The Structure, Evolution, and Gene Expression Within the Caprine Leukocyte Receptor Complex

John C. Schwartz[1], Nicholas D. Sanderson[1,2], Derek M. Bickhart[3], Timothy P. L. Smith[4] and John A. Hammond[1]*

[1] The Pirbright Institute, Woking, United Kingdom, [2] Experimental Medicine Division, Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford, United Kingdom, [3] Cell Wall Biology and Utilization Research, USDA-ARS, Madison, WI, United States, [4] Meat Animal Research Center, USDA-ARS, Clay Center, NE, United States

The leukocyte receptor complex (LRC) encodes a large number of immunoglobulin (Ig)-like receptors involved in the immune response, particularly in modulating natural killer (NK) cell function. The killer cell Ig-like receptors (*KIR*), the leukocyte Ig-like receptors (*LILR*), and a recently described novel Ig-like receptor family are highly variable between species, which is consistent with rapid evolution driven by selection pressure from pathogens. Among the species studied to date, only simians (such as humans) and bovids (such as cattle and goats) have an expanded complement of *KIR* genes and represent an interesting model to study *KIR* evolution. Using recently improved genome assemblies and an assembly of bacterial artificial chromosomes, we describe the structure of the LRC, and the *KIR* region in particular, in goats and compare this to sheep as the assemblies allow. These species diverged from a common ancestor ~10 million years ago and from cattle ~25 million years ago. We identified conserved *KIR* genes common to both goats and sheep and confirm a partial sheep haplotype shared between the Rambouillet and Texel breeds. Goats and sheep have independently expanded two novel *KIR* subgroups, and unlike cattle or any other mammal, they do not appear to possess a functional 3DL-lineage *KIR* gene. Investigation of LRC gene expression using available transcriptomic data for various sheep and goat tissues largely confirmed putative gene annotation and revealed that a relatively conserved *caprinae*-specific *KIR* subgroup is expressed in macrophages. The *LILR* and novel Ig-like receptors were also highly expressed across a diverse range of tissues. This further step toward our understanding of the LRC receptor repertoire will help inform future studies investigating immune response variation in these species.

Keywords: NK cells, LRC, *KIR, LILR*, immunoglobulin-like receptors, goat, sheep

INTRODUCTION

Vertebrates have evolved numerous receptor/ligand systems for monitoring and maintaining organism health. Most notable for the high degree of polymorphism and rapid evolution between species are the major histocompatibility complex (MHC) class I and class I-like molecules and their respective somatically rearranging and germline encoded receptors. As the most rapidly evolving regions of jawed vertebrate genomes, both MHC and their germline receptor complexes

are typically highly polymorphic and variable in gene content. As a result, these gene complexes are often substantially different even between closely related species. One such gene complex is the leukocyte receptor complex (LRC). Within the LRC exist numerous related genes belonging to the immunoglobulin (Ig) superfamily. These include the killer cell Ig-like receptors (*KIR*) expressed by natural killer (NK) cells and a subset of CD8 T cells, and the leukocyte Ig-like receptors (*LILR*) expressed by a larger variety of lymphoid and myeloid cell types. We also recently reported the existence of an additional novel Ig-like gene family in the LRC of many mammalian species, including seven genes in cattle and a pseudogene outside the LRC in humans. Like the *KIR* and *LILR*, these genes also vary in gene content between haplotypes and between species (1). Although little is known about these novel genes, they typically encode a single long-tailed inhibitory receptor and a variable number of short-tailed receptors which potentially antagonize the function of the inhibitory form. Thus, the LRC is a dynamic, quickly evolving region of the genome that encodes multiple related receptors.

The *KIR* genes encoded within the LRC, and the unrelated killer lectin-like receptors (*KLR*) encoded within the natural killer complex (NKC) are the major cell-surface receptors which mediate NK cell function via interaction with MHC class I ligands. In all species studied to date, these gene complexes are located on different chromosomes and have variably expanded or contracted *KIR* and *KLR* content. Mice (*Mus musculus*) have a highly expanded and variable repertoire of *KLRA* genes (2, 3), yet possess only two *KIR*-like genes located outside the LRC on chromosome X with likely alternative functions than NK cell control (4, 5). In contrast, humans have a highly expanded, gene variable and polymorphic *KIR* region, yet only a single non-functional *KLRA* gene (6). An alternative strategy has been adopted by marine carnivores that have maintained single functional *KIR* and *KLRA* genes with no evidence of gene expansion (7).

Species belonging to the *Bovidae* family, such as cattle (*Bos taurus*), goats (*Capra hircus*), and sheep (*Ovis aries*), are so far unique in that they have a highly expanded repertoire of *KLR* genes within the NKC (8, 9), and are also the only non-simian species known to have an expanded complement of *KIR* genes in the LRC (10–12). Importantly, these receptors appear to be differentially expressed in NK cells depending on the MHC genotype of the animal, implying a similar regulation of NK cell function (13). A cattle *KIR* haplotype has been characterized and localized to a telomeric end of chromosome 18, and contains eight functional and 10 non-functional *KIR* genes (11). Except for *KIR2DL1* and closely related pseudogenes, the expanded and likely functional *KIR* genes in cattle have evolved from the 3DX-lineage and are therefore more closely related to *KIR3DX1* in humans. This is in contrast to humans where the functional *KIR* genes belong to the 3DL-lineage (10, 11). The goat *KIR* region, and the LRC in general, appears to be no less complex than cattle, and is disrupted by many sequence gaps and missing sequence in the previous CHIR_2.0 assembly compared to the latest ARS1 genome assembly, in which the LRC is intact on a single contig (14).

In addition to an expanded repertoire of KIR, humans also possess 11 functional LILR, which are expressed by a diverse range of cell types. At least five of these interact with MHC class I ligands, while others appear to interact with non-MHC molecules such as tetherin (15–17). Mice possess 11 paired immunoglobulin-like receptor (*PIR*) genes which are orthologous to and arranged similarly to human *LILR* (18–20). Pigs possess at least 17 *LILR* genes and gene fragments (at least six are functional) that form two distinct phylogenetic clades (1), however their ligands remain unknown. Multiple *LILR* orthologs are also known to exist in cattle (21) although they remain largely uncharacterized. Thus, like the *KIR*, the *LILR* are also highly variable between species, however little is known about their function and polymorphism outside humans.

Goats and sheep diverged from a common ancestor ∼10 Mya and from cattle ∼25 Mya (22). There has been no detailed investigation of the *KIR* or other LRC genes in a non-cattle species of bovid to date and there are no *KIR* mRNA accessions for goats or sheep in GenBank. To better understand the evolution of this region we describe the LRC that assembled as one contiguous sequence in the recent high-quality goat genome assembly, ARS1 (14). We also report the first sequenced and assembled partial sheep *KIR* haplotype, derived from bacterial artificial chromosomes (BACs) and compared to the LRC in both the Oar_v4 reference assembly and the recently available long-read sheep assembly, Oar_rambouillet_v1.0. The expression pattern and likely functionality of these genes was then examined using publicly available transcriptomic data from the sheep gene expression atlas (23) and the goat gene expression atlas (24, 25) produced by the FAANG consortium (26).

