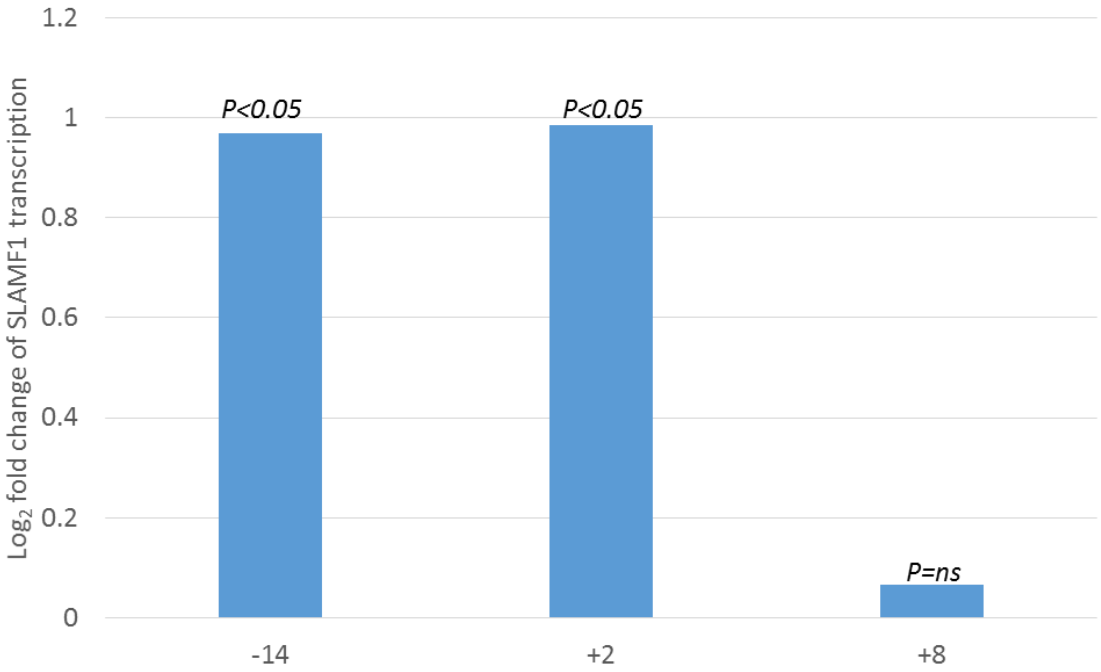


Figure 5-15. Log_2 fold change of *SLAMF1* transcription between unvaccinated and vaccinated animals. Fold change was calculated using the edgeR package in R. A positive fold change indicates an increase in the vaccinated group.

predominantly downregulated at the first two time points in vaccinated animals. At the third time point, 30 genes with B cell related GO terms are downregulated and 20 are upregulated. A small number of NK cell and NK T cell related genes are also differentially expressed at all three time points.

The number of T cell related terms associated with the goat genome is much higher than any other individual immune cell type. To reduce bias towards cell types with more GO term annotations, the percentage of the total GO terms of each cell type differentially expressed was calculated (figure 5-14B). This does not change the overall pattern of differential expression within cell types, but does alter the comparison between cell types. A higher percentage of B cell related genes are downregulated at all three time points compared to T cells. The percentage of NK cell related genes upregulated is higher at the first two time points than the corresponding T cell values.

5.3.10 The PPRV receptor SLAMF1 is significantly upregulated after vaccination

Contained within the differentially expressed genes at the first two time points was the gene encoding for the PPRV receptor SLAMF1 (figure 5-15). Transcription of *SLAMF1* occurs at a level approximately twice as high in the vaccinated animals at the first and second time points. *SLAMF1* is not significantly differentially expressed at the final time point.

5.4 Discussion

Using RNA-Seq data from a publically available goat gene atlas, the pattern of LRC/NKC gene transcription in various gene tissues was established. Total LRC/NKC gene transcription was correlated with *NCR1* transcription to provide an approximation of the contribution of NK cells. RNA-Seq data was generated from unvaccinated and vaccinated goats during a PPRV vaccination study. UniMMap was run on these samples to characterise differences of LRC/NKC gene transcription between the unvaccinated and vaccinated animals. The impact of PPRV infection on LRC/NKC transcription was also assessed. The global transcriptome of these animals was also analysed to determine the relative transcription level of the LRC/NKC genes as well as to understand differences in the wider immune system.

5.4.1 Goat macrophages transcribe *KIR*

Both bone marrow (BM) and alveolar macrophages transcribe numerous *KIR* genes. Macrophages transcribe *1_KIR3DXL* at a higher level than any other tissue, although it would be expected a pure cell type would have higher transcription of a cell specific gene than a more general tissue type would. Transcription/expression of *KIR* has not been shown to be associated with macrophages in any published data to date. Much higher transcription of *NCR1* in the spleen than any of the other tissue types would suggest *NCR1* is transcribed by goat NK cells. The low levels of *NCR1* transcription observed in the macrophage populations, particularly BM macrophages, suggests it is unlikely that *KIR* transcription is due to contamination with NK cells. This is further supported by the pattern of NKC gene transcription in the macrophage populations. Cattle NK cells typically transcribe the majority of NKC genes, particularly those between *KLRI2* and *KLRE1* which are transcribed at a higher level than any of the *KIR* (chapter 3 and 4). Goat macrophages only appear to transcribe a small subset of the NKC genes, at a level below any of the *KIR* genes they transcribe. This suggests goat macrophages utilise *KIR* in some manner, particularly *1_KIR3DXL* which may serve to inhibit macrophage function.

5.4.2 The testis contain a unique LRC/NKC receptor phenotype

Testis tissue possesses the fifth highest level of total LRC gene transcription of the 20 examined tissue/cell types. The total NKC gene transcription of testis tissue is second only to spleen tissue, a tissue know to contain a high proportion of NK cells (Grégoire et al. 2007). Despite this, transcription of the NK marker *NCR1* is particularly low in testis tissue. The amount of total LRC gene transcription relative to *NCR1* transcription is higher only in BM macrophages and is almost identical to alveolar macrophages. The amount of total NKC gene transcription per *NCR1* transcription is much higher in testis tissue than any other tested tissue/cell type. The pattern of NKC gene transcription in testis tissues closely resembles splenic tissue, with the exception of the absence of *KLRJ1* in testis tissue. We have shown that cattle NK cells transcribe *KLRJ* only at a very low level, *KLRJ* transcription predominantly occurs in gamma delta T cells (chapter 4). Despite the transcription pattern of LRC/NKC genes in testis tissue closely resembling that of cattle NK cells, *NCR1* transcription is far below the expected value of NK cells. This suggests that either a non-NK cell type exists in goat testis that transcribe LRC/NKC genes similarly to cattle NK cells, or that NK cells in the testis heavily downregulate *NCR1* transcription.

5.4.3 Changes in LRC/NKC gene transcription vary between individuals after PPRV-challenge

The response of LRC/NKC gene transcription to both PPRV-vaccination and PPRV-challenge is highly variable between animals. This could be due to haplotypic/allelic variation between animals, which has been implicated in differential outcomes to numerous viral infections in humans (Martin et al. 2007; Khakoo et al. 2004; Jost et al. 2011; Warfield et al. 2004). The infection history of an individual can also result in NK populations with different abilities to respond, due to terminal differentiation of NK cells (Strauss-Albee et al. 2015).

Averaging the read counts of the individuals within the unvaccinated and vaccinated groups serves to reduce the individual variation. Comparing average total read counts for both the LRC and NKC genes shows that both vaccination and PPRV-infection alter the total transcription of the genes of both complexes. Over the course of PPRV infection in unvaccinated animals, transcription of LRC/NKC genes increases. Changes in total NKC gene transcription in individuals

closely mirrors changes in *NCR1* transcription. This could be due to either *NCR1* responding in the same manner as the NKC genes or changes in NKC gene transcription occur because of changes in the number of circulating NK cells. Infection with PPRV is associated with leukopenia (Kumar et al. 2014) and so a decrease in NK cell number in infected animals is expected.

5.4.4 There are distinct differences in LRC/NKC gene transcription based on infection and vaccination status

Total LRC and NKC gene transcription increases in the unvaccinated group after infection. The increase is driven by a subset of genes (*3_3DXL*, *KLRI1* and *KLRE1*). This suggests they play a role in the response of the host to PPRV infection. In all three vaccinated animals, total LRC gene transcription is lower than prior to PPRV challenge. This is due predominantly to a decrease in *3_3DXL* transcription. This difference in *3_3DXL* transcription between unvaccinated and vaccinated animals could be explained by a subpopulation of *3_3DXL*$^{+}$ cells responding to PPRV infection. In the unvaccinated animals this population could be in the process of expansion. Whereas in the vaccinated animals expansion has already occurred, the population has reached exhaustion due to cytokine production and is in decline. This could also be occurring with a *KLRK*$^{+}$ population, *KLRK* is transcribed at a higher and lower level post challenge in the unvaccinated group and vaccinated group respectively. In both groups *KLRK* transcription is greatly reduced by the final time point. However, NK cell exhaustion is typically associated with an increase in inhibitory receptor expression (Bi and Tian, 2017). Therefore, an alternate explanation could be that the increase in transcription is occurring in CD8+ T cells, which upregulate inhibitory receptors after encountering antigen (Anfossi et al. 2004). The initial time point in the vaccinated animals could represent this and would explain why transcription of *3_3DXL* is highest at this time point.

It is also interesting to note the difference in total transcription between the two unvaccinated animals. Animal 216 had to be euthanised early after the second time point due to the severity of infection. This correlated with a much greater increase in total NKC gene transcription and lower *NCR1* transcription in animal 216, whereas total LRC gene transcription was almost identical between the two animals. This could suggest an *NCR1*$^{-}$ population underwent expansion.

The total transcription of LRC/NKC genes is remarkably similar between the vaccinated group at the initial time point (pre-PPRV challenge) and the unvaccinated group at the final time point (8 days post-PPRV challenge). This suggests that the vaccine drives a similar LRC/NKC gene response to the virus. By the third time point, the total transcription in vaccinated animals falls to a similar level as the unvaccinated group at the initial time point. This reduction could be caused by NK turnover - in humans labelled NK cells disappear from circulation with a half-life of <10 days (Zhang et al. 2007).

Further experiments with larger sample numbers are required to test these hypotheses. Sorting PBMCs by immune cell type prior to sequencing would provide information as to the contribution of each cell type to significant changes in transcription.

5.4.5 A subset of LRC/NKC genes respond differently to PPRV-challenge between unvaccinated and unvaccinated animals

Comparing the average response of individual receptors to PPRV-challenge between the unvaccinated and vaccinated groups shows that there is variation in the response of only a subset of the genes. The difference in total transcription of LRC genes between the two groups is almost entirely driven by *3_KIR3DXL*. The correlation between *3_KIR3DXL* and *NCR1* was the highest of all the *KIR*, suggesting NK cells account for more of its transcription than they do any other *KIR*. Subsequently this suggests that this response may be occurring in NK cells, rather than CD8+ T cells. The response of *KLRI1*, which in cattle has been shown to be primarily transcribed on NK cells (chapter 4), mirrors that of *NCR1*. This suggests that the change in *KLRI1* transcription is due to differences in the number of circulating NK cells. Future experiments would benefit from FACS analysis to determine the percentage of NK cells present in each PBMC sample.

5.4.6 The largest differential expression occurs 8 days post PPRV-challenge

Vaccination results in statistically significant (*p-value* <0.05) differential expression of 984 genes. Of the 984 genes, 950 remain differentially expressed 2 days post-PPRV challenge. Only 71 additional genes are differentially expressed 2 days post-PPRV challenge. The differential expression of genes changes drastically between 2 and 8 days post-challenge. The number of differentially expressed genes almost triples between the first and third time points. However

only a small fraction of the differentially expressed genes at the final time point were also differentially expressed at the first. This correlates with the timeline of PPRV infection where fever usually develops 3-7 days post infection (dpi) and additional clinical signs develop within 3-5 days after fever establishes (Kumar et al. 2004). At the final time point a large number of T cell and B cell genes are differentially expressed between the unvaccinated and vaccinated group. The proportion of CD4+ T cells has been shown to decrease 4 days post-challenge in unvaccinated but not vaccinated animals. The proportion of CD8+ T cells slightly increases 7 days post-challenge in both unvaccinated and vaccinated animals (Herbert et al. 2014). The passive transfer of immunity to PPRV via colostrum suggests a role for B cells in infection (Ata et al. 1989).

Also contributing to the large number of differentially expressed genes at the final time point was a gender imbalance occurring between the unvaccinated and vaccinated groups. This occurred due to the euthanisation of a control animal prior to the final time point. Prior to euthanisation, the unvaccinated group consisted of one female and one male and the vaccinated group of two males and a female. As the unvaccinated group only contained a female at the final time point, and the vaccinated group contained males, a large number of the differentially expressed genes would have been sex-linked genes. Many of these genes would likely have not been differentially expressed at previous time points.

While outside the scope of this work, further detailed analysis of the immune response using these RNA-Seq datasets could provide valuable information into how the vaccine induces protection. In future experiments, coupling transcriptomics with FACS analysis to determine immune cell percentages would provide valuable insight.

Chapter 6. Discussion

6.1 Summary of findings

The primary aim of the project was the creation of a pipeline to accurately quantify RNA-Seq reads mapping to complex regions of the genome. With a particular focus on the cattle LRC and NKC. These gene complexes have undergone significant expansion in cattle resulting in a large number of highly similar genes. Subsequently, short RNA-Seq reads map to multiple locations within the complexes. Due the high sequence similarity of the genes within each complex, previous attempts to characterise transcription of these regions using qPCR required grouping the most similar genes together - individual gene resolution was not possible. Information regarding transcription of individual genes is important to confirm existing gene models. This necessitates an understanding of which genes are transcribed and the extent to which transcription occurs in various tissues and cell types.

The analysis pipeline UniMMap was created, utilising the concept of mappability to calculate the 'uniqueness' of LRC/NKC coding sequences. After mapping, a score is assigned to each read based on the average mappability of the locations it maps to. Only genes that have reads mapping to regions that are unique are considered to be transcribed. This enables high confidence when determining whether a gene is transcribed. In humans, the receptors encoded by the LRC and NKC genes are primarily NK cell receptors this is also thought to be the case for cattle. In order to characterise the transcription of LRC and NKC genes in cattle, UniMMap was run on RNA-Seq data generated from paired PBMC and NK cell samples from two animals. This provided the first information on the transcriptional status for the highly similar genes that could not be resolved by previous methods, as well as novel genes only recently characterised. It also gave insight into the variation of their transcriptional levels between individuals. The read counts obtained informed future sequencing experiments by indicating the read coverage required to detect LRC/NKC gene transcripts. UniMMap was also used to assess transcription of LRC/NKC genes in

multiple immune cell types, compare transcription between dams and calves and to create an LRC/NKC gene atlas across various tissue types. The utility of UniMMap in analysis of other species was demonstrated through the generation of RNA-Seq data from goats involved in a PPRV vaccination study.

6.1.1 LRC/NKC gene transcription is variable between animals

Transcriptional analysis of the full repertoire of known cattle LRC/NKC genes has been carried out in multiple animals. The amount of total transcription of the genes of each complex is highly variable between individuals. However, higher total transcription does not equate with higher transcription of each individual gene. This could be due to haplotypic variation, particularly with the LRC. There is considerable haplotypic variation of the LRC within humans where two major haplotype groups exist, A and B. The B haplotype can contain up to five *KIR* not found in the A haplotype. Therefore, the total transcription in an individual heterozygous for haplotype A and B would be expected to be lower than an individual homozygous for the group B haplotype. Transcription could also be variable between individuals due to allelic differences. Alleles more divergent from the reference sequence may be represented less in the data due to a higher number of mismatches. Across the NK RNA-Seq datasets from 5 individual cattle analysed here, transcription of all of the genes were consistently observed. Although this is a small sample size, these five animals can be assumed to possess at least one copy of each known *KIR*. Variation in the amount of transcription between individual genes may suggest that haplotypes exist that do not contain all of the *KIR*. This also highlights one of the weaknesses of UniMMap, it is only able to quantify transcription of genes present in the reference.

The genes of the NKC can be divided into two distinct groups based on the amount of transcription occurring. The genes between and including *KLRI2* and *KLRE* are consistently transcribed at a much higher rate than the genes located between and including *KLRA* and *KLRJ*. As with the LRC genes, the highly transcribed genes are transcribed consistently in the five NK

datasets. There is variation however in the genes that are transcribed at a low level. Transcription of *KLRH7*, *KLRC1-6*, *KLRC2-4*, *KLRH5*, *KLRC1-4*, *KLRH3*, *KLRC2-3* and *KLRJ* is absent in at least one of the five animals. This suggests that haplotypic variation may also occur within the cattle NKC. Alternatively, the low transcription of these genes could make them undetectable in RNA-Seq datasets. This may signify that the receptors they encode are not transcribed for under normal conditions, occurring only in response to particular stimuli. Further evidence for the influence of read depth comes from the analysis of PBMC RNA-Seq datasets from an additional six cattle, where variation in presence/absence of gene transcription emerges. Transcription of *3DXL7*, *3DXL4*, *2DS2*, *3DXS2*, *3DXL5*, *3DXL2* and *3DXL1* is absent in at least one animal. It is difficult to determine whether any of these genes are absent/non-transcribed, or simply undetectable due to the low read depth.

6.1.2 Genes predicted to be non-functional are transcribed in multiple species

Transcription of *KIR* pseudogenes/putatively non-functional genes/alleles is reported by UniMMap in humans, cattle and goats. In humans, mRNA of both *KIR* pseudogenes has been observed in multiple individuals. Coupled with confidence in the accuracy of UniMMap, this suggests that the observed transcription of predicted non-functional genes in cattle and goats is accurate. It is unlikely to be a result of unaccounted multi-mapping. In the case of the human pseudogene *3DP1*, it can encode for a secreted receptor in a minority of individuals (Gómez-Lozano et al. 2005). It could therefore be possible that one or more of the cattle/goat *KIR* gene products are secreted, and that encoding for a truncated transcript is not necessarily an indicator of non-functional status. Putatively non-functional *KLR* genes are also reported in cattle and goats. Transcription of human *KLR* was not analysed as the human *KLR* have not expanded to the same extent as cattle, making any comparisons inaccurate. As non-functional status in cattle and goats is based off the individual used for either respective genome, it is possible that there are functional alleles of these

genes. Transcription does not always result in translation, it could be that transcription of non-functional genes is a mechanism of regulating the transcription of other genes, rather than encoding a protein. Another possibility is that the non-functional status is derived from errors in the genome assembly. The genes of the LRC/NKC are highly repetitive and pose enormous difficult for assembly algorithms, even with the use of long-read PacBio sequencing.

6.1.3 Transcription of LRC/NKC genes is not exclusive to NK cells

Transcription of LRC/NKC genes is present to a varying extent in all tested cattle immune cell types. Transcription of *KIR* occurs predominantly in cattle NK cells and CD8+ T cells and to a lesser extent, gamma delta T cells. Low level transcription of some, but not all *KIR*, was observed in CD4+ T cells, B cells and monocytes. Subsets of human CD8+ T cells express KIR, but the pattern of expression is different to NK cells. Just one inhibitory KIR is expressed by ~90% of KIR expressing CD8+ T cells. In comparison ~75% of NK cells express one inhibitory KIR, the remainder express 2 or more (Bjorkstrom et al. 2012). Although inhibitory KIR can independently inhibit CD8+ T cell functions, activating KIR have a co-stimulatory role in conjunction with the T-cell receptor. In cattle, transcription of *KIR* at a level comparable to NK cells suggests that they could play a role in the functional control of CD8+ T cells. Human gamma delta T cells expressing KIR have also been observed in humans. Expression of 2DL1 by gamma delta T cells is associated with non-responsiveness to malaria. In contrast, KLRA is more frequently expressed on gamma delta T cells that respond to malaria. Transcription of *KLRA* is not particularly high in cattle gamma delta T cells. They do however transcribe *KLRJ* at a level much higher than any other tested cell type, including NK cells. Transcription of *KLRJ* has previously been demonstrated to be lowly transcribed in NK cells (Boysen et al. 2006). As *KLRJ* lacks either an inhibitory or activating signalling component, Schwartz et al (2017) suggest it could form a heterodimer with an unknown partner. Elevated transcription of *KLRK* in cattle gamma delta T cells, could suggest KLRK is

the partner. Alternatively, as KLRJ lacks the classical signalling component, it could mean it is a secreted receptor.

Although transcribed at low levels in the various cell types, the majority of NKC gene transcription occurs in NK cells. With the exception of the aforementioned *KLRJ* and *KLRK*. Transcription of *KLRK* is also observed in CD8+ T cells at a comparable level to NK cells. In humans KLRK is constitutively expressed on CD8+ T cells (Bauer et al. 1999), in mice it is only expressed on activated CD8+ T cells (Ho et al. 1998). Whilst in NK cells activation of KLRK alone is sufficient to induce killing, in CD8+ T cells the T-cell receptor must also be activated (Maasho et al. 2005). The observation of transcription of *KLRK* at a similar level to NK cells in cattle suggests that KLRK may also be involved in activation of their CD8+ T cells. Although thought to be non-functional based on the sequence in the reference, *KLRI2* is transcribed at a consistently high level in cattle PBMC/NK datasets and goat PBMC datasets. It is transcribed at a very low level or is absent in all other examined immune cell types, including goat macrophages. Depending on whether or not it is expressed on the cell surface, KLRI2 could potentially be used as an additional NK cell marker in cattle. Individual immune cell types would need to be interrogated before the same could be said of goats. However the high *KLRI2* transcription in PBMC data suggests it could be a possibility.

In goats, the transcription of LRC and NKC genes in bone marrow macrophages and alveolar macrophages was characterised. They were found to transcribe a number of *KIR*, particularly *1_3DXL*. Transcription or expression of *KIR* in macrophages has yet to be shown in any other species. This suggests a unique function for KIR in goats, potentially 1_3DXL could be used to inhibit macrophage functions. RNA-Seq analysis of cattle macrophages populations, as well as of other goat immune cell types would provide insight into similarities and differences in gene usage between the two species.

6.1.4 LRC/NKC gene transcription is observed in every analysed tissue type

Comparisons of LRC/NKC gene transcription between the cattle and goat atlases is complicated by the differences in tissue types selected. However, transcription of at least one LRC and NKC gene was observed in every tissue type in both species. Tissues isolated from the brain transcribe at the lowest level, which is not surprising given its immune-privileged status. NK cells have been observed in the brains of healthy mice (Poli et al. 2013). The CNS is also patrolled by T cells where they carry out immune surveillance (Ferretti et al. 2016). This could provide the explanation for transcription of LRC/NKC genes in brain tissue of cattle/goats. Alternatively, tissue samples could have been contaminated with blood during isolation.

Relatively high transcription of LRC/NKC genes in cattle lymph node, lung and bone marrow as well as goat spleen is unsurprising. In other species these tissues contain a high proportion of NK cells as well as other KIR expressing immune cells. Surprisingly, transcription was also high in cattle mammary gland tissue. Bacterial infection in mammary gland tissue of cattle has been shown to induce NK cell migration (Sipka et al. 2016). This also highlights some of the difficulties in using pre-generated data for comparisons of this nature. A number of tissues that contain unique NK phenotypes in humans, such as spleen, lung, liver and spleen, are absent from one of the gene atlases.

The pattern of LRC/NKC gene transcription is particularly interesting in the goat. An approximation of LRC/NKC gene transcription by NK cells was provided by comparing with transcription of the NK marker *NCR1*. With the exception of the macrophages, testis tissue has the highest ratio of total LRC gene transcription per *NCR1* read. It also has the highest ratio of total NKC gene transcription per *NCR1* read. In comparison, cattle testis tissue does not transcribe the genes of either complex particularly highly. The high level of NKC gene transcription suggests NK cells are a major contributor to the total transcription. However the low amount of *NCR1*

transcription suggests that either testis NK cells downregulate NCR1 or that there is non-NK cell type that expresses NKC genes in a similar manner to NK cells. Characterisation of immune cell types in goats, as has been done in cattle, would allow a more thorough understanding of LRC/NKC gene phenotypes. Additionally, expanding UniMMap to analyse more than just LRC/NKC genes would be highly beneficial for this analysis. Including CD8/CD4 as well as markers for other cell types would allow a more robust interrogation of the cell types contributing to total LRC/NKC transcription.

6.1.5 UniMMap provides utility in interrogating the response to infection

UniMMap was used to characterise differences in LRC/NKC gene transcription between uninfected and animals infected with *M. bovis*. As this was publically available data and not created for the purpose of analysis of LRC/NKC gene transcription, there are a number of limiting factors. Firstly the data is not of suitable read depth for analysis of the relatively poorly transcribed LRC/NKC genes. This meant datasets from multiple animals had to be combined prior to analysis. Secondly, the LRC/NKC gene content of the animals is not characterised and the uninfected and infected groups are made up of separate animals. This makes comparisons difficult as the extent to which haplotypic variation contributes is unknown. Nonetheless, analysis of these datasets can be used to infer either differences in transcription between uninfected/infected animals or further evidence of haplotypic variation. Although transcription of the NKC genes is highly similar between the two groups, there is variation in LRC gene transcription. Notably, transcription of *3DXL7* is absent in the control group and *3DXL5* is absent in the infected. Absence of *3DXL7* was observed in four out of six PBMC RNA-Seq datasets, in three of those four individuals transcription of *3DXL5* was also absent. This suggests that *3DXL5*/*3DXL7* are either frequently absent/non-transcribed, or that they are often transcribed at an undetectable level in PBMCs. The latter could be investigated by resequencing at a much higher read depth. Transcription of all *KIR* is higher with the exception of *3DXL7* and *2DL1* is higher in the control population. This suggests that *M. bovis*

infection could induce an increase in *KIR* transcription. Alternatively, *M. bovis* infection may induce NK cells/CD8+ T cells to migrate from the blood, reducing their cell count in the blood, and subsequently overall observed transcription. Further work into changes to LRC/NKC gene transcription in response to infection would benefit from FACS profiling of the cell populations present in each individual.

UniMMap was also used to investigate changes to LRC/NKC gene transcription in goats from a PPRV vaccination study. The response of the LRC/NKC genes seems to be limited to a subset when comparing average read counts from unvaccinated and vaccinated animals. Within the LRC, the major driver of the difference in total transcription observed between the two groups is *3_3DXL*, involved to a lesser extent are *5_3DXL*, *7*_3DXS, 9*_4DXL1* and *10_3DXL*. Interestingly, with the exception of *7_3DXS*, these are all inhibitory receptors. They decrease post-challenge in the vaccinated and increase in the unvaccinated. This could hint that decrease transcription of inhibitory receptors in the vaccinated animals could contribute to the favourable outcome by enabling more NK cell activation. This is far from conclusive however. A larger scale study would be required to provide statistical support. As mentioned previously, FACS data would increase the power of RNA-Seq analysis. Quantifying the percentage of various immune cells within the PBMCs at each time point would enable the contribution of the various cell types to be assessed.

6.1.6 Conclusions

UniMMap, the pipeline for the quantification of RNA-Seq short reads from complex regions of the genome, enables confidence in analysis results. This is clearly demonstrated through the accuracy of read counts from cattle simulated reads and from human NK cell RNA-Seq data mapped to their respective LRC/NKC genes (chapter 2). The utility of UniMMap was examined in various analysis scenarios and used to provide valuable information on transcription of individual LRC/NKC genes in multiple individuals, cell types, tissue types, infection statuses and species.

The amount of total transcription of the genes of both complexes is highly variable between individuals. This is most likely due to one of two mechanisms. Haplotypic variation may result in differences in gene content between individuals, subsequently transcription levels will vary. Alternately, or perhaps additionally, varying amounts of divergence of an individual's alleles from the reference may result in differences in the percentage of reads successfully mapped.

Genes predicted to be non-functional based on the reference sequence are consistently transcribed in both cattle and goats. This raises multiple possibilities. Functional alleles of some of these genes may exist in the population. Some may encode for secreted or otherwise non-prototypical receptors. Transcripts from non-functional genes may also provide a mechanism of regulation transcription of the functional genes.

Cattle/LRC genes are transcribed in immune cell types other than NK cells. LRC gene transcription occurs at a very similar level in CD8+ T cells to NK cells, and at a lower level in gamma delta T cells. The majority of the NKC genes are transcribed only at a low level in other immune cell types. The exception to this are *KLRK* and *KLRJ*. Transcription of *KLRK* is observed at an elevated level in CD8+ T cell and gamma delta T cells. Gamma delta T cells also transcribe *KLRK* at an elevated level. Of all of the LRC/NKC genes analysed here, *KLRI2* is the only gene essentially exclusive to NK cells that is transcribed at relatively high level.

UniMMap is demonstrated as viable for the analysis of RNA-Seq data from infection or vaccination studies. Early insight into the transcriptional response of individual LRC/NKC genes in both *M. bovis* infection in cattle and PPRV infection in goats is provided. The utility of UniMMap as part of a larger collection of analysis tools to deconstruct the immune response is indicated.

6.2 Future work

Achieving the main aim of this project enables further characterisation of LRC/NKC genes in more individuals as well as species. Due to the design of UniMMap, it can be quickly and simply adapted to new species and gene

complexes. Many of the results generated with UniMMap open up a number of new avenues for interrogation that would previously not have been possible.

6.2.1 Expansion of the UniMMap pipeline

The accuracy and usefulness of the information provided by UniMMap indicates that expansion of its functions would provide additional value. Expansion of the genes analysed by UniMMap simply requires the addition of their coding sequences to the input reference. As the ligands of the majority of inhibitory KIR/KLR are the MHC class I, their addition to the pipeline would allow the interrogation of receptor and ligand under various conditions. The addition of other important gene complexes, such as the T-cell receptor complex, would enable the parallel analysis of multiple immune cell types. UniMMap coupled with FACS analysis of proportions of cell types would be a powerful tool for investigating future vaccination or infection studies. This would be of particular importance in a vaccination study, where the exact nature of the immune response is often poorly understood.

In addition to expanding the repertoire of genes included in a typical UniMMap analysis, there is potential further use of the unique regions that are identified as part of the pipeline. Extracting these reads would facilitate assembly of the transcripts of each gene, without contamination of reads from other genes. Although this would be complicated by the likely presence of two copies of each gene, information on alleles could be elucidated. It would also provide valuable information on the potential proteins encoded for by the genes that are non-functional in the reference genome.

6.2.2 UniMMap analysis of other data types

It would also be possible to analyse DNA-Seq data, particularly with a previous version of UniMMap that was described in chapter 2. This could facilitate accurate high-throughput genotyping of a large number of genomes. Additional genotype/haplotype information is crucial to

understanding these regions, as well as improving the accuracy of UniMMap. As with RNA-Seq data, it could be used to improve the assembly of individual genes. It could also be used to extract reads mapping to a certain location of each gene, facilitating a comparison of a particular exon such as an individual immunoglobulin domain. This could provide an early indication of the scale of diversity.