METHODS

Genome Assemblies and BAC Clones

A comparison of the assembly across the LRC between the previous the goat reference assembly, CHIR_2.0, and the current assembly, ARS1, is discussed elsewhere (14). The goat chromosome 18 scaffold of ARS1 is available in GenBank (CM004579). The sheep chromosome 14 scaffolds within the Oar_v4 and the Oar_rambouillet_v1.0 assemblies were acquired from GenBank (NC_019471 and CM008485, respectively). Although we manually annotated the LRC within Oar_v4, the large number of sequence gaps (*n* = 43) across this region and likely misassembled contigs was determined to be less informative than the Oar_rambouillet_v1.0 assembly which we include in this manuscript.
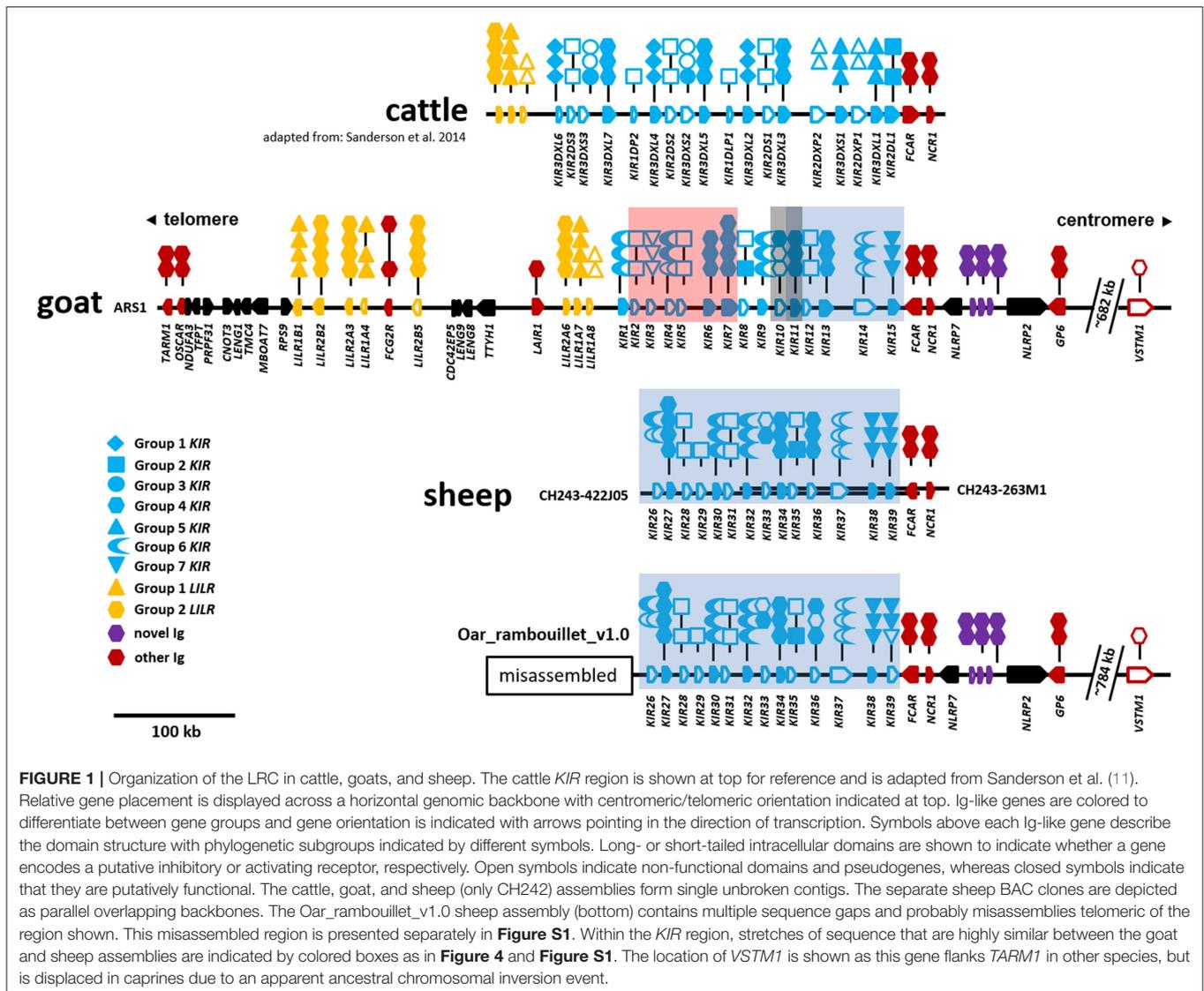
The Children's Hospital Oakland Research Institute (CHORI)-243 BAC library (https://bacpacresources.org/) is derived from a Texel ram, and used in part to generate the sheep genome reference assembly, Oar_v4 (27). BAC end-sequences for this library were screened *in silico* for sequence similarity to the *KIR* region. Two BAC clones were determined to overlap the sheep *KIR* region: CH243-263M01 and CH243-422J05. LRC sequence data derived from these clones, as well as other closely related assemblies were then used to identify more BAC clones

over the sheep *KIR* region, but none were identified. These clones were expanded overnight and DNA was purified using the Qiagen Large Construct Kit and manufacturer's protocol. The DNA was then sequenced using the Pacific Biosciences RSII platform at GATC Biotech (Konstanz, Germany). Read filtering and assembly were conducted using the Pacific Biosciences single molecule real-time (SMRT) analysis software (version 2.3.0; http://www.pacb.com/devnet/). The cloning vector sequence (GenBank: AY487253) was identified and removed, resulting in consensus sequences of ~223 and 172 kb for CH243-422J05 and CH243-263M01, respectively. Alignment of these insert sequences generated an overlap of ~147 kb which contained no differences between the two clones, so they almost certainly belong to the same haplotype. Together, the two assembled clones form a contiguous sequence of 248,501 bp. This assembly was annotated and deposited in GenBank (MH680527).

Identification and Annotation of LRC Genes

Genomic scaffolds and BAC clones were queried for their gene content using a combination of the basic local alignment search tool (BLAST) (28), the BLAST-like alignment tool (BLAT) (29), HMMgene (30), and NCBI conserved domain search (31). Transmembrane (TM) domains were predicted using TMHMM (32). Identified genes and gene fragments were then manually annotated using Artemis 17.0.1 (33). Exonic structure for the *KIR* genes was determined based on alignment to the previously described cattle *KIR* gene sequences (11). These were visually verified to confirm accuracy, and splice junction donor acceptor sites were identified based on the GT/AG motif (34). Translation of the putative coding region sequence (CDS) enabled prediction of functionally intact genes by identifying frameshifts and nonsense codons.