Single cell RNA-Seq is an obvious next step of the work presented here, especially given the nature of variegated expression of NK receptors. Understanding the nature of NK receptor transcription in individual NK cells would provide additional resolution whilst also improving the interpretation of RNA-Seq from a large population of cells.

7. Appendices

7.1 Chapter 2 Appendix

7.1.1 Awk script for read simulation

Pool creation

cat **annotation.gtf** | awk -F '\t' 'function output_seq(){if (!first) {if (str=="-") {print seq |& rc; rc |& getline seq} print seq} return} BEGIN{rc="rev|com"; retr="gem-retriever **genome_index.gem**"; first=1} {if (state==1) {if ($3=="exon") {print $1"\t"$7"\t"$4"\t"$5 |& retr; retr |& getline exon; seq=seq exon} else {state=0}} if ($3=="transcript") {output_seq(); first=0; split($0,s,"\";|\""); print ">"s[length(s)-1]; str=$7; seq=""; state=1}} END{output_seq()}' > transcripts.fa && cat transcripts.fa | awk '{nr=(NR-1)%2; if (nr==0) {name=substr($0,2)} else {if (pool!="") pool=pool" "; pool=pool $0}} END{print ">pool\n"pool}' > pool.fa && gem-indexer --complement-size-threshold 0 -i pool.fa -o pool

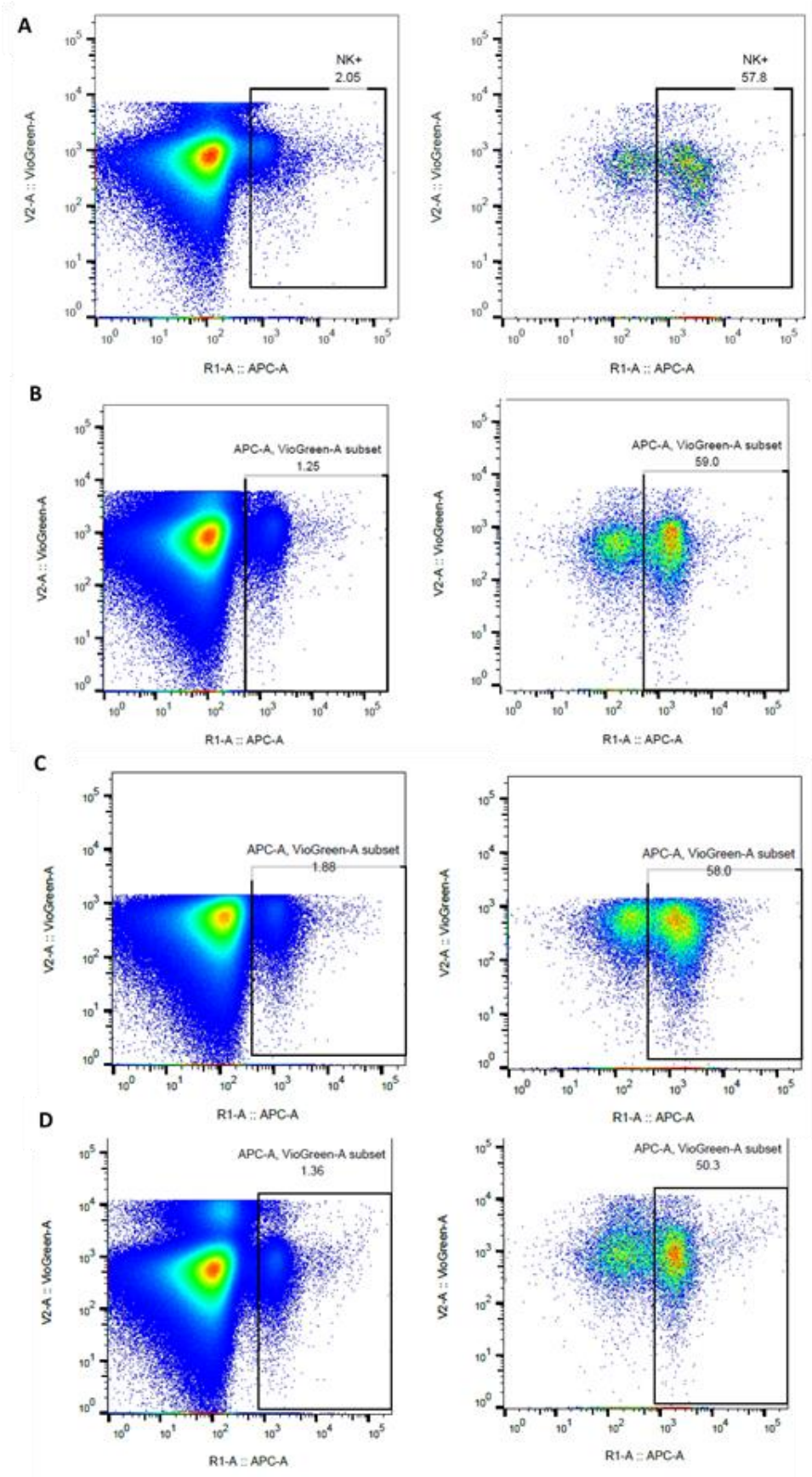
Read creation

for READ_LENGTH in **150**; do READ_NUMBER=**200000**; cat pool.fa | awk '{if (NR==2) pool_len=length($0)} END{nr=-1; tb=0; while (getline line < "transcripts.fa") {if ((++nr)%2==0) {name=substr(line,2); if (nr>0) ++tb} else {len=length(line); for (off=1;off<=len;++off) {je[tb+off]=name"."off} tb+=len}} if (tb!=pool_len) {print "ERROR" > "/dev/stderr"; exit 1} for (i=1;i<='$READ_LENGTH';++i) qua=qua"I"; retr="gem-retriever pool.gem"; while (cntr<'$READ_NUMBER') {idx=int(rand()*pool_len)+1; str=rand()>=0.5?"-":"+"; print "pool\t"str"\t"idx"\t"(idx+'$READ_LENGTH'-1) |& retr; retr |& getline seq; if (seq!=""&&seq!~" ") {print "@"je[idx]"."str"."++cntr"\n"seq"\n+\n"qua}}}' > $READ_LENGTH.fastq; done

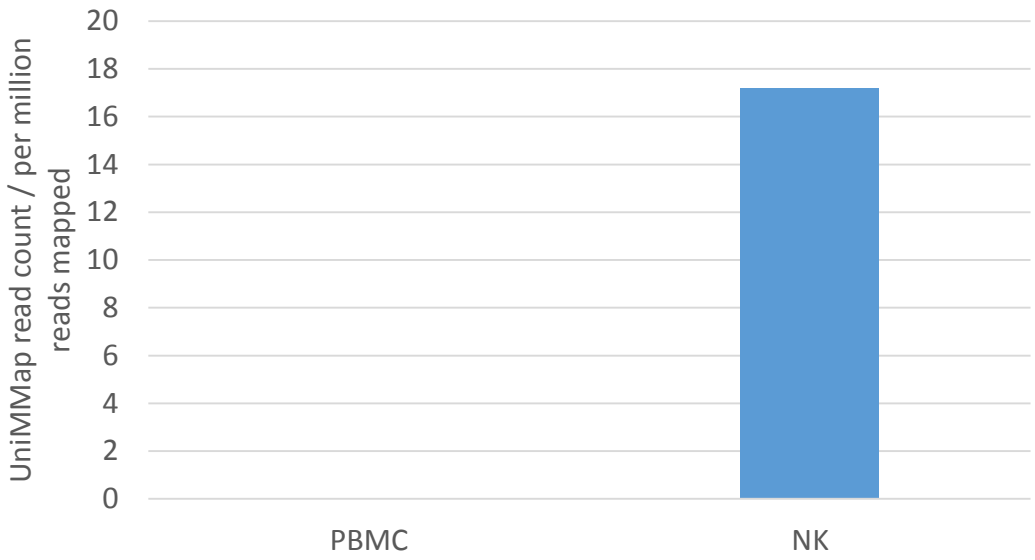
7.1.2 Per exon mappability

bedtools intersect -wb -a **mappability.bg** -b **unique_regions.bed** | grep 'chrLRC' | awk '{OFS = "\t"} ; {print $1, $2, $3, $4, $3-$2, $8}' > exon_mappability
cat exon_mappability | awk '{OFS = "\t"} ; {print $1, $2, $3, $4, $5, $4*$5, $6}' > temp; mv temp exon_mappability
awk 'NR==FNR { total[$7] += $5; total2[$7] += $6; ++n[$7] } NR>FNR { print $0, total2[$7]/total[$7]}' exon_mappability exon_mappability | awk '{print $7"\t"$8}' | uniq > exon_mappability.txt

7.2 Chapter 3 Appendix

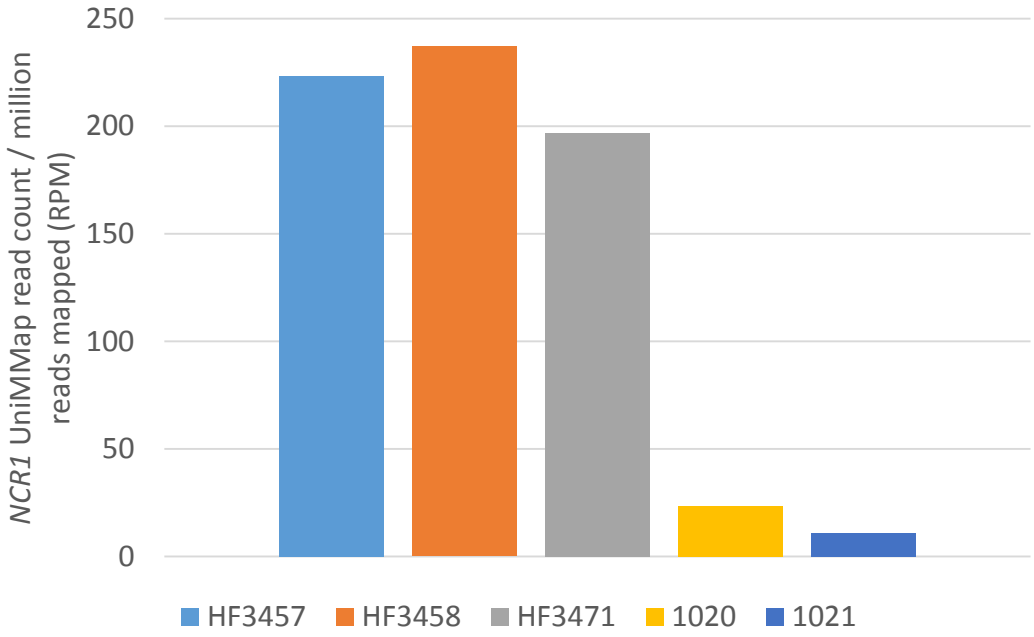


Supplementary figure 3-1. FACS plots showing percentage of NCR1+ cells in the PBMC population (left) and enriched cell population (right) for 1020 sample 1 (A), 1020 sample 2 (B), 1021 sample 1 (C) and 1021 sample 2 (D).

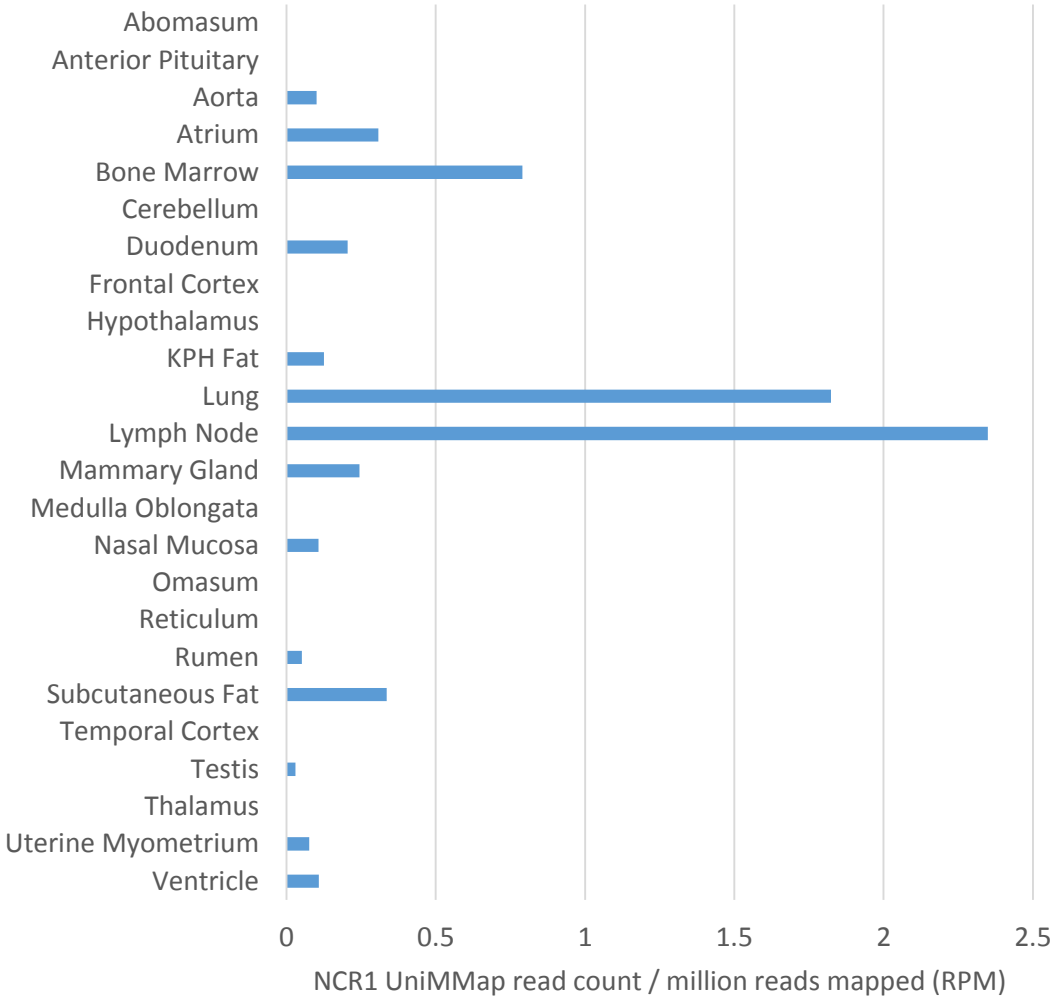


Supplementary figure 3-2. Normalised average UniMMap read count of *NCR1* from PBMC and NK RNA-Seq. UniMMap was run on two cattle PBMC and two cattle NK 150bp, paired-end RNA-Seq datasets and the resulting read counts normalised to RPM. Read counts across each cell type were averaged.

7.3 Chapter 4 Appendix



Supplementary figure 4-1. Comparison of *NCR1* transcription in five animals. UniMMap was run on RNA-Seq data from NK cells isolated from five cattle. NK cells from HF347, HF3458 and HF3471 were isolated using a combination of negative and positive sorting by FACS. MACS beads were used to positively select for NK cells from animals 1020 and 1021. Read counts from UniMMap were normalised based on the number of reads mapped.



Supplementary figure 4-2. Comparison of NCR1 transcription in 24 tissue types. RNA-Seq data was downloaded from the sequence read archive (Accession: PRJNA379574). The BioProject contains RNA-Seq from 23 Dominette tissues as well as the testis of SuperBull 99375. Read counts from UniMMap were normalised based on the number of reads mapped.

Bibliography

Adams, M D, J M Kelley, J D Gocayne, M Dubnick, M H Polymeropoulos, H Xiao, C R Merril, A Wu, B Olde, and R F Moreno. 1991. “Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project.” *Science (New York, N.Y.)* 252 (5013): 1651–56. http://www.ncbi.nlm.nih.gov/pubmed/2047873.

Allan, Alasdair J, Nicholas D Sanderson, Simon Gubbins, Shirley A Ellis, and John A Hammond. 2015. “Cattle NK Cell Heterogeneity and the Influence of MHC Class I.” *Journal of Immunology (Baltimore, Md. : 1950)* 195 (5). The American Association of Immunologists, Inc.: 2199–2206. doi:10.4049/jimmunol.1500227.

Almeida-Oliveira, Aline, Monique Smith-Carvalho, Luis Cristovão Porto, Juliana Cardoso-Oliveira, Aline dos Santos Ribeiro, Rosângela Rosa Falcão, Eliana Abdelhay, et al. 2011. “Age-Related Changes in Natural Killer Cell Receptors from Childhood through Old Age.” *Human Immunology* 72 (4). Elsevier: 319–29. doi:10.1016/J.HUMIMM.2011.01.009.

Alter, Galit, Maureen P. Martin, Nickolas Teigen, William H. Carr, Todd J. Suscovich, Arne Schneidewind, Hendrik Streeck, et al. 2007. “Differential Natural Killer Cell–mediated Inhibition of HIV-1 Replication Based on Distinct KIR/HLA Subtypes.” *The Journal of Experimental Medicine* 204 (12): 3027–36. doi:10.1084/jem.20070695.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215 (3): 403–10. doi:10.1016/S0022-2836(05)80360-2.

Ata, F A, H S al Sumry, G J King, S I Ismaili, and A A Ata. 1989. “Duration of Maternal Immunity to Peste Des Petits Ruminants.” *The Veterinary Record* 124 (22): 590–91. http://www.ncbi.nlm.nih.gov/pubmed/2773199.

Barin, Jobert G, G Christian Baldeviano, Monica V Talor, Lei Wu, SuFey Ong, DeLisa Fairweather, Djahida Bedja, et al. 2013. “Fatal Eosinophilic Myocarditis Develops in the Absence of IFN-γ and IL-17A.” *Journal of Immunology (Baltimore, Md. : 1950)* 191 (8): 4038–47. doi:10.4049/jimmunol.1301282.

Barter, Rebecca, and Bin Yu. 2017. “Superheat: A Graphical Tool for Exploring Complex Datasets Using Heatmaps.”

Baruzzo, Giacomo, Katharina E Hayer, Eun Ji Kim, Barbara Di Camillo, Garret A FitzGerald, and Gregory R Grant. 2017. “Simulation-Based Comprehensive Benchmarking of RNA-Seq Aligners.” *Nature Methods* 14 (2). NIH Public Access: 135–39. doi:10.1038/nmeth.4106.

Batista, Mariana D., Emily L. Ho, Peter J. Kuebler, Jeffrey M. Milush, Lewis L. Lanier, Esper G. Kallas, Vanessa A. York, et al. 2013. “Skewed Distribution of Natural Killer Cells in Psoriasis Skin Lesions.” *Experimental Dermatology* 22 (1): 64–66. doi:10.1111/exd.12060.

Battistini, L, G Borsellino, G Sawicki, F Poccia, M Salvetti, G Ristori, and C F Brosnan. 1997. “Phenotypic and Cytokine Analysis of Human Peripheral Blood Gamma Delta T Cells Expressing NK Cell Receptors.” *Journal of Immunology (Baltimore, Md. : 1950)* 159 (8). American Association of Immunologists: 3723–30. http://www.ncbi.nlm.nih.gov/pubmed/9378958.

Bauer, S, V Groh, J Wu, A Steinle, J H Phillips, L L Lanier, and T Spies. 1999. “Activation of NK Cells and T Cells by NKG2D, a Receptor for Stress-Inducible MICA.” *Science (New York, N.Y.)* 285 (5428): 727–29. http://www.ncbi.nlm.nih.gov/pubmed/10426993.

Belanger, S., M. M. Tu, M. M. A. Rahim, A. B. Mahmoud, R. Patel, L.-H. Tai, A. D. Troke, et al. 2012. “Impaired Natural Killer Cell Self-Education and "missing-Self" Responses in Ly49-Deficient Mice.” *Blood* 120 (3): 592–602. doi:10.1182/blood-2012-02-408732.

Benton, M. J., and P. C. J. Donoghue. 2006. “Paleontological Evidence to Date the Tree of Life.” *Molecular Biology and Evolution* 24 (1): 26–53. doi:10.1093/molbev/msl150.

Berke, G. 1995. “The CTL’s Kiss of Death.” *Cell* 81 (1): 9–12. http://www.ncbi.nlm.nih.gov/pubmed/7536631.

Beziat, V., J. A. Traherne, L. L. Liu, J. Jayaraman, M. Enqvist, S. Larsson, J. Trowsdale, and K.-J. Malmberg. 2013. “Influence of KIR Gene Copy Number on Natural Killer Cell Education.” *Blood* 121 (23): 4703–7. doi:10.1182/blood-2012-10-461442.

Béziat, Vivien, Hugo G. Hilton, Paul J. Norman, and James A. Traherne. 2017. “Deciphering the Killer-Cell Immunoglobulin-like Receptor System at Super-Resolution for Natural Killer and T-Cell Biology.” *Immunology* 150 (3). Wiley/Blackwell (10.1111): 248–64. doi:10.1111/imm.12684.

Bi, Jiacheng, and Zhigang Tian. 2017. “NK Cell Exhaustion.” *Frontiers in Immunology* 8 (June). Frontiers: 760. doi:10.3389/fimmu.2017.00760.

Bickhart, Derek M, Benjamin D Rosen, Sergey Koren, Brian L Sayre, Alex R Hastie, Saki Chan, Joyce Lee, et al. 2017. “Single-Molecule Sequencing and Chromatin Conformation Capture Enable de Novo Reference Assembly of the Domestic Goat Genome.” *Nature Genetics* 49 (4). Nature Publishing Group: 643–50. doi:10.1038/ng.3802.

Bjorkstrom, N. K., V. Beziat, F. Cichocki, L. L. Liu, J. Levine, S. Larsson, R. A. Koup, S. K. Anderson, H.-G. Ljunggren, and K.-J. Malmberg. 2012. “CD8 T Cells Express Randomly Selected KIRs with Distinct Specificities Compared with NK Cells.” *Blood* 120 (17): 3455–65. doi:10.1182/blood-2012-03-416867.

Boysen, Preben, Ingrid Olsen, Ingvild Berg, Siri Kulberg, Grethe M Johansen, and Anne K Storset. 2006. “Bovine CD2-/NKp46+ Cells Are Fully Functional Natural Killer Cells with a High Activation Status.” *BMC Immunology* 7 (April). BioMed Central: 10. doi:10.1186/1471-2172-7-10.

Brodin, P., T. Lakshmikanth, S. Johansson, K. Karre, and P. Hoglund. 2009. “The Strength of Inhibitory Input during Education Quantitatively Tunes the Functional Responsiveness of Individual Natural Killer Cells.” *Blood* 113 (11): 2434–41. doi:10.1182/blood-2008-05-156836.

Bubeník, Jan. 2003. “Tumour MHC Class I Downregulation and Immunotherapy (Review).” *Oncology Reports* 10 (6). Spandidos Publications: 2005–8. doi:10.3892/or.10.6.2005.

Bukowski, J F, J F Warner, G Dennert, and R M Welsh. 1985. “Adoptive Transfer Studies Demonstrating the Antiviral Effect of Natural Killer Cells in Vivo.” *The Journal of Experimental Medicine* 161 (1). The Rockefeller University Press: 40–52. http://www.ncbi.nlm.nih.gov/pubmed/2981954.

Bukowski, J F, B A Woda, and R M Welsh. 1984. “Pathogenesis of Murine Cytomegalovirus Infection in Natural Killer Cell-Depleted Mice.” *Journal of Virology* 52 (1): 119–28. http://www.ncbi.nlm.nih.gov/pubmed/6207307.

Bush, Stephen J, Mary E B McCulloch, Charity Muriuki, Mazdak Salavati, Gemma M Davis, Iseabail L Farquhar, Zofia M Lisowski, Alan L Archibald, David A Hume, and Emily L Clark. 2019. “Comprehensive Transcriptional Profiling of the Gastrointestinal Tract of Ruminants from Birth to Adulthood Reveals Strong Developmental Stage Specific Gene Expression.” *G3 (Bethesda, Md.)* 9 (2). Genetics Society of America: 359–73. doi:10.1534/g3.118.200810.

Caligiuri, Michael A. 2008. “Human Natural Killer Cells.” *Blood* 112 (3). American Society of Hematology: 461–69. doi:10.1182/blood-2007-09-077438.

Carretero, Marta, Claudia Cantoni, Teresa Bellón, Cristina Bottino, Roberto Biassoni, Antonio Rodríguez, Juan J. Pérez-Villar, Lorenzo Moretta, Alessandro Moretta, and Miguel López-Botet. 1997. "The CD94 and NKG2-A C-Type Lectins Covalently Assemble to Form a Natural Killer Cell Inhibitory Receptor for HLA Class I Molecules." *European Journal of Immunology* 27 (2). Wiley-Blackwell: 563–67. doi:10.1002/eji.1830270230.

Carter, D L, T M Shieh, R L Blosser, K R Chadwick, J B Margolick, J E Hildreth, J E Clements, and M C Zink. 1999. "CD56 Identifies Monocytes and Not Natural Killer Cells in Rhesus Macaques." *Cytometry* 37 (1): 41–50. http://www.ncbi.nlm.nih.gov/pubmed/10451505.

Catalfamo, Marta, and Pierre A Henkart. 2003. "Perforin and the Granule Exocytosis Cytotoxicity Pathway." *Current Opinion in Immunology* 15 (5): 522–27. http://www.ncbi.nlm.nih.gov/pubmed/14499260.

Cerboni, C., A. Zingoni, M. Cippitelli, M. Piccoli, L. Frati, and A. Santoni. 2007. "Antigen-Activated Human T Lymphocytes Express Cell-Surface NKG2D Ligands via an ATM/ATR-Dependent Mechanism and Become Susceptible to Autologous NK- Cell Lysis." *Blood* 110 (2): 606–15. doi:10.1182/blood-2006-10-052720.

Chan, Huei-Wei, Zoya B Kurago, C Andrew Stewart, Michael J Wilson, Maureen P Martin, Brian E Mace, Mary Carrington, John Trowsdale, and Charles T Lutz. 2003. "DNA Methylation Maintains Allele-Specific KIR Gene Expression in Human Natural Killer Cells." *The Journal of Experimental Medicine* 197 (2): 245–55. http://www.ncbi.nlm.nih.gov/pubmed/12538663.

Cichocki, F., R. J. Hanson, T. Lenvik, M. Pitt, V. McCullar, H. Li, S. K. Anderson, and J. S. Miller. 2009. "The Transcription Factor c-Myc Enhances KIR Gene Transcription through Direct Binding to an Upstream Distal Promoter Element." *Blood* 113 (14): 3245–53. doi:10.1182/blood-2008-07-166389.

Coca, S, J Perez-Piqueras, D Martinez, A Colmenarejo, M A Saez, C Vallejo, J A Martos, and M Moreno. 1997. "The Prognostic Significance of Intratumoral Natural Killer Cells in Patients with Colorectal Carcinoma." *Cancer* 79 (12): 2320–28. http://www.ncbi.nlm.nih.gov/pubmed/9191519.

Coles, Mark C., Christopher W. McMahon, Hisao Takizawa, and David H. Raulet. 2000. "Memory CD8 T Lymphocytes Express Inhibitory MHC-Specific Ly49 Receptors." *European Journal of Immunology* 30 (1): 236–44. doi:10.1002/1521-4141(200001)30:1<236::AID-IMMU236>3.0.CO;2-X.

Colonna, Marco. 2009. “Interleukin-22–Producing Natural Killer Cells and Lymphoid Tissue Inducer-like Cells in Mucosal Immunity.” *Immunity* 31 (1): 15–23. doi:10.1016/j.immuni.2009.06.008.

Colucci, Francesco. 2017. “The Role of KIR and HLA Interactions in Pregnancy Complications.” *Immunogenetics* 69 (8-9). Springer: 557–65. doi:10.1007/s00251-017-1003-9.

Cooper, M A, T A Fehniger, and M A Caligiuri. 2001. “The Biology of Human Natural Killer-Cell Subsets.” *Trends in Immunology* 22 (11): 633–40. http://www.ncbi.nlm.nih.gov/pubmed/11698225.

Crispe, Ian Nicholas. 2003. “Hepatic T Cells and Liver Tolerance.” *Nature Reviews Immunology* 3 (1): 51–62. doi:10.1038/nri981.

Cui, Peng, Qiang Lin, Feng Ding, Chengqi Xin, Wei Gong, Lingfang Zhang, Jianing Geng, et al. 2010. “A Comparison between Ribo-Minus RNA-Sequencing and polyA-Selected RNA-Sequencing.” *Genomics* 96 (5): 259–65. doi:10.1016/j.ygeno.2010.07.010.

D’Andrea, A, C Chang, J H Phillips, and L L Lanier. 1996. “Regulation of T Cell Lymphokine Production by Killer Cell Inhibitory Receptor Recognition of Self HLA Class I Alleles.” *The Journal of Experimental Medicine* 184 (2). Rockefeller University Press: 789–94. doi:10.1084/JEM.184.2.789.

De Re, Valli, Laura Caggiari, Mariangela De Zorzi, Ombretta Repetto, Anna Linda Zignego, Francesco Izzo, Maria Lina Tornesello, et al. 2015. “Genetic Diversity of the KIR/HLA System and Susceptibility to Hepatitis C Virus-Related Diseases.” *PloS One* 10 (2). Public Library of Science: e0117420. doi:10.1371/journal.pone.0117420.

Degli-Esposti, Mariapia. 1999. “To Die or Not to Die-the Quest of the TRAIL Receptors.” *Journal of Leukocyte Biology* 65 (5). Wiley-Blackwell: 535–42. doi:10.1002/jlb.65.5.535.

Denis, Michel, Denise L. Keen, Natalie A. Parlane, Anne K. Storset, and Bryce M. Buddle. 2007. “Bovine Natural Killer Cells Restrict the Replication of Mycobacterium Bovis in Bovine Macrophages and Enhance IL-12 Release by Infected Macrophages.” *Tuberculosis* 87 (1): 53–62. doi:10.1016/j.tube.2006.03.005.

Derrien, Thomas, Jordi Estellé, Santiago Marco Sola, David G. Knowles, Emanuele Raineri, Roderic Guigó, and Paolo Ribeca. 2012. “Fast Computation and Applications of Genome Mappability.” Edited by Christos A. Ouzounis. *PLoS ONE* 7 (1). Public Library of Science: e30377. doi:10.1371/journal.pone.0030377.

Dhanasekaran, Sakthivel, Moanaro Biswas, Ambothi R. Vignesh, R. Ramya, Gopal Dhinakar Raj, Krishnaswamy G. Tirumurugaan, Angamuthu Raja, Ranjit S. Kataria, Satya Parida, and Elankumaran Subbiah. 2014. "Toll-Like Receptor Responses to Peste Des Petits Ruminants Virus in Goats and Water Buffalo." Edited by Nagendra Hegde. *PLoS ONE* 9 (11). Public Library of Science: e111609. doi:10.1371/journal.pone.0111609.

Diallo, A., C. Minet, C. Le Goff, G. Berhe, E. Albina, G. Libeau, and T. Barrett. 2007. "The Threat of Peste Des Petits Ruminants: Progress in Vaccine Development for Disease Control." *Vaccine* 25 (30). Elsevier: 5591–97. doi:10.1016/J.VACCINE.2007.02.013.