Genes encoding inhibitory receptors were identified in part based on their potential to produce a protein containing

FIGURE 1 | Organization of the LRC in cattle, goats, and sheep. The cattle *KIR* region is shown at top for reference and is adapted from Sanderson et al. (11). Relative gene placement is displayed across a horizontal genomic backbone with centromeric/telomeric orientation indicated at top. Ig-like genes are colored to differentiate between gene groups and gene orientation is indicated with arrows pointing in the direction of transcription. Symbols above each Ig-like gene describe the domain structure with phylogenetic subgroups indicated by different symbols. Long- or short-tailed intracellular domains are shown to indicate whether a gene encodes a putative inhibitory or activating receptor, respectively. Open symbols indicate non-functional domains and pseudogenes, whereas closed symbols indicate that they are putatively functional. The cattle, goat, and sheep (only CH242) assemblies form single unbroken contigs. The separate sheep BAC clones are depicted as parallel overlapping backbones. The Oar_rambouillet_v1.0 sheep assembly (bottom) contains multiple sequence gaps and probably misassembles telomeric of the region shown. This misassembled region is presented separately in **Figure S1**. Within the *KIR* region, stretches of sequence that are highly similar between the goat and sheep assemblies are indicated by colored boxes as in **Figure 4** and **Figure S1**. The location of *VSTM1* is shown as this gene flanks *TARM1* in other species, but is displaced in caprines due to an apparent ancestral chromosomal inversion event.

immunoreceptor tyrosine-based inhibitory motifs (ITIMs), which have a loosely conserved amino acid sequence of (hydrophobic)-x-Y-x-x-(hydrophobic) in their intracellular tail, where the hydrophobic residues are typically either valine, leucine, or isoleucine, although serine may occasionally be found at the N-terminal position (35). In contrast, genes for activating receptors were identified based on their possession of a short intracellular tail and a charged residue, such as arginine or lysine, in the predicted TM domain that would enable interaction with an activating adapter molecule such as CD3ζ, FcRγ chain, DAP10, or DAP12. Thus, for the purposes of this manuscript, "long-tailed" and "short-tailed" receptors are synonymous with inhibitory and activating receptors, respectively.

Phylogenetic Analyses and Subgroup Definitions for LILR and KIR

We previously defined two distinct phylogenetic clades (groups 1 and 2) of *LILR* in the pig (1) and we preserve this nomenclature in the present manuscript. Cattle *KIR* gene groups 1 through five were previously defined (11), and these definitions are also used in the present manuscript. Phylogenetic analysis of goat, sheep, and cattle *KIR* genes was used to place individual goat and sheep *KIR* genes into either one of these five groups, or into the novel presently defined groups 6 and 7. *KIR* gene phylogenetic analysis was performed using the exons encoding the leader and the D0 and D1 Ig domains. The D2 Ig domain, TM, and intracellular regions are less variable and more similar to each

other even between distantly related *KIR* and do not provide enough resolution for phylogenetic analyses (10). Sequences were aligned using MUSCLE (36). Phylogenetic trees were constructed using the neighbor-joining method, the Tamura-3 parameter model (37), pairwise deletion, and 1,000 bootstrap iterations within MEGA7 (38).

Transcriptomic Analyses of Goat and Sheep LRC Ig-Like Genes

Recently generated sheep (23) and goat (24, 25) expression atlas data (BioProject accessions: PRJEB19199 and PRJEB23196, respectively) and T cell expression data (BioProject: PRJEB27455) were used to investigate the expression of the LRC genes identified in this study. Data for alveolar macrophages, bone marrow macrophages, CD4+ cells, CD8+ cells, kidney, liver, spleen, and thymus, were downloaded from both datasets. There are currently no available NK cell (i.e., NCR1+) transcriptomes available for the goat or sheep, so this population was not specifically assessed. Reads corresponding to putative LRC Ig-like genes were identified using BLAST+ (39) and an E-value cut-off of 1×10^{-10}. We assessed the relative expression profiles of identified transcripts by combining detected read counts from each tissue per species. Raw reads per transcript were then normalized using the RUVg method of the RUVSeq R package (40) using the transcriptional read counts of the *SDHA*, *PPIA*, *GAPDH*, *ACTB*, *RPL13A*, and *YWHAZ* genes in each tissue as controls and by setting the value of "k" to 1.
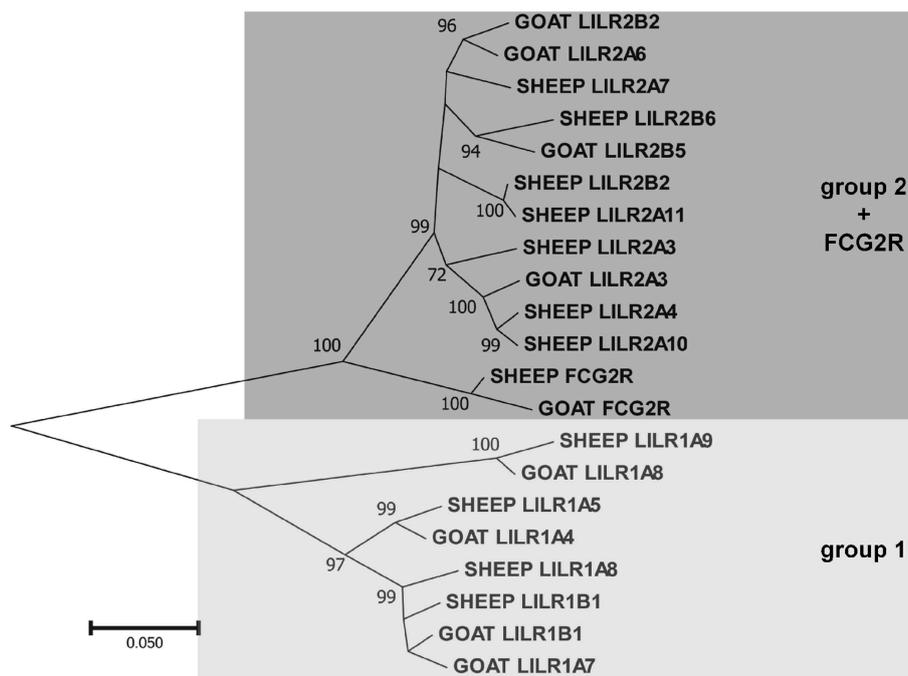