Dighe, A S, E Richards, L J Old, and R D Schreiber. 1994. "Enhanced in Vivo Growth and Resistance to Rejection of Tumor Cells Expressing Dominant Negative IFN Gamma Receptors." *Immunity* 1 (6): 447–56. http://www.ncbi.nlm.nih.gov/pubmed/7895156.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21. doi:10.1093/bioinformatics/bts635.

Dorner, Brigitte G, Hamish R C Smith, Anthony R French, Sungjin Kim, Jennifer Poursine-Laurent, Diana L Beckman, Jeanette T Pingel, Richard A Kroczek, and Wayne M Yokoyama. 2004. "Coordinate Expression of Cytokines and Chemokines by NK Cells during Murine Cytomegalovirus Infection." *Journal of Immunology (Baltimore, Md. : 1950)* 172 (5): 3119–31. http://www.ncbi.nlm.nih.gov/pubmed/14978118.

Dunphy, S.E., C.M. Sweeney, G. Kelly, A.M. Tobin, B. Kirby, and C.M. Gardiner. 2017. "Natural Killer Cells from Psoriasis Vulgaris Patients Have Reduced Levels of Cytotoxicity Associated Degranulation and Cytokine Production." *Clinical Immunology* 177 (April). Academic Press: 43–49. doi:10.1016/J.CLIM.2015.10.004.

Fabregat, Antonio, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, et al. 2018. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 46 (D1): D649–55. doi:10.1093/nar/gkx1132.

Fernandez, Nadine C., Anne Lozier, Caroline Flament, Paola Ricciardi-Castagnoli, Dominique Bellet, Mark Suter, Michel Perricaudet, Thomas Tursz, Eugene Maraskovsky, and Laurence Zitvogel. 1999. "Dendritic Cells Directly Trigger NK Cell Functions: Cross-Talk Relevant in Innate Anti-Tumor Immune Responses in Vivo." *Nature Medicine* 5 (4): 405–11. doi:10.1038/7403.

Ferretti, M.T., M. Merlini, C. Späni, C. Gericke, N. Schweizer, G. Enzmann, B. Engelhardt, L. Kulic, T. Suter, and R.M. Nitsch. 2016. “T-Cell Brain Infiltration and Immature Antigen-Presenting Cells in Transgenic Models of Alzheimer’s Disease-like Cerebral Amyloidosis.” *Brain, Behavior, and Immunity* 54 (May). Academic Press: 211–25. doi:10.1016/J.BBI.2016.02.009.

Ferrini, Silvano, Anna Cambiaggi, Raffaella Meazza, Sabrina Sforzini, Sabrina Marciano, Maria Cristina Mingari, and Lorenzo Moretta. 1994. “T Cell Clones Expressing the Natural Killer Cell-Related p58 Receptor Molecule Display Heterogeneity in Phenotypic Properties and p58 Function.” *European Journal of Immunology* 24 (10). John Wiley & Sons, Ltd: 2294–98. doi:10.1002/eji.1830241005.

Fort, M M, M W Leach, and D M Rennick. 1998. “A Role for NK Cells as Regulators of CD4+ T Cells in a Transfer Model of Colitis.” *Journal of Immunology (Baltimore, Md. : 1950)* 161 (7): 3256–61. http://www.ncbi.nlm.nih.gov/pubmed/9759840.

Freud, Aharon G., Akihiko Yokohama, Brian Becknell, Melissa T. Lee, Hsiaoyin C. Mao, Amy K. Ferketich, and Michael A. Caligiuri. 2006. “Evidence for Discrete Stages of Human Natural Killer Cell Differentiation in Vivo.” *The Journal of Experimental Medicine* 203 (4): 1033–43. doi:10.1084/jem.20052507.

Fujihara, Mitsuhiro, Masashi Muroi, Ken-ichi Tanamoto, Tsuneo Suzuki, Hiroshi Azuma, and Hisami Ikeda. 2003. “Molecular Mechanisms of Macrophage Activation and Deactivation by Lipopolysaccharide: Roles of the Receptor Complex.” *Pharmacology & Therapeutics* 100 (2): 171–94. http://www.ncbi.nlm.nih.gov/pubmed/14609719.

Futas, Jan, and Petr Horin. 2013. “Natural Killer Cell Receptor Genes in the Family Equidae: Not Only Ly49.” *PloS One* 8 (5). Public Library of Science: e64736. doi:10.1371/journal.pone.0064736.

Gaynor, Louise M, and Francesco Colucci. 2017. “Uterine Natural Killer Cells: Functional Distinctions and Influence on Pregnancy in Humans and Mice.” *Frontiers in Immunology* 8. Frontiers Media SA: 467. doi:10.3389/fimmu.2017.00467.

Ge, Ning, Yasuhiko Nishioka, Yoichi Nakamura, Yoshio Okano, Kazuo Yoneda, Hirohisa Ogawa, Akemi Sugita, Hiroaki Yanagawa, and Saburo Sone. 2004. “Synthesis and Secretion of Interleukin-15 by Freshly Isolated Human Bronchial Epithelial Cells.” *International Archives of Allergy and Immunology* 135 (3): 235–42. doi:10.1159/000081309.

Gibbs, Richard A, Jeffrey Rogers, Michael G Katze, Roger Bumgarner, George M Weinstock, Elaine R Mardis, Karin A Remington, et al. 2007. "Evolutionary and Biomedical Insights from the Rhesus Macaque Genome." *Science* 316 (5822): 222 LP – 234. doi:10.1126/science.1139247.

Glenn, Travis C. 2011. "Field Guide to next-Generation DNA Sequencers." *Molecular Ecology Resources* 11 (5). Wiley/Blackwell (10.1111): 759–69. doi:10.1111/j.1755-0998.2011.03024.x.

Gómez-Lozano, Natalia, Ernesto Estefanía, Fionnuala Williams, Iris Halfpenny, Derek Middleton, Rosario Solís, and Carlos Vilches. 2005. "The Silent *KIR3DP1* Gene (CD158c) Is Transcribed and Might Encode a Secreted Receptor in a Minority of Humans, in Whom the *KIR3DP1* , *KIR2DL4* and *KIR3DL1* / *KIR3DS1* Genes Are Duplicated." *European Journal of Immunology* 35 (1): 16–24. doi:10.1002/eji.200425493.

Gosselin, P, A P Makrigiannis, R Nalewaik, and S K Anderson. 2000. "Characterization of the Ly49I Promoter." *Immunogenetics* 51 (4-5): 326–31. http://www.ncbi.nlm.nih.gov/pubmed/10803845.

Graham, Elizabeth M., Michelle L. Thom, Chris J. Howard, Preben Boysen, Anne K. Storset, Paul Sopp, and Jayne C. Hope. 2009. "Natural Killer Cell Number and Phenotype in Bovine Peripheral Blood Is Influenced by Age." *Veterinary Immunology and Immunopathology* 132 (2-4). Elsevier: 101–8. doi:10.1016/J.VETIMM.2009.05.002.

Grégoire, Claude, Lionel Chasson, Carmelo Luci, Elena Tomasello, Frédéric Geissmann, Eric Vivier, and Thierry Walzer. 2007. "The Trafficking of Natural Killer Cells." *Immunological Reviews* 220 (1). Wiley/Blackwell (10.1111): 169–82. doi:10.1111/j.1600-065X.2007.00563.x.

Guethlein, Lisbeth A., Laurent Abi-Rached, John A. Hammond, and Peter Parham. 2007. "The Expanded Cattle KIR Genes Are Orthologous to the Conserved Single-Copy KIR3DX1 Gene of Primates." *Immunogenetics* 59 (6). Springer-Verlag: 517–22. doi:10.1007/s00251-007-0214-x.

Guia, S., B. N. Jaeger, S. Piatek, S. Mailfert, T. Trombik, A. Fenis, N. Chevrier, et al. 2011. "Confinement of Activating Receptors at the Plasma Membrane Controls Natural Killer Cell Tolerance." *Science Signaling* 4 (167): ra21–ra21. doi:10.1126/scisignal.2001608.

Gutierrez-A, N. 2018. "Economic Constraints on Sheep and Goat Production in Developing Countries." Accessed July 21. http://www.fao.org/docrep/009/ah221e/AH221E13.htm.

Guzman, Efrain, James R. Birch, and Shirley A. Ellis. 2010. "Cattle MIC Is a Ligand for the Activating NK Cell Receptor NKG2D." *Veterinary Immunology and Immunopathology* 136 (3-4). Elsevier: 227–34. doi:10.1016/J.VETIMM.2010.03.012.

Hamilton, Carly A., Suman Mahan, Charlotte R. Bell, Bernardo Villarreal-Ramos, Bryan Charleston, Gary Entrican, and Jayne C. Hope. 2017. "Frequency and Phenotype of Natural Killer Cells and Natural Killer Cell Subsets in Bovine Lymphoid Compartments and Blood." *Immunology* 151 (1). John Wiley & Sons, Ltd (10.1111): 89–97. doi:10.1111/imm.12708.

Hammond, John A, Lisbeth A Guethlein, Laurent Abi-Rached, Achim K Moesta, and Peter Parham. 2009. "Evolution and Survival of Marine Carnivores Did Not Require a Diversity of Killer Cell Ig-like Receptors or Ly49 NK Cell Receptors." *Journal of Immunology (Baltimore, Md. : 1950)* 182 (6). NIH Public Access: 3618–27. doi:10.4049/jimmunol.0803026.

Hansen, Kasper D, Steven E Brenner, and Sandrine Dudoit. 2010. "Biases in Illumina Transcriptome Sequencing Caused by Random Hexamer Priming." *Nucleic Acids Research* 38 (12). Oxford University Press: e131. doi:10.1093/nar/gkq224.

Hayakawa, Yoshihiro, Nicholas D. Huntington, Stephen L. Nutt, and Mark J. Smyth. 2006. "Functional Subsets of Mouse Natural Killer Cells." *Immunological Reviews* 214 (1). Wiley/Blackwell (10.1111): 47–55. doi:10.1111/j.1600-065X.2006.00454.x.

He, Yuke, and Zhigang Tian. 2017. "NK Cell Education via Nonclassical MHC and Non-MHC Ligands." *Cellular & Molecular Immunology* 14 (4). Nature Publishing Group: 321–30. doi:10.1038/cmi.2016.26.

Heaney, J, T Barrett, and S L Cosby. 2002. "Inhibition of in Vitro Leukocyte Proliferation by Morbilliviruses." *Journal of Virology* 76 (7). American Society for Microbiology Journals: 3579–84. doi:10.1128/JVI.76.7.3579-3584.2002.

Held, Werner, and Béatrice Kunz. 1998. "An Allele-Specific, Stochastic Gene Expression Process Controls the Expression of multipleLy49family Genes and Generates a Diverse, MHC-Specific NK Cell Receptor Repertoire." *European Journal of Immunology* 28 (8): 2407–16. doi:10.1002/(SICI)1521-4141(199808)28:08<2407::AID-IMMU2407>3.0.CO;2-D.

Herbert, Rebecca, Jana Baron, Carrie Batten, Michael Baron, and Geraldine Taylor. 2014. "Recombinant Adenovirus Expressing the Haemagglutinin of Peste Des Petits Ruminants Virus (PPRV) Protects Goats against Challenge with Pathogenic Virus; a DIVA Vaccine for PPR." *Veterinary Research* 45 (1): 24. doi:10.1186/1297-9716-45-24.

Herbert, Zachary T., Jamie P. Kershner, Vincent L. Butty, Jyothi Thimmapuram, Sulbha Choudhari, Yuriy O. Alekseyev, Jun Fan, et al. 2018. “Cross-Site Comparison of Ribosomal Depletion Kits for Illumina RNAseq Library Construction.” *BMC Genomics* 19 (1). BioMed Central: 199. doi:10.1186/s12864-018-4585-1.

Hiby, Susan E., James J. Walker, Kevin M. O’Shaughnessy, Christopher W.G. Redman, Mary Carrington, John Trowsdale, and Ashley Moffett. 2004. “Combinations of Maternal KIR and Fetal HLA-C Genes Influence the Risk of Preeclampsia and Reproductive Success.” *The Journal of Experimental Medicine* 200 (8): 957–65. doi:10.1084/jem.20041214.

Hiendleder, S, H Lewalski, R Wassmuth, and A Janke. 1998. “The Complete Mitochondrial DNA Sequence of the Domestic Sheep (Ovis Aries) and Comparison with the Other Major Ovine Haplotype.” *Journal of Molecular Evolution* 47 (4): 441–48. http://www.ncbi.nlm.nih.gov/pubmed/9767689.

Higuchi, D A, P Cahan, J Gao, S T Ferris, J Poursine-Laurent, T A Graubert, and W M Yokoyama. 2010. “Structural Variation of the Mouse Natural Killer Gene Complex.” *Genes and Immunity* 11 (8). NIH Public Access: 637–48. doi:10.1038/gene.2010.48.

Ho, E L, J W Heusel, M G Brown, K Matsumoto, A A Scalzo, and W M Yokoyama. 1998. “Murine Nkg2d and Cd94 Are Clustered within the Natural Killer Complex and Are Expressed Independently in Natural Killer Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (11): 6320–25. http://www.ncbi.nlm.nih.gov/pubmed/9600963.

Hoelsbrekken, Sigurd E, Øyvind Nylenna, Per C Saether, Imer O Slettedal, James C Ryan, Sigbjørn Fossum, and Erik Dissen. 2003. “Cutting Edge: Molecular Cloning of a Killer Cell Ig-like Receptor in the Mouse and Rat.” *Journal of Immunology (Baltimore, Md. : 1950)* 170 (5): 2259–63. http://www.ncbi.nlm.nih.gov/pubmed/12594244.

Hsu, Katharine C, Shohei Chida, Daniel E Geraghty, and Bo Dupont. 2002. “The Killer Cell Immunoglobulin-like Receptor (KIR) Genomic Region: Gene-Order, Haplotypes and Allelic Polymorphism.” *Immunological Reviews* 190 (December): 40–52. doi:10.1034/J.1600-065X.2002.19004.X.

Huard, Bertrand, and Lars Karlsson. 2000. “KIR Expression on Self-Reactive CD8+ T Cells Is Controlled by T-Cell Receptor Engagement.” *Nature* 403 (6767): 325–28. doi:10.1038/35002105.

Huhn, Oisin, Olympe Chazara, Martin Ivarsson, Christelle Retiere, Tim Venkatesan, Hormas Ghadially, Ashley Moffett, Andrew Sharkey, and Francesco Colucci. 2018. “High-Resolution Genetic and Phenotypic Analysis of KIR2DL1 Alleles and Their Association with Pre-Eclampsia.” *bioRxiv*, May. Cold Spring Harbor Laboratory, 330803. doi:10.1101/330803.

Ishigami, S, S Natsugoe, K Tokuda, A Nakajo, X Che, H Iwashige, K Aridome, S Hokita, and T Aikou. 2000. “Prognostic Value of Intratumoral Natural Killer Cells in Gastric Carcinoma.” *Cancer* 88 (3): 577–83. http://www.ncbi.nlm.nih.gov/pubmed/10649250.

Jain, Miten, Sergey Koren, Karen H Miga, Josh Quick, Arthur C Rand, Thomas A Sasani, John R Tyson, et al. 2018. “Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads.” *Nature Biotechnology* 36 (4). Nature Publishing Group: 338–45. doi:10.1038/nbt.4060.

Jain, Miten, John R. Tyson, Matthew Loose, Camilla L.C. Ip, David A. Eccles, Justin O’Grady, Sunir Malla, et al. 2017. “MinION Analysis and Reference Consortium: Phase 2 Data Release and Analysis of R9.0 Chemistry.” *F1000Research* 6 (May): 760. doi:10.12688/f1000research.11354.1.

Jamieson, Amanda M, Patricia Isnard, Jeffrey R Dorfman, Mark C Coles, and David H Raulet. 2004. “Turnover and Proliferation of NK Cells in Steady State and Lymphopenic Conditions.” *Journal of Immunology (Baltimore, Md. : 1950)* 172 (2): 864–70. http://www.ncbi.nlm.nih.gov/pubmed/14707057.

Jones, D. C., V. Kosmoliaptsis, R. Apps, N. Lapaque, I. Smith, A. Kono, C. Chang, et al. 2011. “HLA Class I Allelic Sequence and Conformation Regulate Leukocyte Ig-Like Receptor Binding.” *The Journal of Immunology* 186 (5): 2990–97. doi:10.4049/jimmunol.1003078.

Jost, Stephanie, Heloise Quillay, Jeff Reardon, Eric Peterson, Rachel P. Simmons, Blair A. Parry, Nancy N. P. Bryant, William D. Binder, and Marcus Altfeld. 2011. “Changes in Cytokine Levels and NK Cell Activation Associated with Influenza.” Edited by Douglas F. Nixon. *PLoS ONE* 6 (9): e25060. doi:10.1371/journal.pone.0025060.

Kadri, Nadir, Thuy Luu Thanh, and Petter Höglund. 2015. “Selection, Tuning, and Adaptation in Mouse NK Cell Education.” *Immunological Reviews* 267 (1): 167–77. doi:10.1111/imr.12330.

Kanda, T, T Tanaka, K Sekiguchi, Y Seta, M Kurimoto, J E Wilson McManus, R Nagai, D Yang, B M McManus, and I Kobayashi. 2000. “Effect of Interleukin-18 on Viral Myocarditis: Enhancement of Interferon-Gamma and Natural Killer Cell Activity.” *Journal of Molecular and Cellular Cardiology* 32 (12): 2163–71. doi:10.1006/jmcc.2000.1242.

Kelley, James, Lutz Walter, and John Trowsdale. 2005. “Comparative Genomics of Natural Killer Cell Receptor Gene Clusters.” *PLoS Genetics* 1 (2): e27. doi:10.1371/journal.pgen.0010027.

Khakoo, S. I., Chloe L Thio, Maureen P Martin, Collin R Brooks, Xiaojiang Gao, Jacquie Astemborski, Jie Cheng, et al. 2004. “HLA and NK Cell Inhibitory Receptor Genes in Resolving Hepatitis C Virus Infection.” *Science* 305 (5685): 872–74. doi:10.1126/science.1097670.

Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg. 2013. “TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions.” *Genome Biology* 14 (4). BioMed Central: R36. doi:10.1186/gb-2013-14-4-r36.

Kim, Sungjin, John B. Sunwoo, Liping Yang, Taewoong Choi, Yun-Jeong Song, Anthony R. French, Anna Vlahiotis, et al. 2008. “HLA Alleles Determine Differences in Human Natural Killer Cell Responsiveness and Potency.” *Proceedings of the National Academy of Sciences* 105 (8). National Academy of Sciences: 3053–58. doi:10.1073/PNAS.0712229105.

Kinsella, R. J., A. Kahari, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, et al. 2011. “Ensembl BioMarts: A Hub for Data Retrieval across Taxonomic Space.” *Database* 2011 (0): bar030–bar030. doi:10.1093/database/bar030.

Krumsiek, J., R. Arnold, and T. Rattei. 2007. “Gepard: A Rapid and Sensitive Tool for Creating Dotplots on Genome Scale.” *Bioinformatics* 23 (8). Oxford University Press: 1026–28. doi:10.1093/bioinformatics/btm039.

Kubota, A, S Kubota, S Lohwasser, D L Mager, and F Takei. 1999. “Diversity of NK Cell Receptor Repertoire in Adult and Neonatal Mice.” *Journal of Immunology (Baltimore, Md. : 1950)* 163 (1): 212–16. http://www.ncbi.nlm.nih.gov/pubmed/10384118.

Küçük, Can, Xiaozhou Hu, Qiang Gong, Bei Jiang, Adam Cornish, Philippe Gaulard, Timothy McKeithan, and Wing C. Chan. 2016. “Diagnostic and Biological Significance of KIR Expression Profile Determined by RNA-Seq in Natural Killer/T-Cell Lymphoma.” *The American Journal of Pathology* 186 (6): 1435–41. doi:10.1016/j.ajpath.2016.02.011.

Kulski, Jerzy K, Takashi Shiina, Tatsuya Anzai, Sakae Kohara, and Hidetoshi Inoko. 2002. “Comparative Genomic Analysis of the MHC: The Evolution of Class I Duplication Blocks, Diversity and Complexity from Shark to Man.” *Immunological Reviews* 190 (December): 95–122. http://www.ncbi.nlm.nih.gov/pubmed/12493009.

Kumar, Naveen, Sunil Maherchandani, Sudhir Kumar Kashyap, Shoor Vir Singh, Shalini Sharma, Kundan Kumar Chaubey, and Hinh Ly. 2014. “Peste Des Petits Ruminants Virus Infection of Small Ruminants: A Comprehensive Review.” *Viruses* 6 (6). Multidisciplinary Digital Publishing Institute (MDPI): 2287–2327. doi:10.3390/v6062287.

Kumar, P., B. N. Tripathi, A. K. Sharma, R. Kumar, B. P. Sreenivasa, R. P. Singh, P. Dhar, and S. K. Bandyopadhyay. 2004. “Pathological and Immunohistochemical Study of Experimental Peste Des Petits Ruminants Virus Infection in Goats.” *Journal of Veterinary Medicine Series B* 51 (4): 153–59. doi:10.1111/j.1439-0450.2004.00747.x.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10 (3). BioMed Central: R25. doi:10.1186/gb-2009-10-3-r25.

Lanier, L L, A M Le, C I Civin, M R Loken, and J H Phillips. 1986. “The Relationship of CD16 (Leu-11) and Leu-19 (NKH-1) Antigen Expression on Human Peripheral Blood NK Cells and Cytotoxic T Lymphocytes.” *Journal of Immunology (Baltimore, Md. : 1950)* 136 (12). American Association of Immunologists: 4480–86. http://www.ncbi.nlm.nih.gov/pubmed/3086432.

Lanier, L L, and J H Phillips. 1996. “Inhibitory MHC Class I Receptors on NK Cells and T Cells.” *Immunology Today* 17 (2): 86–91. http://www.ncbi.nlm.nih.gov/pubmed/8808056.

Lanier, Lewis L. 2003. “Natural Killer Cell Receptor Signaling.” *Current Opinion in Immunology* 15 (3): 308–14. http://www.ncbi.nlm.nih.gov/pubmed/12787756.

Lanier, Lewis L., Brian C. Corliss, Jun Wu, Clement Leong, and Joseph H. Phillips. 1998. “Immunoreceptor DAP12 Bearing a Tyrosine-Based Activation Motif Is Involved in Activating NK Cells.” *Nature* 391 (6668): 703–7. doi:10.1038/35642.

Lappalainen, Tuuli, Michael Sammeth, Marc R Friedländer, Peter A C ’t Hoen, Jean Monlong, Manuel A Rivas, Mar Gonzàlez-Porta, et al. 2013. “Transcriptome and Genome Sequencing Uncovers Functional Variation in Humans.” *Nature* 501 (7468). NIH Public Access: 506–11. doi:10.1038/nature12531.

Lassen, Matthew G, John R Lukens, Joseph S Dolina, Michael G Brown, and Young S Hahn. 2010. “Intrahepatic IL-10 Maintains NKG2A+Ly49- Liver NK Cells in a Functionally Hyporesponsive State.” *Journal of Immunology (Baltimore, Md. : 1950)* 184 (5). NIH Public Access: 2693–2701. doi:10.4049/jimmunol.0901362.

Le Garff-Tavernier, Magali, Vivien Béziat, Julie Decocq, Virginie Siguret, Frédérique Gandjbakhch, Eric Pautas, Patrice Debré, Hélène Merle-Beral, and Vincent Vieillard. 2010. “Human NK Cells Display Major Phenotypic and Functional Changes over the Life Span.” *Aging Cell* 9 (4): 527–35. doi:10.1111/j.1474-9726.2010.00584.x.

Leinonen, Rasko, Hideaki Sugawara, Martin Shumway, and International Nucleotide Sequence Database Collaboration. 2011. “The Sequence Read Archive.” *Nucleic Acids Research* 39 (Database issue). Oxford University Press: D19–21. doi:10.1093/nar/gkq1019.

León, Francisco, Ernesto Roldán, Laura Sanchez, Cristina Camarero, Alfredo Bootello, and Garbiñe Roy. 2003. “Human Small-Intestinal Epithelium Contains Functional Natural Killer Lymphocytes.” *Gastroenterology* 125 (2): 345–56. http://www.ncbi.nlm.nih.gov/pubmed/12891535.

Li, H., and R. Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60. doi:10.1093/bioinformatics/btp324.

Li, H, P W Wright, M McCullen, and S K Anderson. 2016. “Characterization of KIR Intermediate Promoters Reveals Four Promoter Types Associated with Distinct Expression Patterns of KIR Subtypes.” *Genes and Immunity* 17 (1). NIH Public Access: 66–74. doi:10.1038/gene.2015.56.

Li, Hongchuan, Véronique Pascal, Maureen P. Martin, Mary Carrington, and Stephen K. Anderson. 2008. “Genetic Control of Variegated KIR Gene Expression: Polymorphisms of the Bi-Directional KIR3DL1 Promoter Are Associated with Distinct Frequencies of Gene Expression.” Edited by Derry C. Roopenian. *PLoS Genetics* 4 (11): e1000254. doi:10.1371/journal.pgen.1000254.

Liang, Hui-Ling, Shu-Juan Ma, and Hong-Zhuan Tan. 2017. “Association between Killer Cell Immunoglobulin-like Receptor (KIR) Polymorphisms and Systemic Lupus Erythematosus (SLE) in Populations: A PRISMA-Compliant Meta-Analysis.” *Medicine* 96 (10). Wolters Kluwer Health: e6166. doi:10.1097/MD.0000000000006166.

Liao, Y., G. K. Smyth, and W. Shi. 2014. “featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30. doi:10.1093/bioinformatics/btt656.

Ljunggren, H G, and K Kärre. 1985. “Host Resistance Directed Selectively against H-2-Deficient Lymphoma Variants. Analysis of the Mechanism.” *The Journal of Experimental Medicine* 162 (6). The Rockefeller University Press: 1745–59. http://www.ncbi.nlm.nih.gov/pubmed/3877776.

Lockhart, David J., Helin Dong, Michael C. Byrne, Maximillian T. Follettie, Michael V. Gallo, Mark S. Chee, Michael Mittmann, et al. 1996. “Expression Monitoring by Hybridization to High-Density Oligonucleotide Arrays.” *Nature Biotechnology* 14 (13): 1675–80. doi:10.1038/nbt1296-1675.

Lopez-Verges, S., J. M. Milush, S. Pandey, V. A. York, J. Arakawa-Hoyt, H. Pircher, P. J. Norris, D. F. Nixon, and L. L. Lanier. 2010. “CD57 Defines a Functionally Distinct Population of Mature NK Cells in the Human CD56dimCD16+ NK-Cell Subset.” *Blood* 116 (19). American Society of Hematology: 3865–74. doi:10.1182/blood-2010-04-282301.

Lutz, Charles T., Mikel B. Moore, Sarah Bradley, Brent J. Shelton, and Susan K. Lutgendorf. 2005. “Reciprocal Age Related Change in Natural Killer Cell Receptors for MHC Class I.” *Mechanisms of Ageing and Development* 126 (6-7): 722–31. doi:10.1016/j.mad.2005.01.004.

Maasho, Kerima, Jessica Opoku-Anane, Alina I Marusina, John E Coligan, and Francisco Borrego. 2005. “NKG2D Is a Costimulatory Receptor for Human Naive CD8+ T Cells.” *Journal of Immunology (Baltimore, Md. : 1950)* 174 (8): 4480–84. http://www.ncbi.nlm.nih.gov/pubmed/15814668.

Mahmoud, Medhat, Marek Zywicki, Tomasz Twardowski, and Wojciech M. Karlowski. 2017. “Efficiency of PacBio Long Read Correction by 2nd Generation Illumina Sequencing.” *Genomics*, December. Academic Press. doi:10.1016/J.YGENO.2017.12.011.

Marco-Sola, Santiago, Michael Sammeth, Roderic Guigó, and Paolo Ribeca. 2012. “The GEM Mapper: Fast, Accurate and Versatile Alignment by Filtration.” *Nature Methods* 9 (12). Nature Publishing Group: 1185–88. doi:10.1038/nmeth.2221.

Martin, Maureen P, Ying Qi, Xiaojiang Gao, Eriko Yamada, Jeffrey N Martin, Florencia Pereyra, Sara Colombo, et al. 2007. “Innate Partnership of HLA-B and KIR3DL1 Subtypes against HIV-1.” *Nature Genetics* 39 (6): 733–40. doi:10.1038/ng2035.

McCullen, M V, H Li, M Cam, S K Sen, D W McVicar, and S K Anderson. 2016. “Analysis of Ly49 Gene Transcripts in Mature NK Cells Supports a Role for the Pro1 Element in Gene Activation, Not Gene Expression.” *Genes and Immunity* 17 (6). NIH Public Access: 349–57. doi:10.1038/gene.2016.31.

McLoughlin, Kirsten E, Nicolas C Nalpas, Kévin Rue-Albrecht, John A Browne, David A Magee, Kate E Killick, Stephen D E Park, et al. 2014. "RNA-Seq Transcriptional Profiling of Peripheral Blood Leukocytes from Cattle Infected with Mycobacterium Bovis." *Frontiers in Immunology* 5. Frontiers Media SA: 396. doi:10.3389/fimmu.2014.00396.

McQueen, Karina L., Brian T. Wilhelm, Kristin D. Harden, and Dixie L. Mager. 2002. "Evolution of NK Receptors: A Single Ly49 and Multiple KIR Genes in the Cow." *European Journal of Immunology* 32 (3): 810. doi:10.1002/1521-4141(200203)32:3<810::AID-IMMU810>3.0.CO;2-P.

Meresse, Bertrand, Shane A Curran, Cezary Ciszewski, Gerasim Orbelyan, Mala Setty, Govind Bhagat, Leanne Lee, et al. 2006. "Reprogramming of CTLs into Natural Killer-like Cells in Celiac Disease." *The Journal of Experimental Medicine* 203 (5). Rockefeller University Press: 1343–55. doi:10.1084/jem.20060028.

Middleton, Derek, and Faviel Gonzelez. 2010. "The Extensive Polymorphism of KIR Genes." *Immunology* 129 (1). Wiley-Blackwell: 8–19. doi:10.1111/j.1365-2567.2009.03208.x.

Migalska, M, A Sebastian, M Konczal, P Kotlík, and J Radwan. 2017. "De Novo Transcriptome Assembly Facilitates Characterisation of Fast-Evolving Gene Families, MHC Class I in the Bank Vole (Myodes Glareolus)." *Heredity* 118 (4). Nature Publishing Group: 348–57. doi:10.1038/hdy.2016.105.

Miller, J S, V McCullar, Monia Draghi, Fotini Partheniou, Ann-Margaret Little, and Peter Parham. 2001. "Human Natural Killer Cells with Polyclonal Lectin and Immunoglobulinlike Receptors Develop from Single Hematopoietic Stem Cells with Preferential Expression of NKG2A and KIR2DL2/L3/S2." *Blood* 98 (3). American Society of Hematology: 705–13. doi:10.1182/blood.v98.3.705.