FIGURE 2 | Phylogenetic relationships of goat and sheep *LILR* nucleotide coding region sequences for the extracellular Ig-like domains. *FCG2R* is shown as it is monophyletic with the *LILR*. These two *LILR* groups shown here are based on those previously described in the pig (1). Branch node values indicate the percentage of trees (out of 1,000 iterations) in which the associated sequences cluster together; only values ≥50 are shown. Branch length scale units are number of substitutions per site.
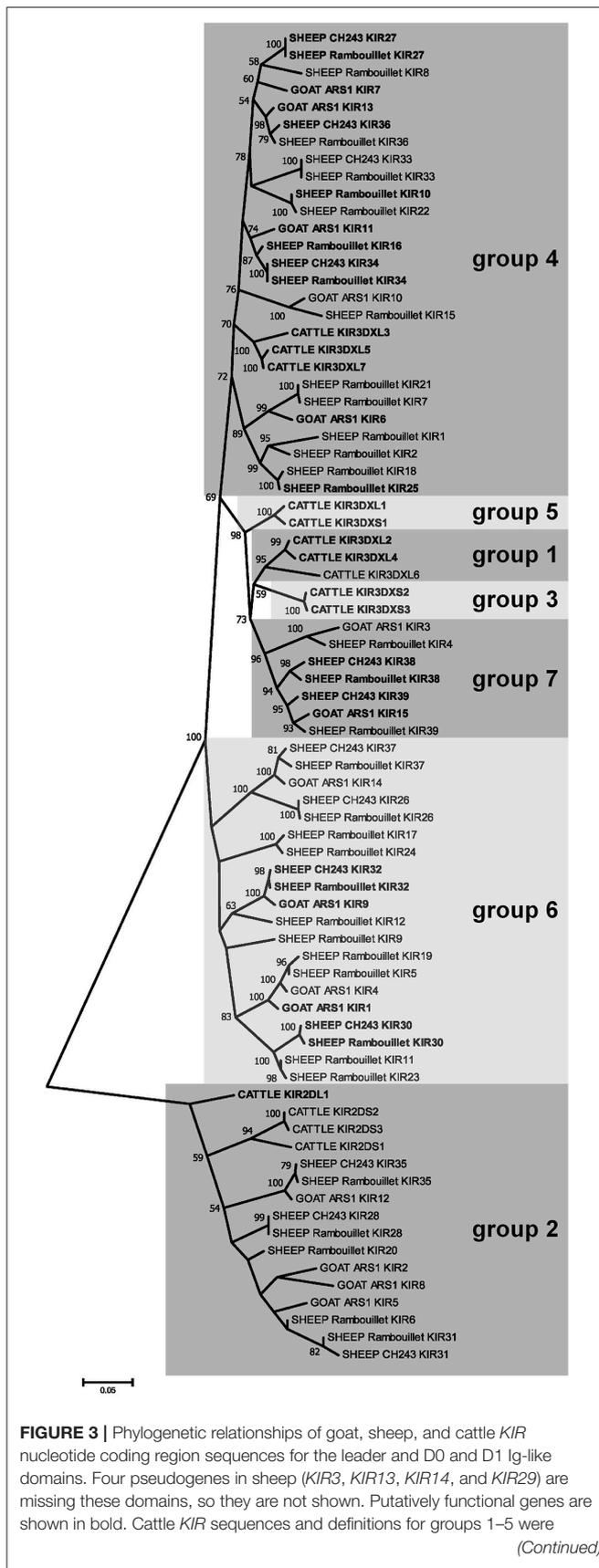
Normalized read count heat maps were generated using the R ggplot2 package.

RESULTS

Gene Content and Structure of the Goat LRC

In the ARS1 assembly, the goat LRC is on the long arm of the telocentric chromosome 18 and ~2.4 Mb from the telomere. The region is flanked by *TARM1* on the telomeric end and *GP6* on the centromeric end, is ~746 kb in size, and contains no sequence gaps (**Figure 1**). No additional Ig-like genes were identified within 500 kb beyond these genes, suggesting the region is complete. However, the Ig-like gene *VSTM1*, which is adjacent to *TARM1* in pigs and humans, was identified on the opposite end of the LRC and ~682 kb from *GP6*. With the exception of *VSTM1*, the LRC is inverted relative to the centromere compared to other species. These observations are indicatative of a chromosomal inversion with a breakpoint <26 kb downstream (i.e., telomeric in goats) from *TARM1*. The Ig domain of *VSTM1* is also truncated in the goat, possibly rendering that gene non-functional. The gene encoding the Fc receptor for IgA, *FCAR*, which flanks the centromeric end of the *KIR* region, is also inverted in the goat relative to other species, such as cattle, and in the opposite orientation as the *KIR* and *NCR1*. There are a total of 15 *KIR* genes, 8 *LILR* genes, and 3 novel two-domain Ig-like genes in the goat LRC (**Figure 1**), and of these, seven *KIR*, five *LILR*, and all three novel genes are putatively functional. Detailed information regarding gene structure and genome coordinates for the *LILR*, *KIR*, and novel Ig-like genes are presented in **Table S1**.

Goat LILR Subgroups Are Conserved

As in other species, the goat *LILR* are present in two distinct clusters separated by the *CDC42EP5-LENG9-LENG8-TTYH1-LAIR1* gene cluster (**Figure 1**). There are two distinct phylogenetic groups of *LILR* in goats and sheep (**Figure 2**), with two functional group 1 genes (one inhibitory and one activating) and three functional group 2 genes (one inhibitory and two activating). The unique bovid Fc receptor *FCG2R* is encoded in the midst of the *KIR*-distal *LILR* cluster and is phylogenetically most similar to the group 2 *LILR* genes (**Figure 2**); however, unlike the usual four-domain structure of the *LILR*, *FCG2R* only encodes an Ig1 domain and an Ig4 domain (**Figure 1**).

Novel Goat KIR Genes and Subgroups

Two genes with unusual domain organization were identified. An apparently functional four-domain inhibitory *KIR* gene (*CahiKIR7*) with the domain structure D0-D1-D1-D2, and an

FIGURE 3 | Phylogenetic relationships of goat, sheep, and cattle *KIR* nucleotide coding region sequences for the leader and D0 and D1 Ig-like domains. Four pseudogenes in sheep (*KIR3*, *KIR13*, *KIR14*, and *KIR29*) are missing these domains, so they are not shown. Putatively functional genes are shown in bold. Cattle *KIR* sequences and definitions for groups 1–5 were

(Continued)

inhibitory *KIR* pseudogene (*CahiKIR14*) which contains a ∼7 kb retrovirus insertion in the D2 domain. The goat appears to lack a functional group 2 gene, such as *KIR2DL1*, which is present in many other species. Furthermore, the goat lacks genes belonging to phylogenetic groups 1, 3, and 5, which together comprise the majority of the functional *KIR* genes in cattle. Of the previously described clades in cattle, goats only possess functional copies from group 4. In contrast, we identified two novel phylogenetic clades of *KIR* in the goat, which we presently define as groups 6 and 7 (**Figure 3**). Four *KIR* genes from these novel subgroups appear functional and three are either null alleles on this haplotype or pseudogenes.