Mingari, M C, F Schiavetti, M Ponte, C Vitale, E Maggi, S Romagnani, J Demarest, G Pantaleo, A S Fauci, and L Moretta. 1996. "Human CD8+ T Lymphocyte Subsets That Express HLA Class I-Specific Inhibitory Receptors Represent Oligoclonally or Monoclonally Expanded Cell Populations." *Proceedings of the National Academy of Sciences of the United States of America* 93 (22). National Academy of Sciences: 12433–38. doi:10.1073/PNAS.93.22.12433.

Moffett-King, Ashley. 2002. "Natural Killer Cells and Pregnancy." *Nature Reviews Immunology* 2 (9). Nature Publishing Group: 656–63. doi:10.1038/nri886.

Morandi, Barbara, Gwenola Bougras, William A. Muller, Guido Ferlazzo, and Christian Münz. 2006. "NK Cells of Human Secondary Lymphoid Tissues Enhance T Cell Polarizationvia IFN-γ Secretion." *European Journal of Immunology* 36 (9): 2394–2400. doi:10.1002/eji.200636290.

Morandi, Barbara, Lorenzo Mortara, Laura Chiossone, Roberto S. Accolla, Maria Cristina Mingari, Lorenzo Moretta, Alessandro Moretta, and Guido Ferlazzo. 2012. “Dendritic Cell Editing by Activated Natural Killer Cells Results in a More Protective Cancer-Specific Immune Response.” Edited by Francesco Dieli. *PLoS ONE* 7 (6): e39170. doi:10.1371/journal.pone.0039170.

Mortazavi, Ali, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. 2008. “Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq.” *Nature Methods* 5 (7). Nature Publishing Group: 621–28. doi:10.1038/nmeth.1226.

Nakajima, H, H Tomiyama, and M Takiguchi. 1995. “Inhibition of Gamma Delta T Cell Recognition by Receptors for MHC Class I Molecules.” *Journal of Immunology (Baltimore, Md. : 1950)* 155 (9). American Association of Immunologists: 4139–42. http://www.ncbi.nlm.nih.gov/pubmed/7594567.

Namekawa, Takashi, Ulf G. Wagner, Jorg J. Goronzy, and Cornelia M. Weyand. 1998. “Functional Subsets of CD4 T Cells in Rheumatoid Synovitis.” *Arthritis & Rheumatism* 41 (12): 2108–16. doi:10.1002/1529-0131(199812)41:12<2108::AID-ART5>3.0.CO;2-Q.

Nandakumar, Subhadra, Stacie N Woolard, Dorothy Yuan, Barry T Rouse, and Uday Kumaraguru. 2008. “Natural Killer Cells as Novel Helpers in Anti-Herpes Simplex Virus Immune Response.” *Journal of Virology* 82 (21). American Society for Microbiology: 10820–31. doi:10.1128/JVI.00365-08.

Naper, Christian, Shigenari Hayashi, Guro Løvik, Lise Kveberg, Eréne C Niemi, Bent Rolstad, Erik Dissen, James C Ryan, and John T Vaage. 2002. “Characterization of a Novel Killer Cell Lectin-like Receptor (KLRH1) Expressed by Alloreactive Rat NK Cells.” *Journal of Immunology (Baltimore, Md. : 1950)* 168 (10): 5147–54. http://www.ncbi.nlm.nih.gov/pubmed/11994469.

Nedvetzki, Shlomo, Stefanie Sowinski, Robert A Eagle, James Harris, Frédéric Vély, Daniela Pende, John Trowsdale, Eric Vivier, Siamon Gordon, and Daniel M Davis. 2007. “Reciprocal Regulation of Human Natural Killer Cells and Macrophages Associated with Distinct Immune Synapses.” *Blood* 109 (9). American Society of Hematology: 3776–85. doi:10.1182/blood-2006-10-052977.

Nenadic, Oleg, and Michael Greenacre. 2007. “Correspondence Analysis in R, with Two- and Three-Dimensional Graphics: The ca Package.” *Journal of Statistical Software* 20 (3): 1–13. doi:10.18637/jss.v02

Nicholson, Allen W. 2014. “Ribonuclease III Mechanisms of Double-Stranded RNA Cleavage.” *Wiley Interdisciplinary Reviews: RNA* 5 (1): 31–48. doi:10.1002/wrna.1195.

O'Callaghan, Christopher A. 2000. "Natural Killer Cell Surveillance of Intracellular Antigen Processing Pathways Mediated by Recognition of HLA-E and Qa-1b by CD94/NKG2 Receptors." *Microbes and Infection* 2 (4). Elsevier Masson: 371–80. doi:10.1016/S1286-4579(00)00330-0.

Ong, SuFey, Davinna L Ligons, Jobert G Barin, Lei Wu, Monica V Talor, Nicola Diny, Jillian A Fontes, et al. 2015. "Natural Killer Cells Limit Cardiac Inflammation and Fibrosis by Halting Eosinophil Infiltration." *The American Journal of Pathology* 185 (3): 847–61. doi:10.1016/j.ajpath.2014.11.023.

Pando, Marcelo J, Clair M Gardiner, Michael Gleimer, Karina L McQueen, and Peter Parham. 2003. "The Protein Made from a Common Allele of KIR3DL1 (3DL1*004) Is Poorly Expressed at Cell Surfaces due to Substitution at Positions 86 in Ig Domain 0 and 182 in Ig Domain 1." *Journal of Immunology (Baltimore, Md. : 1950)* 171 (12): 6640–49. http://www.ncbi.nlm.nih.gov/pubmed/14662867.

Parham, Peter. 2005. "MHC Class I Molecules and Kirs in Human History, Health and Survival." *Nature Reviews Immunology* 5 (3): 201–14. doi:10.1038/nri1570.

Parham, Peter, and Ashley Moffett. 2013. "Variable NK Cell Receptors and Their MHC Class I Ligands in Immunity, Reproduction and Human Evolution." *Nature Reviews Immunology* 13 (2). Nature Publishing Group: 133–44. doi:10.1038/nri3370.

Parkhomchuk, Dmitri, Tatiana Borodina, Vyacheslav Amstislavskiy, Maria Banaru, Linda Hallen, Sylvia Krobitsch, Hans Lehrach, and Alexey Soldatov. 2009. "Transcriptome Analysis by Strand-Specific Sequencing of Complementary DNA." *Nucleic Acids Research* 37 (18). Oxford University Press: e123–e123. doi:10.1093/nar/gkp596.

Parks, Olivia B, Derek A Pociask, Zerina Hodzic, Jay K Kolls, and Misty Good. 2015. "Interleukin-22 Signaling in the Regulation of Intestinal Health and Disease." *Frontiers in Cell and Developmental Biology* 3. Frontiers Media SA: 85. doi:10.3389/fcell.2015.00085.

Pascal, Véronique, Michael J. Stulberg, and Stephen K. Anderson. 2006. "Regulation of Class I Major Histocompatibility Complex Receptor Expression in Natural Killer Cells: One Promoter Is Not Enough!" *Immunological Reviews* 214 (1). John Wiley & Sons, Ltd (10.1111): 9–21. doi:10.1111/j.1600-065X.2006.00452.x.

Pasman, Yfke, Daniele Merico, and Azad K. Kaushik. 2017. "Preferential Expression of IGHV and IGHD Encoding Antibodies with Exceptionally Long CDR3H and a Rapid Global Shift in Transcriptome Characterizes Development of Bovine Neonatal Immunity." *Developmental & Comparative Immunology* 67 (February): 495–507. doi:10.1016/j.dci.2016.08.020.

Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford. 2017. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression.” *Nature Methods* 14 (4). NIH Public Access: 417–19. doi:10.1038/nmeth.4197.

Phillips, J H, C Chang, J Mattson, J E Gumperz, P Parham, and L L Lanier. 1996. “CD94 and a Novel Associated Protein (94AP) Form a NK Cell Receptor Involved in the Recognition of HLA-A, HLA-B, and HLA-C Allotypes.” *Immunity* 5 (2). Elsevier: 163–72. doi:10.1016/S1074-7613(00)80492-6.

Phillips, J H, J E Gumperz, P Parham, and L L Lanier. 1995. “Superantigen-Dependent, Cell-Mediated Cytotoxicity Inhibited by MHC Class I Receptors on T Lymphocytes.” *Science (New York, N.Y.)* 268 (5209): 403–5. http://www.ncbi.nlm.nih.gov/pubmed/7716542.

Poli, Aurélie, Justyna Kmiecik, Olivia Domingues, François Hentges, Mathieu Bléry, Martha Chekenya, José Boucraut, and Jacques Zimmer. 2013. “NK Cells in Central Nervous System Disorders.” *Journal of Immunology (Baltimore, Md. : 1950)* 190 (11). American Association of Immunologists: 5355–62. doi:10.4049/jimmunol.1203401.

Pollok, K E, Y J Kim, Z Zhou, J Hurtado, K K Kim, R T Pickard, and B S Kwon. 1993. “Inducible T Cell Antigen 4-1BB. Analysis of Expression and Function.” *Journal of Immunology (Baltimore, Md. : 1950)* 150 (3). American Association of Immunologists: 771–81. doi:10.4049/jimmunol.172.8.4779.

Poptsova, Maria S, Irina A Il’icheva, Dmitry Yu Nechipurenko, Larisa A Panchenko, Mingian V Khodikov, Nina Y Oparina, Robert V Polozov, Yury D Nechipurenko, and Sergei L Grokhovsky. 2014. “Non-Random DNA Fragmentation in next-Generation Sequencing.” *Scientific Reports* 4 (March). Nature Publishing Group: 4532. doi:10.1038/srep04532.

Portevin, Damien, Laura E. Via, Seokyong Eum, and Douglas Young. 2012. “Natural Killer Cells Are Recruited during Pulmonary Tuberculosis and Their *Ex Vivo* Responses to Mycobacteria Vary between Healthy Human Donors in Association with KIR Haplotype.” *Cellular Microbiology* 14 (11). Wiley/Blackwell (10.1111): 1734–44. doi:10.1111/j.1462-5822.2012.01834.x.

Porwit, Anja., Jeffrey McCullough, and Wendy N. Erber. 2011. *Blood and Bone Marrow Pathology*. Churchill Livingstone/Elsevier.

Putney, Scott D., Walter C. Herlihy, and Paul Schimmel. 1983. “A New Troponin T and cDNA Clones for 13 Different Muscle Proteins, Found by Shotgun Sequencing.” *Nature* 302 (5910). Nature Publishing Group: 718–21. doi:10.1038/302718a0.

Pydi, Satya Sudheer, Sharada Ramaseri Sunder, Sambasivan Venkatasubramanian, Srinivas Kovvali, Subbanna Jonnalagada, and Vijaya Lakshmi Valluri. 2013. “Killer Cell Immunoglobulin like Receptor Gene Association with Tuberculosis.” *Human Immunology* 74 (1): 85–92. doi:10.1016/j.humimm.2012.10.006.

Pyo, Chul-Woo, Lisbeth A Guethlein, Quyen Vu, Ruihan Wang, Laurent Abi-Rached, Paul J Norman, Steven G E Marsh, Jeffrey S Miller, Peter Parham, and Daniel E Geraghty. 2010. “Different Patterns of Evolution in the Centromeric and Telomeric Regions of Group A and B Haplotypes of the Human Killer Cell Ig-like Receptor Locus.” Edited by Hiroaki Matsunami. *PloS One* 5 (12): e15115. doi:10.1371/journal.pone.0015115.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42. doi:10.1093/bioinformatics/btq033.

Raulet, David H., Werner Held, Isabel Correa, Jeffrey R. Dorfman, Ming-Fan Wu, and Laura Corral. 1997. “Specificity, Tolerance and Developmental Regulation of Natural Killer Cells Defined by Expression of Class I-Specific Ly49 Receptors.” *Immunological Reviews* 155 (1). Wiley/Blackwell (10.1111): 41–52. doi:10.1111/j.1600-065X.1997.tb00938.x.

Raulet, David H., and Russell E. Vance. 2006. “Self-Tolerance of Natural Killer Cells.” *Nature Reviews Immunology* 6 (7): 520–31. doi:10.1038/nri1863.

Rhoads, Anthony, and Kin Fai Au. 2015. “PacBio Sequencing and Its Applications.” *Genomics, Proteomics & Bioinformatics* 13 (5): 278–89. doi:10.1016/j.gpb.2015.08.002.

Robinson, James, Jason A Halliwell, James D Hayhurst, Paul Flicek, Peter Parham, and Steven G E Marsh. 2015. “The IPD and IMGT/HLA Database: Allele Variant Databases.” *Nucleic Acids Research* 43 (Database issue). Oxford University Press: D423–31. doi:10.1093/nar/gku1161.

Robinson, James, Kavita Mistry, Hamish McWilliam, Rodrigo Lopez, and Steven G E Marsh. 2010. “IPD--the Immuno Polymorphism Database.” *Nucleic Acids Research* 38 (Database issue). Oxford University Press: D863–69. doi:10.1093/nar/gkp879.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. “edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26 (1): 139–40. doi:10.1093/bioinformatics/btp616.

Rosenow, Carsten, Rini Mukherjee Saxena, Mark Durst, and Thomas R. Gingeras. 2001. “Prokaryotic RNA Preparation Methods Useful for High Density Array Analysis: Comparison of Two Approaches.” *Nucleic Acids Research* 29 (22). Oxford University Press: e112. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC92579/.

Roy, Sugata, Peter F Barnes, Ankita Garg, Shiping Wu, David Cosman, and Ramakrishna Vankayalapati. 2008. “NK Cells Lyse T Regulatory Cells That Expand in Response to an Intracellular Pathogen.” *Journal of Immunology (Baltimore, Md. : 1950)* 180 (3): 1729–36. http://www.ncbi.nlm.nih.gov/pubmed/18209070.

Rudis, Bob. 2015. “Streamgraph: Streamgraph Is an Htmlwidget for Building Streamgraph Visualizations.” http://github.com/hrbrmstr/streamgraph.

Sambrook, Jennifer G., Harminder Sehra, Penny Coggill, Sean Humphray, Sophie Palmer, Sarah Sims, Haru-Hisa Takamatsu, Thomas Wileman, Alan L. Archibald, and Stephan Beck. 2006. “Identification of a Single Killer Immunoglobulin-like Receptor (KIR) Gene in the Porcine Leukocyte Receptor Complex on Chromosome 6q.” *Immunogenetics* 58 (5-6): 481–86. doi:10.1007/s00251-006-0110-9.

Sanderson, Nicholas D, Paul J Norman, Lisbeth A Guethlein, Shirley A Ellis, Christina Williams, Matthew Breen, Steven D E Park, et al. 2014. “Definition of the Cattle Killer Cell Ig-like Receptor Gene Family: Comparison with Aurochs and Human Counterparts.” *Journal of Immunology (Baltimore, Md. : 1950)* 193 (12). The American Association of Immunologists, Inc.: 6016–30. doi:10.4049/jimmunol.1401980.

Sanderson, Nicholas D, Paul J Norman, Lisbeth A Guethlein, Shirley A Ellis, Christina Williams, Matthew Breen, Steven D E Park, et al. 2014. “Definition of the Cattle Killer Cell Ig-like Receptor Gene Family: Comparison with Aurochs and Human Counterparts.” *Journal of Immunology (Baltimore, Md. : 1950)* 193 (12). The American Association of Immunologists, Inc.: 6016–30. doi:10.4049/jimmunol.1401980.

Santourlidis, Simeon, Hans-Ingo Trompeter, Sandra Weinhold, Britta Eisermann, Klaus L Meyer, Peter Wernet, and Markus Uhrberg. 2002. “Crucial Role of DNA Methylation in Determination of Clonally Distributed Killer Cell Ig-like Receptor Expression Patterns in NK Cells.” *Journal of Immunology (Baltimore, Md. : 1950)* 169 (8): 4253–61. http://www.ncbi.nlm.nih.gov/pubmed/12370356.

Schena, M, D Shalon, R W Davis, and P O Brown. 1995. “Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray.” *Science (New York, N.Y.)* 270 (5235): 467–70. http://www.ncbi.nlm.nih.gov/pubmed/7569999.

Schwartz, John C, Mark S Gibson, Dorothea Heimeier, Sergey Koren, Adam M Phillippy, Derek M Bickhart, Timothy P L Smith, Juan F Medrano, and John A Hammond. 2017. “The Evolution of the Natural Killer Complex; a Comparison between Mammals Using New High-Quality Genome Assemblies and Targeted Annotation.” *Immunogenetics* 69 (4). Springer: 255–69. doi:10.1007/s00251-017-0973-y.

Schwartz, John C, Nicholas D Sanderson, and John A Hammond. n.d. “The Structure, Evolution, and Expression of the Caprine Leukocyte Receptor Complex.” *Manuscript in Preparation*.

Seaman, W E, M Sleisenger, E Eriksson, and G C Koo. 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.” *Journal of Immunology (Baltimore, Md. : 1950)* 138 (12): 4539–44. http://www.ncbi.nlm.nih.gov/pubmed/3584981.

Sipka, Anja, Brianna Pomeroy, Suzanne Klaessig, and Ynte Schukken. 2016. “Bovine Natural Killer Cells Are Present in Escherichia Coli Infected Mammary Gland Tissue and Show Antimicrobial Activity in Vitro.” *Comparative Immunology, Microbiology and Infectious Diseases* 48 (October). Pergamon: 54–60. doi:10.1016/J.CIMID.2016.08.001.

Sipka, Anja, Brianna Pomeroy, Suzanne Klaessig, and Ynte Schukken. 2016. “Bovine Natural Killer Cells Are Present in Escherichia Coli Infected Mammary Gland Tissue and Show Antimicrobial Activity in Vitro.” *Comparative Immunology, Microbiology and Infectious Diseases* 48 (October). Pergamon: 54–60. doi:10.1016/J.CIMID.2016.08.001.

Smith, Hamish R C, Jonathan W Heusel, Indira K Mehta, Sungjin Kim, Brigitte G Dorner, Olga V Naidenko, Koho Iizuka, et al. 2002. “Recognition of a Virus-Encoded Ligand by a Natural Killer Cell Activation Receptor.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (13). National Academy of Sciences: 8826–31. doi:10.1073/pnas.092258599.

Stein-Streilein, J, M Bennett, D Mann, and V Kumar. 1983. “Natural Killer Cells in Mouse Lung: Surface Phenotype, Target Preference, and Response to Local Influenza Virus Infection.” *Journal of Immunology (Baltimore, Md. : 1950)* 131 (6): 2699–2704. http://www.ncbi.nlm.nih.gov/pubmed/6644021.

Storset, Anne K., Siri Kulberg, Ingvild Berg, Preben Boysen, Jayne C. Hope, and Erik Dissen. 2004. “NKp46 Defines a Subset of Bovine Leukocytes with Natural Killer Cell Characteristics.” *European Journal of Immunology* 34 (3): 669–76. doi:10.1002/eji.200324504.

Storset, Anne K., Imer Ö. Slettedal, John L. Williams, Andy Law, and Erik Dissen. 2003. “Natural Killer Cell Receptors in Cattle: A Bovine Killer Cell Immunoglobulin-like Receptor Multigene Family Contains Members with Divergent Signaling Motifs.” *European Journal of Immunology* 33 (4): 980–90. doi:10.1002/eji.200323710.

Strauss-Albee, Dara M, Julia Fukuyama, Emily C Liang, Yi Yao, Justin A Jarrell, Alison L Drake, John Kinuthia, et al. 2015. “Human NK Cell Repertoire Diversity Reflects Immune Experience and Correlates with Viral Susceptibility.” *Science Translational Medicine* 7 (297). NIH Public Access: 297ra115. doi:10.1126/scitranslmed.aac5722.

Stulberg, M J, P W Wright, H Dang, R J Hanson, J S Miller, and S K Anderson. 2007. “Identification of Distal KIR Promoters and Transcripts.” *Genes and Immunity* 8 (2): 124–30. doi:10.1038/sj.gene.6364363.

Sun, Joseph C, and Lewis L Lanier. 2009. “The Natural Selection of Herpesviruses and Virus-Specific NK Cell Receptors.” *Viruses* 1 (3). Multidisciplinary Digital Publishing Institute (MDPI): 362. doi:10.3390/v1030362.

Takahashi, Tomoko, Makoto Yawata, Terje Raudsepp, Teri L. Lear, Bhanu P. Chowdhary, Douglas F. Antczak, and Masanori Kasahara. 2004. “Natural Killer Cell Receptors in the Horse: Evidence for the Existence of Multiple Transcribed *LY49* Genes.” *European Journal of Immunology* 34 (3): 773–84. doi:10.1002/eji.200324695.

Teng, Mingxiang, Michael I. Love, Carrie A. Davis, Sarah Djebali, Alexander Dobin, Brenton R. Graveley, Sheng Li, et al. 2016. “A Benchmark for RNA-Seq Quantification Pipelines.” *Genome Biology* 17 (1). BioMed Central: 74. doi:10.1186/s13059-016-0940-1.

Trapani, Joseph A, and Vivien R Sutton. 2003. “Granzyme B: Pro-Apoptotic, Antiviral and Antitumor Functions.” *Current Opinion in Immunology* 15 (5): 533–43. http://www.ncbi.nlm.nih.gov/pubmed/14499262.

Uhrberg, M, N M Valiante, B P Shum, H G Shilling, K Lienert-Weidenbach, B Corliss, D Tyan, L L Lanier, and P Parham. 1997. “Human Diversity in Killer Cell Inhibitory Receptor Genes.” *Immunity* 7 (6). Elsevier: 753–63. doi:10.1016/S1074-7613(00)80394-5.

Uhrberg, M, N M Valiante, N T Young, L L Lanier, J H Phillips, and P Parham. 2001. “The Repertoire of Killer Cell Ig-like Receptor and CD94:NKG2A Receptors in T Cells: Clones Sharing Identical Alpha Beta TCR Rearrangement Express Highly Diverse Killer Cell Ig-like Receptor Patterns.” *Journal of Immunology (Baltimore, Md. : 1950)* 166 (6): 3923–32. http://www.ncbi.nlm.nih.gov/pubmed/11238637.

Valiante, N M, M Uhrberg, H G Shilling, K Lienert-Weidenbach, K L Arnett, A D'Andrea, J H Phillips, L L Lanier, and P Parham. 1997. "Functionally and Structurally Distinct NK Cell Receptor Repertoires in the Peripheral Blood of Two Human Donors." *Immunity* 7 (6). Elsevier: 739–51. doi:10.1016/S1074-7613(00)80393-3.

van Bergen, Jeroen, and Frits Koning. 2010. "The Tortoise and the Hare: Slowly Evolving T-Cell Responses Take Hastily Evolving KIR." *Immunology* 131 (3). Wiley-Blackwell: 301–9. doi:10.1111/j.1365-2567.2010.03337.x.

van Bergen, Jeroen, Allan Thompson, Arno van der Slik, Tom H M Ottenhoff, Jacobijn Gussekloo, and Frits Koning. 2004. "Phenotypic and Functional Characterization of CD4 T Cells Expressing Killer Ig-like Receptors." *Journal of Immunology (Baltimore, Md. : 1950)* 173 (11). American Association of Immunologists: 6719–26. doi:10.4049/JIMMUNOL.173.11.6719.

van Dijk, Erwin L., Yan Jaszczyszyn, Delphine Naquin, and Claude Thermes. 2018. "The Third Revolution in Sequencing Technology." *Trends in Genetics* 34 (9): 666–81. doi:10.1016/j.tig.2018.05.008.

Vankayalapati, Ramakrishna, Peter Klucar, Benjamin Wizel, Stephen E Weis, Buka Samten, Hassan Safi, Homayoun Shams, and Peter F Barnes. 2004. "NK Cells Regulate CD8+ T Cell Effector Function in Response to an Intracellular Pathogen." *Journal of Immunology (Baltimore, Md. : 1950)* 172 (1): 130–37. http://www.ncbi.nlm.nih.gov/pubmed/14688318.

Vankayalapati, Ramakrishna, Benjamin Wizel, Stephen E Weis, Hassan Safi, David L Lakey, Ofer Mandelboim, Buka Samten, Angel Porgador, and Peter F Barnes. 2002. "The NKp46 Receptor Contributes to NK Cell Lysis of Mononuclear Phagocytes Infected with an Intracellular Bacterium." *Journal of Immunology (Baltimore, Md. : 1950)* 168 (7): 3451–57. http://www.ncbi.nlm.nih.gov/pubmed/11907104.

Velculescu, V E, L Zhang, B Vogelstein, and K W Kinzler. 1995. "Serial Analysis of Gene Expression." *Science (New York, N.Y.)* 270 (5235): 484–87. http://www.ncbi.nlm.nih.gov/pubmed/7570003.

Vivier, Eric, Elena Tomasello, Myriam Baratin, Thierry Walzer, and Sophie Ugolini. 2008. "Functions of Natural Killer Cells." *Nature Immunology* 9 (5). Nature Publishing Group: 503–10. doi:10.1038/ni1582.

Wallin, Robert P. A., Valentina Screpanti, Jakob Michaëlsson, Alf Grandien, and Hans-Gustaf Ljunggren. 2003. "Regulation of Perforin-Independent NK Cell-Mediated Cytotoxicity." *European Journal of Immunology* 33 (10): 2727–35. doi:10.1002/eji.200324070.

Walzer, T., M. Blery, J. Chaix, N. Fuseri, L. Chasson, S. H. Robbins, S. Jaeger, et al. 2007. “Identification, Activation, and Selective in Vivo Ablation of Mouse NK Cells via NKp46.” *Proceedings of the National Academy of Sciences* 104 (9): 3384–89. doi:10.1073/pnas.0609692104.

Walzer, Thierry, and Eric Vivier. 2011. “G-Protein-Coupled Receptors in Control of Natural Killer Cell Migration.” *Trends in Immunology* 32 (10): 486–92. doi:10.1016/j.it.2011.05.002.

Wang, Jian, Fengqi Li, Meijuan Zheng, Rui Sun, Haiming Wei, and Zhigang Tian. 2012. “Lung Natural Killer Cells in Mice: Phenotype and Response to Respiratory Infection.” *Immunology* 137 (1): 37–47. doi:10.1111/j.1365-2567.2012.03607.x.

Warfield, Kelly L., Jeremy G. Perkins, Dana L. Swenson, Emily M. Deal, Catharine M. Bosio, M. Javad Aman, Wayne M. Yokoyama, Howard A. Young, and Sina Bavari. 2004. “Role of Natural Killer Cells in Innate Protection against Lethal Ebola Virus Infection.” *The Journal of Experimental Medicine* 200 (2): 169–79. doi:10.1084/jem.20032141.

Wauquier, Nadia, Cindy Padilla, Pierre Becquart, Eric Leroy, and Vincent Vieillard. 2010. “Association of KIR2DS1 and KIR2DS3 with Fatal Outcome in Ebola Virus Infection.” *Immunogenetics* 62 (11-12): 767–71. doi:10.1007/s00251-010-0480-x.

Wehner, Rebekka, Bärbel Löbel, Martin Bornhäuser, Knut Schäkel, Marc Cartellieri, Michael Bachmann, Ernst Peter Rieber, and Marc Schmitz. 2009. “Reciprocal Activating Interaction between 6-Sulfo LacNAc $^{+}$ Dendritic Cells and NK Cells.” *International Journal of Cancer* 124 (2): 358–66. doi:10.1002/ijc.23962.

Weirather, Jason L, Mariateresa de Cesare, Yunhao Wang, Paolo Piazza, Vittorio Sebastiano, Xiu-Jie Wang, David Buck, and Kin Fai Au. 2017. “Comprehensive Comparison of Pacific Biosciences and Oxford Nanopore Technologies and Their Applications to Transcriptome Analysis.” *F1000Research* 6 (February): 100. doi:10.12688/f1000research.10571.1.

Welte, Stefan A., Christian Sinzger, Stefan Z. Lutz, Harpreet Singh-Jasuja, Kerstin Laib Sampaio, Ute Eknigk, Hans-Georg Rammensee, and Alexander Steinle. 2003. “Selective Intracellular Retention of Virally Induced NKG2D Ligands by the Human Cytomegalovirus UL16 Glycoprotein.” *European Journal of Immunology* 33 (1). Wiley-Blackwell: 194–203. doi:10.1002/immu.200390022.

Westgaard, Ingunn H., Siri F. Berg, John T. Vaage, Lawrence L. Wang, Wayne M. Yokoyama, Erik Dissen, and Sigbjørn Fossum. 2004. “Rat NKp46 Activates Natural Killer Cell Cytotoxicity and Is Associated with FcεRIγ and CD3ζ.” *Journal of Leukocyte Biology* 76 (6): 1200–1206. doi:10.1189/jlb.0903428.

Wickham, Hadley. 2009. *Elegant Graphics for Data Analysis*. *Media*. Vol. 35. doi:10.1007/978-0-387-98141-3.

Wilhelm, B T, K L McQueen, J D Freeman, F Takei, and D L Mager. 2001. “Comparative Analysis of the Promoter Regions and Transcriptional Start Sites of Mouse Ly49 Genes.” *Immunogenetics* 53 (3): 215–24. http://www.ncbi.nlm.nih.gov/pubmed/11398966.

Wright, P W, H Li, A Huehn, G M O’Connor, S Cooley, J S Miller, and S K Anderson. 2014. “Characterization of a Weakly Expressed KIR2DL1 Variant Reveals a Novel Upstream Promoter That Controls KIR Expression.” *Genes & Immunity* 15 (7): 440–48. doi:10.1038/gene.2014.34.

Yawata, Makoto, Nobuyo Yawata, Monia Draghi, Ann-Margaret Little, Fotini Partheniou, and Peter Parham. 2006. “Roles for HLA and KIR Polymorphisms in Natural Killer Cell Repertoire Selection and Modulation of Effector Function.” *The Journal of Experimental Medicine* 203 (3). The Rockefeller University Press: 633–45. doi:10.1084/jem.20051884.

Yokoyama, Wayne M., and Sungjin Kim. 2006. “Licensing of Natural Killer Cells by Self-Major Histocompatibility Complex Class I.” *Immunological Reviews* 214 (1): 143–54. doi:10.1111/j.1600-065X.2006.00458.x.

Zhang, Yan, Diana L Wallace, Catherine M de Lara, Hala Ghattas, Becca Asquith, Andrew Worth, George E Griffin, et al. 2007. “In Vivo Kinetics of Human Natural Killer Cells: The Effects of Ageing and Acute and Chronic Viral Infection.” *Immunology* 121 (2). Wiley-Blackwell: 258–65. doi:10.1111/j.1365-2567.2007.02573.x.

Zytnicki, Matthias. 2017. “Mmquant: How to Count Multi-Mapping Reads?” *BMC Bioinformatics* 18 (1). BioMed Central: 411. doi:10.1186/s12859-017-1816-4.