The Structure and Gene Organization of the Sheep and Goat LRC Appears Largely Conserved

Although the sheep genome assemblies are not as complete as the goat, it is possible to observe some common characteristics with the current level of assembly. The sheep LRC is also inverted on the long arm of the telocentric chromosome 14, ∼2.9 Mb from the telomere. Despite a recent update and addition of long-read PacBio sequence, the region between *GP6* and *TARM1* within the Oar_v4 (Chr14: 59,516,600–60,255,000) reference assembly is disrupted by 43 sequence gaps and likely incorrectly placed contigs and missing sequence (GenBank: NC_019471). We therefore focused our analyses on the Oar_ramboullet_v1.0 assembly which contains only five sequence gaps across the same region (**Figure S1**). Within this assembly, the *LILR* are contained on a single contig and are similarly organized as in the goat, although sheep are missing an inhibitory group 2 pseudogene and have an additional activating group 2 gene. Peculiarly, there are two *LILR* genes telomeric of *TARM1* and several gene fragments on an adjacent contig (**Figure S1**). As far as we are aware, this unusual *LILR* placement is not described in other species, nor is it found in Oar_v4, suggesting that this arrangement is most likely the result of misassembly. Overall, those regions that are well-assembled confirm the goat assembly and provide some evidence of species-specific diversification.

The Sheep Genome Assembly Contains Two Distinct KIR Haplotypes

Manual annotation of the Oar_ramboullet_v1.0 assembly revealed that all five sequence gaps are within the *KIR* region and these gaps separate contigs which contain either *KIR* genes or (presumably) incorrectly placed novel Ig-like genes. There are 39 *KIR* genes and gene fragments within this assembly. These genes are mostly assembled onto two large contigs at the telomeric (14 genes) and centromeric (18 genes) ends of the *KIR* region and seven additional *KIR* genes are assembled to three much smaller contigs (**Figure 4**, **Figure S1**). A four-domain *KIR* gene is found on both the telomeric and centromeric contigs, although it is uncertain if this is due to the incorrect assembly of different haplotypes or if it is a real duplication. Phylogenetic analysis (**Figure 3**) identified at least six pairs of genes that appear to be duplicated across contigs (*ovarKIR17/ovarKIR24*, *ovarKIR18/ovarKIR25*, *ovarKIR19/ovarKIR5*, *ovarKIR21/*
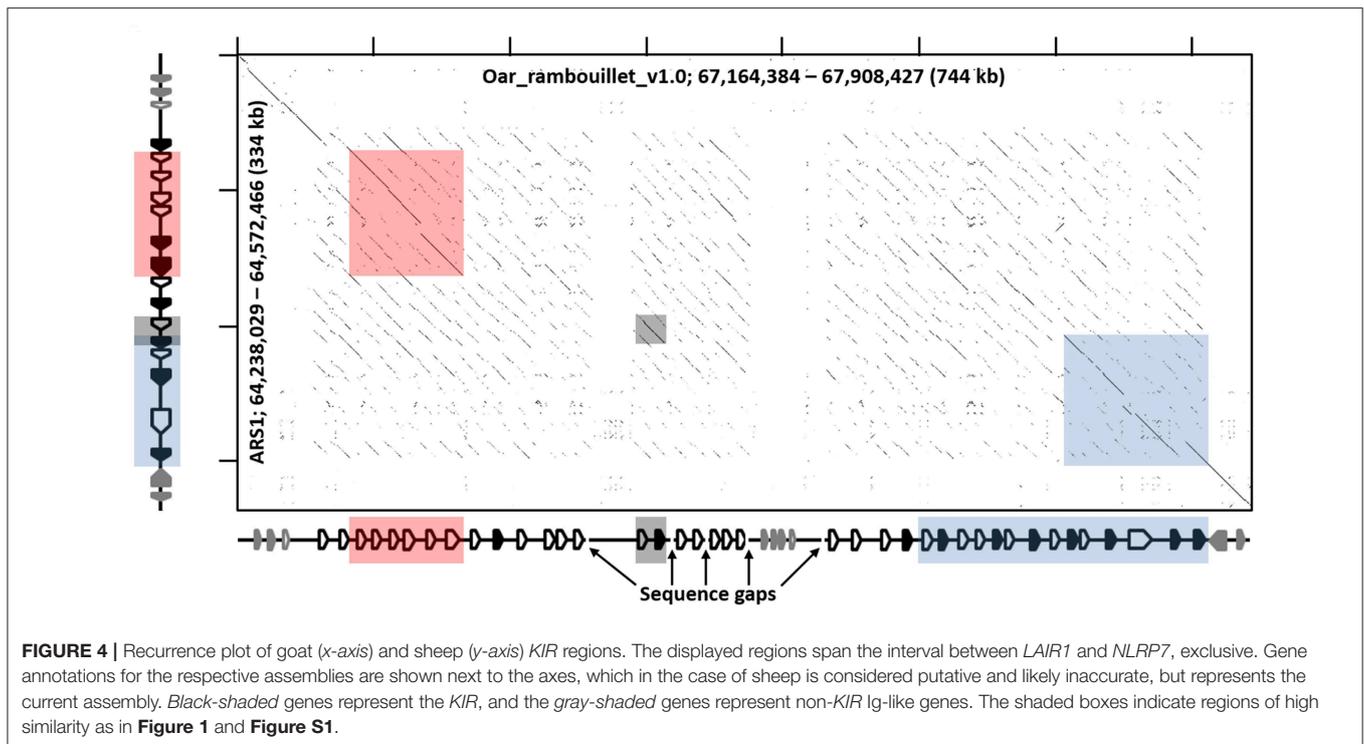
ovarKIR7, *ovarKIR22/ovarKIR10*, and *ovarKIR23/ovarKIR11*). Overall these observations strongly suggest that two distinct and incorrectly phased haplotypes have been at least partially assembled in Oar_ramboullet_v1.0, and that the sequence gaps within the *KIR* region are the result of the impossible assembly of two distinct haplotypes into one reference sequence.

BAC Assembly of the Sheep KIR Region

To further investigate the sheep *KIR* region, we identified and sequenced two overlapping BAC clones derived from a Texel ram. The 248.5 kb BAC assembly contains 14 *KIR* genes which, although derived from a different breed, match the centromeric *KIR* found in the Oar_ramboullet_v1.0 assembly, thus confirming that haplotype structure (**Figure 1**). Seven of these *KIR* genes are putatively functional and only one of these is activating (*OvarKIR38*). Although the overall structure is the same, the genes are generally polymorphic between the BAC assembly and the Oar_ramboullet_v1.0 assembly. Of these polymorphisms, two putative indels that could affect gene function are present in *OvarKIR36* and *OvarKIR39* in the Oar_ramboullet_v1.0 but not the BAC assembly. These are likely the result of sequencing errors, of which indels are common in long-read sequencing data.

Transcriptome Analysis Reveals Unusual Expression of Group 7 KIR

To assess the transcription of goat and sheep LRC genes, we queried the sheep and goat gene expression atlases (BioProjects: PRJEB19199 and PRJEB23196, respectively) and goat CD4 and CD8 T cell expression data (BioProject: PRJEB27455). We then normalized the expression based on the read counts for endogenous reference genes (**Figure 5**, **Table S2**). Given that these data are based on short reads, which can make distinguishing between highly similar genes problematic, we focused our analyses on the gene subgroups for *KIR* and *LILR*, and on the inhibitory or activating designations for the novel two-domain genes. This revealed that the tissue distribution of LRC gene expression was broadly similar between goats and sheep and largely consistent with leukocyte-associated receptors (**Figure 5**, **Table S2**). The *LILR* and the Fc receptor-encoding gene, *FCAR*, for example, were consistently and expectedly expressed by macrophage populations, as were the novel Ig-like genes. Unexpectedly however, the novel *FCAR*-proximal group 7 *KIR* genes were highly expressed in the macrophage populations, whereas the other *KIR* genes were not. While the expression of *NCR1* in all sheep tissues was minimal, goat spleen and CD8 samples were relatively high in *NCR1* expression. However, *NCR1* was not expressed in the macrophage pools, suggesting minimal NK cell contamination in the macrophage preparations (**Figure 5**). In contrast the *KIR* subgroups were approximately equally expressed in the spleen, while subgroup 4 *KIR* were more highly expressed in the CD8 samples. Additionally, novel Ig-like gene expression, while highest in alveolar macrophages, was higher in CD8 cells compared to CD4 cells. As far as our analyses could determine, the transcriptomic data further confirmed the

FIGURE 4 | Recurrence plot of goat (*x-axis*) and sheep (*y-axis*) *KIR* regions. The displayed regions span the interval between *LAIR1* and *NLRP7*, exclusive. Gene annotations for the respective assemblies are shown next to the axes, which in the case of sheep is considered putative and likely inaccurate, but represents the current assembly. *Black-shaded* genes represent the *KIR*, and the *gray-shaded* genes represent non-*KIR* Ig-like genes. The shaded boxes indicate regions of high similarity as in **Figure 1** and **Figure S1**.

annotations and gene models that our *in silico* analyses had predicted.

DISCUSSION

The gross organization of the goat and sheep LRC is typical of other described species, albeit with variation in the number of genes within particular gene groups, such as the *KIR*, *LILR*, and novel Ig-like receptors. The *KIR* were found to be as expanded as in cattle (11), but with major differences in subgroup phylogeny, indicating that cattle *KIR* and caprine *KIR* have distinct evolutionary histories with independent expansions and/or contractions. Particular *KIR* subgroups consequently have very little shared homology. Of note, goats and sheep lack a functional group 2 (i.e., 3DL-lineage) gene, such as *BotaKIR2DL1*. This gene lineage is found in a wide variety of other species, such as cattle, pigs, horses, cats, seals, and sea lions (7), and is most closely related to the expanded *KIR* genes found in primates (10). The existence of a four-domain *KIR* gene in goats and sheep is also unusual, although four-domain *KIR* genes have also been described in the New World night monkey *Aotus vociferans* (41).
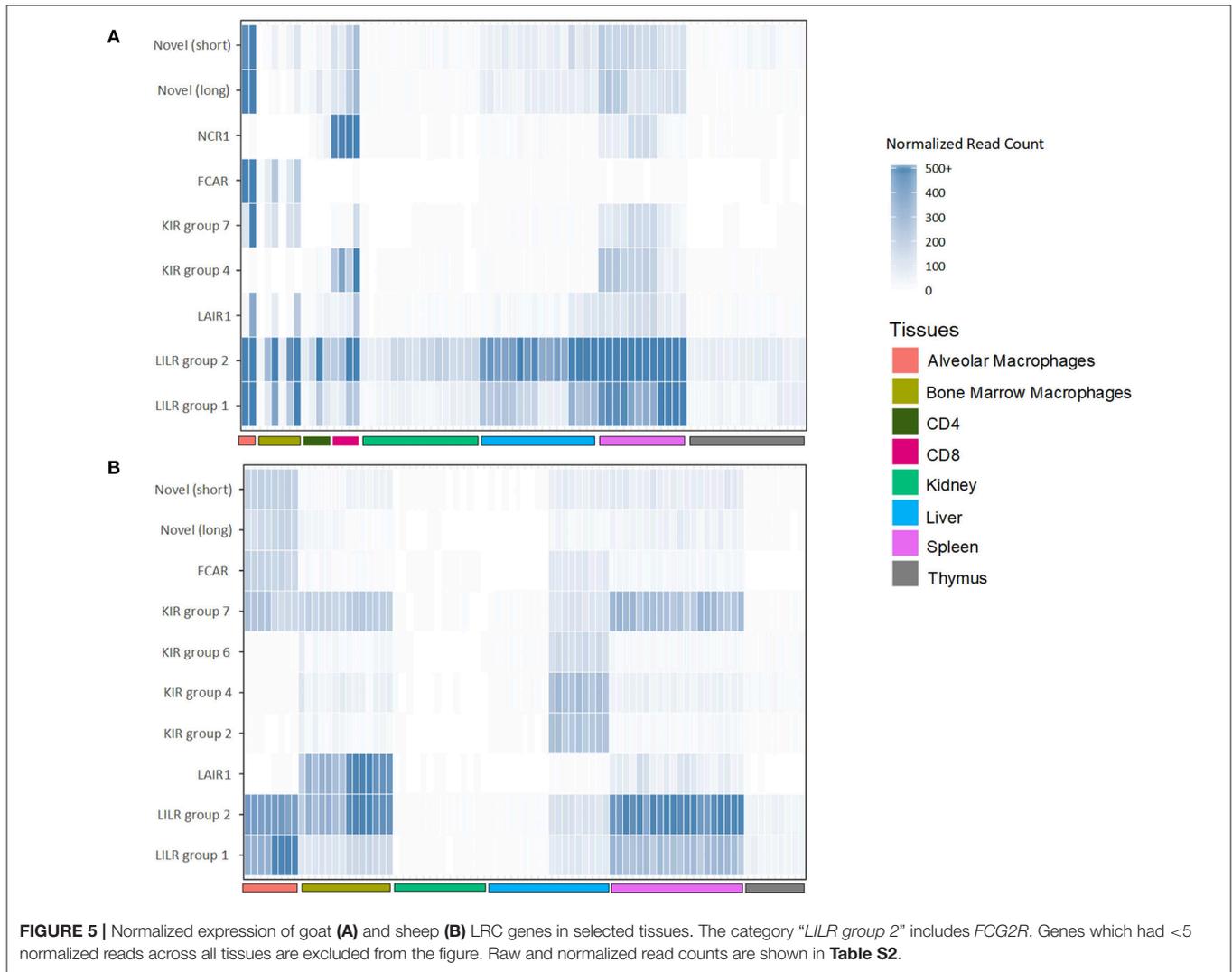
The novel group 7 *KIR* genes proximal to *FCAR* were found to be particularly conserved between goats and sheep, albeit absent from cattle. Transcripts for these receptors were found to be expressed at high levels in macrophages in both sheep and goats. Expression of the group 7 *KIR* correlated with high expression of *FCAR* in the macrophage populations (42). In contrast, expression of *NCR1* (i.e., *NKp46*) which is an NK cell marker (43), was absent, indicating that NK cell contamination was minimal. The coincident expression in macrophages of the

group 7 *KIR* and adjacent *FCAR* may suggest that these genes are under similar transcriptional control. Further work is needed to determine the ligands for these *KIR*, and whether these *KIR* genes are specifically found on macrophages, or if they are also present on other distinct cell types, such as NK cell subsets.

As we previously reported for the pig (1), the *LILR* also clade into two distinct subgroups in both sheep and goats. The transcript expression of these receptors is also similar to the pig, in that they were relatively highly expressed in leukocyte-rich tissues. The group 1 genes, however, were less expressed, likely due to the fact that there are fewer functional copies. The gene encoding FCG2R is so far unique to bovid species and is nearly indistinguishable from the group 2 *LILR* genes, particularly in short read sequencing data, and as such we could not differentiate between them.

The novel two-domain Ig-like genes that we previously described in the pig are present as three copies in both the sheep and goat genome assemblies. These genes vary considerably in content between species and can differ between haplotypes (1). Indeed, a misplaced contig in the midst of the *KIR* region in the Oar_rambouillet_v1.0 assembly suggests that sheep haplotypes may also vary in content. Similar to the observed high expression in porcine peripheral blood (1), we presently found these genes to be highly expressed in goat and sheep tissues. This expression seemed to mirror that of the *LILR*, suggesting expression by similar cell types. The function of these novel Ig-like genes remains unknown, but they are widely distributed and conserved across mammalian species, suggesting a conserved function.

Our analyses identified a large number of non-functional *KIR* pseudogenes or null alleles in both the goat and sheep

FIGURE 5 | Normalized expression of goat **(A)** and sheep **(B)** LRC genes in selected tissues. The category "*LILR group 2*" includes *FCG2R*. Genes which had <5 normalized reads across all tissues are excluded from the figure. Raw and normalized read counts are shown in **Table S2**.

assemblies and a large majority of these have the potential to encode activating receptors. This observation is consistent with the cattle *KIR* (11), human *KIR*, mouse *KLRA* (44), *KLRC* in bovids and horses (9), and *LILR* in pigs (1). Even in humans, relatively little is known about the ligand binding of activating KIR as they tend to have poor affinity for MHC class I ligands compared to the inhibitory KIR. It has recently been shown, however, that the affinity of human KIR2DS1 for HLA-C is greatly increased following human cytomegalovirus infection (45), and that human KIR2DS2 specifically recognizes conserved flaviviral peptides presented by HLA-C*0102 (46). The expansion, diversification, and eventual loss of activating receptors may therefore be driven in part by the evolutionary arms race between the host and its pathogens. It is hypothesized that once they are no longer beneficial for host survival and reproduction, activating genes are quickly lost as their retention may permit inappropriate cytotoxicity and autoimmunity (44).

The ligands for the presently discussed LRC receptors are currently unknown outside of primates and rodents. However, only recently has a host ligand for human NKp46 been identified (47), and the ligands for several human LILR are also currently unknown (17), thus highlighting a remaining challenge in leukocyte biology. Due to advances in genome sequencing and assembly, our understanding of the repetitive immune-related gene complexes in non-human species is greatly accelerating. These advancements and the knowledge of the receptors that exist and their diversity are fundamental prerequisites to identifying their ligands, and by extension, their function and involvement in immunity.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in GenBank, MH680527.

AUTHOR CONTRIBUTIONS

JH, JS, NS, DB, and TS conceived the experimental design. JS, NS, and DB performed the experiments and the analyses. JS and JH wrote and edited the manuscript. The paper was reviewed and agreed upon by all authors.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02302/full#supplementary-material

Table S1 | Gene structure of goat and sheep *KIR, LILR*, and novel Ig-like genes.

Table S2 | Raw and normalized gene expression values for LRC genes in goat and sheep tissues and cell types.

REFERENCES

1. Schwartz JC, Hammond JA. The unique evolution of the pig LRC, a single KIR but expansion of LILR and a novel Ig receptor family. *Immunogenetics*. (2018) 70:661–9. doi: 10.1007/s00251-018-1067-1
2. Higuchi DA, Cahan P, Gao J, Ferris ST, Poursine-Laurent J, Graubert TA, et al. Structural variation of the mouse natural killer gene complex. *Genes Immun*. (2010) 11:637–48. doi: 10.1038/gene.2010.48
3. Anderson SK, Ortaldo JR, Mcvicar DW. The ever-expanding Ly49 gene family: repertoire and signaling. *Immunol Rev*. (2001) 181:79–89. doi: 10.1034/j.1600-065X.2001.1810106.x
4. Bryceson YT, Foster JA, Kuppusamy SP, Herkenham M, Long EO. Expression of a killer cell receptor-like gene in plastic regions of the central nervous system. *J Neuroimmunol*. (2005) 161:177–82. doi: 10.1016/j.jneuroim.2004.11.018
5. Hoelsbrekken SE, Nylenna O, Saether PC, Slettedal IO, Ryan JC, Fossum S, et al. Cutting edge: molecular cloning of a killer cell Ig-like receptor in the mouse and rat. *J Immunol*. (2003) 170:2259–63. doi: 10.4049/jimmunol.170.5.2259
6. Trowsdale J, Barten R, Haude A, Stewart CA, Beck S, Wilson MJ. The genomic context of natural killer receptor extended gene families. *Immunol Rev*. (2001) 181:20–38. doi: 10.1034/j.1600-065X.2001.1810102.x
7. Hammond JA, Guethlein LA, Abi-Rached L, Moesta AK, Parham P. Evolution and survival of marine carnivores did not require a diversity of killer cell Ig-like receptors or Ly49 NK cell receptors. *J Immunol*. (2009) 182:3618–27. doi: 10.4049/jimmunol.0803026
8. Birch J, Ellis S. Complexity in the cattle CD94/NKG2 gene families. *Immunogenetics*. (2007) 59:273–80. doi: 10.1007/s00251-006-0189-z
9. Schwartz JC, Gibson MS, Heimeier D, Koren S, Phillippy AM, Bickhart DM, et al. The evolution of the natural killer complex; a comparison between mammals using new high-quality genome assemblies and targeted annotation. *Immunogenetics*. (2017) 69:255–69. doi: 10.1007/s00251-017-0973-y
10. Guethlein LA, Abi-Rached L, Hammond JA, Parham P. The expanded cattle KIR genes are orthologous to the conserved single-copy KIR3DX1 gene of primates. *Immunogenetics*. (2007) 59:517–22. doi: 10.1007/s00251-007-0214-x
11. Sanderson ND, Norman PJ, Guethlein LA, Ellis SA, Williams C, Breen M, et al. Definition of the cattle killer cell Ig–like receptor gene family: comparison with aurochs and human counterparts. *J Immunol*. (2014) 193:6016–30. doi: 10.4049/jimmunol.1401980

12. Storset AK, Slettedal IÖ, Williams JL, Law A, Dissen E. Natural killer cell receptors in cattle: a bovine killer cell immunoglobulin-like receptor multigene family contains members with divergent signaling motifs. *Eur J Immunol*. (2003) 33:980–90. doi: 10.1002/eji.200323710
13. Allan AJ, Sanderson ND, Gubbins S, Ellis SA, Hammond JA. Cattle NK cell heterogeneity and the influence of MHC class I. *J Immunol*. (2015) 195:2199–206. doi: 10.4049/jimmunol.1500227
14. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat Genet*. (2017) 49:643–50. doi: 10.1038/ng.3802
15. Cao W, Bover L, Cho M, Wen X, Hanabuchi S, Bao M, et al. Regulation of TLR7/9 responses in plasmacytoid dendritic cells by BST2 and ILT7 receptor interaction. *J Exp Med*. (2009) 206:1603–14. doi: 10.1084/jem.20090547
16. López-Álvarez MR, Jones DC, Jiang W, Traherne JA, Trowsdale J. Copy number and nucleotide variation of the LILR family of myelomonocytic cell activating and inhibitory receptors. *Immunogenetics*. (2014) 66:73–83. doi: 10.1007/s00251-013-0742-5
17. Zhang J, Mai S, Chen HM, Kang K, Li XC, Chen SH, et al. Leukocyte immunoglobulin-like receptors in human diseases: an overview of their distribution, function, and potential application for immunotherapies. *J Leukoc Biol*. (2017) 102:351–60. doi: 10.1189/jlb.5MR1216-534R
18. Kelley J, Walter L, Trowsdale J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet*. (2005) 1:129–39. doi: 10.1371/journal.pgen.0010027
19. Takai T. A novel recognition system for MHC class I molecules constituted by PIR. *Adv Immunol*. (2005) 88:161–92. doi: 10.1016/S0065-2776(05)88005-8
20. Yamashita Y, Fukuta D, Tsuji A, Nagabukuro A, Matsuda Y, Nishikawa Y, et al. Genomic structures and chromosomal location of p91, a novel murine regulatory receptor family. *J Biochem*. (1998) 123:358–68. doi: 10.1093/oxfordjournals.jbchem.a021945
21. Hogan L, Bhuju S, Jones DC, Laing K, Trowsdale J, Butcher P, et al. Characterisation of bovine leukocyte Ig-like receptors. *PLoS ONE*. (2012) 7:e34291. doi: 10.1371/journal.pone.0034291
22. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. (2017) 34:1812–9. doi: 10.1093/molbev/msx116

23. Clark EL, Bush SJ, Mcculloch MEB, Farquhar IL, Young R, Lefevre L, et al. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). *PLoS Genet.* (2017) 13:e1006997. doi: 10.1371/journal.pgen. 1006997

24. Bush SJ, Muriuki C, Mcculloch MEB, Farquhar IL, Clark EL, Hume DA. Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genet Sel Evol.* (2018) 50:20. doi: 10.1186/s12711-018-0391-0

25. Muriuki C, Bush SJ, Salavati M, Mcculloch MEB, Lisowski ZM, Agaba M, et al. A mini-atlas of gene expression for the domestic goat (Capra hircus) reveals transcriptional differences in immune signatures between sheep and goats. *bioRxiv [preprint].* doi: 10.1101/711127

26. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* (2015) 16:57. doi: 10.1186/s13059-015-0622-4

27. The International Sheep Genomics Consortium, Archibald AL, Cockett NE, Dalrymple BP, Faraut T, Kijas JW, et al. The sheep genome reference sequence: a work in progress. *Anim Genet.* (2010) 41:449–53. doi: 10.1111/j.1365-2052.2010.02100.x

28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

29. Kent WJ. BLAT - the BLAST-like alignment tool. *Genome Res.* (2002) 12:656–64. doi: 10.1101/gr.229202

30. Krogh A. Two methods for improving performance of an HMM and their application for gene finding. *Proc Int Conf Intell Syst Mol Biol.* (1997) 5:179–86.

31. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* (2015) 43:D222–6. doi: 10.1093/nar/gku1221

32. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* (2001) 305:567–80. doi: 10.1006/jmbi.2000.4315

33. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics.* (2000) 16:944–5. doi: 10.1093/bioinformatics/16.10.944

34. Burset M, Seledtsov IA, Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.* (2000) 28:4364–75. doi: 10.1093/nar/28.21.4364

35. Daëron M, Jaeger S, Du Pasquier L, Vivier E. Immunoreceptor tyrosine-based inhibition motifs: a quest in the past and future. *Immunol Rev.* (2008) 224:11–43. doi: 10.1111/j.1600-065X.2008.00666.x

36. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* (2004) 32:1792–7. doi: 10.1093/nar/gkh340

37. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol.* (1992) 9:678–87.

38. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* (2016) 33:1870–4. doi: 10.1093/molbev/msw054

39. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* (2009) 10:421. doi: 10.1186/1471-2105-10-421

40. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol.* (2014) 32:896–902. doi: 10.1038/nbt.2931

41. Garzón-Ospina D, López C, Cadavid LF, Patarroyo ME, Patarroyo MA. Identification and diversity of killer cell ig-like receptors in aotus vociferans, a New World Monkey. *PLoS ONE.* (2013) 8:e79731. doi: 10.1371/journal.pone.0079731

42. Monteiro RC, Van De Winkel JG. IgA Fc receptors. *Annu Rev Immunol.* (2003) 21:177–204. doi: 10.1146/annurev.immunol.21.120601.141011

43. Connelley T, Storset AK, Pemberton A, Machugh N, Brown J, Lund H, et al. NKp46 defines ovine cells that have characteristics corresponding to NK cells. *Vet Res.* (2011) 42:37. doi: 10.1186/1297-9716-42-37

44. Abi-Rached L, Parham P. Natural selection drives recurrent formation of activating killer cell immunoglobulin-like receptor and Ly49 from inhibitory homologues. *J Exp Med.* (2005) 201:1319–32. doi: 10.1084/jem.20042558

45. Van Der Ploeg K, Chang C, Ivarsson MA, Moffett A, Wills MR, Trowsdale J. Modulation of human leukocyte antigen-C by human cytomegalovirus stimulates KIR2DS1 recognition by natural killer cells. *Front Immunol.* (2017) 8:298. doi: 10.3389/fimmu.2017.00298

46. Naiyer MM, Cassidy SA, Magri A, Cowton V, Chen K, Mansour S, et al. KIR2DS2 recognizes conserved peptides derived from viral helicases in the context of HLA-C. *Sci Immunol.* (2017) 2:eaal5296. doi: 10.1126/sciimmunol.aal5296

47. Narni-Mancinelli E, Gauthier L, Baratin M, Guia S, Fenis A, Deghmane A-E, et al. Complement factor P is a ligand for the natural killer cell–activating receptor NKp46. *Sci immunol.* (2017) 2:eaam9628. doi: 10.1126/sciimmunol.aam9628