Large scale transcriptional analysis of MHC class I haplotype diversity in sheep

Received: 25 October 2023 | Revised: 15 December 2023 | Accepted: 9 January 2024

DOI: 10.1111/tan.15356

ORIGINAL ARTICLE

HLA Immune Response Genetics WILEY

Large scale transcriptional analysis of MHC class I haplotype diversity in sheep

Deepali Vasoya[1] | **Timothy Connelley**[1] | **Thomas Tzelos**[1,2] | **Helen Todd**[2] | **Keith T. Ballingall**[2]

[1]Division of Infection and Immunity, The Roslin Institute, The University of Edinburgh, Scotland, UK

[2]Moredun Research Institute, Pentlands Science Park, Scotland, UK

Correspondence
Deepali Vasoya, Division of Infection and Immunity, The Roslin Institute, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG Scotland, UK.
Email: dvasoya@ed.ac.uk

Present address
Deepali Vasoya, UK Dementia Research Institute at University of Edinburgh, Scotland, UK.

Funding information
European Union's Horizon 2020 VetBioNet project, Grant/Award Number: 731014; Innovation Fund of the Moredun Research Institute

Domestic sheep (*Ovis aries*) have been an important component of livestock agricultural production for thousands of years. Preserving genetic diversity within livestock populations maintains a capacity to respond to changing environments and rapidly evolving pathogens. MHC genetic diversity can influence immune functionality at individual and population levels. Here, we focus on defining functional MHC class I haplotype diversity in a large cohort of Scottish Blackface sheep pre-selected for high levels of MHC class II *DRB1* diversity. Using high-throughput amplicon sequencing with three independent sets of barcoded primers we identified 134 MHC class I transcripts within 38 haplotypes. Haplotypes were identified with between two and six MHC class I genes, plus variable numbers of conserved sequences with very low read frequencies. One or two highly transcribed transcripts dominate each haplotype indicative of two highly polymorphic, classical MHC class I genes. Additional clusters of medium, low, and very low expressed transcripts are described, indicative of lower transcribed classical, non-classical and genes whose function remains to be determined.

KEYWORDS
amplicon, haplotypes, MHC class-I, ovine, polymorphic, sequencing

1 | INTRODUCTION

The most recent statistics from the Food and Agriculture Organisation of the United Nations (FAO) estimate that there are currently 1.2 billion domestic sheep (*Ovis aries*), producing animal protein, fibre (wool), and dairy products across a range of agricultural systems worldwide (https://www.fao.org/livestock-systems/global-distributions/sheep/en/).

Conventional selective breeding for different production traits within a broad range of environmental conditions has resulted in large numbers of specialised sheep breeds, each with different phenotypic characteristics and local adaptations.[1] Many of these breeds are now maintained in small numbers by dedicated farmers and breed societies due to the increasing dominance in mainstream farming of a small number of improved breeds with impressive production traits. Consequently, many local breeds are threatened with extinction or have already been lost.

Preserving genetic diversity within livestock populations maintains a capacity to adapt to changing

environments and to respond to rapidly evolving, novel, or reemerging pathogens. The major histocompatibility complex (MHC) is an important source of immunogenetic diversity within the genome of jawed vertebrates.[2] Covering between 2 and 3 million bases on chromosome 20 in sheep,[3] the MHC region contains the highly polymorphic MHC class I and MHC class II genes and approximately 160 other genes, many of which are involved in innate or adaptive immunity.[3,4]

The MHC class I and class II genes encode cell surface proteins which are recognised by antigen specific receptors on T-cells.[5,6] The T-cell receptor (TCR) recognises the MHC molecule in combination with an antigenic peptide loaded within a binding groove on the surface of the MHC molecule. Recognition by the TCR of the combination of MHC molecule and antigenic peptide provides the initial T cell activation signal with CD4+T cells recognising MHC class II/peptide antigen combinations while CD8+ T cells recognise the combination of MHC class I/peptide antigen.[7] Generally, antigens in combination with MHC class II molecules originate from the extracellular environment while those in combination with MHC class I molecules are derived internally and may include viral, bacterial or intracellular parasite antigens.

Gene duplication and high levels of allelic polymorphism are features of the MHC in most vertebrate species. The Human MHC or HLA includes three functional class I genes, HLA-A, B and C present on all haplotypes, each with classical features including high levels of transcription, high levels of allelic polymorphic and an almost ubiquitous transcriptional distribution. These genes provide the classical MHC class I proteins which present endogenous peptide antigens to CD8+ve T-cells. Other functional HLA class I genes found in all haplotypes include *HLA-E*, *F*, and *G*. These have non-classical features including lower levels of transcription, limited allelic polymorphism, a more restricted tissue distribution and more specialised functions including immune modulation through NK cell regulation and maternofoetal tolerance.[8,9]

Twelve HLA-class I pseudogenes and gene fragments are also listed in the IPD-IMGT/HLA Database (https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/) from which no protein products or function has yet been assigned. This reflects the rapid evolution of the MHC class I region through the birth and death of genes through duplication and recombination. Consequently, direct functional orthologues of classical and non-classical HLA class I genes are not found in other mammalian species.[10]

The nucleotide sequence of many thousands of alleles of each of the genes within the HLA complex are listed in the IPD-IMGT/HLA Database, which provides the official nomenclature for each allele and is used as the template for MHC nomenclature in most other species including sheep.[11]

We have a long-term interest in MHC diversity in sheep to assist our understanding of immune variation to infection and to support development of new and improved vaccines against diseases that impact livestock production. The ability of evaluate the induction of MHC class I restricted T-cell responses to novel vaccine antigens requires tools to monitor antigen specific T cell activation and proliferation over time. MHC tetramer reagents allow the dynamics of antigen-specific T cell activation and proliferation to be followed at the single cell level.[12,13] These reagents are not generally available for livestock species as they require animals with well-defined MHC haplotypes to maintain histocompatibility between the tetramer reagent and the antigen specific TCR. By identifying common MHC class I haplotypes within the Scottish Blackface breed, we hope to be able to generate panels of tetramer reagents based on common MHC molecules within defined haplotypes, allowing long-term applications unrestrained by individual animals' MHC repertoire.

To begin to understand the complexity of the MHC region in sheep we have previously generated MHC homozygous animals representing four haplotypes, through sire daughter mating in the Scottish Blackface breed.[14] The genetic resource provided by these animals delivered many of the reference sequences used to develop MHC nomenclature for sheep and to populate the allelic databases maintained in IPD-MHC, https://www.ebi.ac.uk/ipd/mhc/group/OLA/.[15–17] These studies used RT-PCR, amplicon cloning and Sanger sequencing to characterise the functional MHC class I and class II diversity in four haplotypes, however, this approach is not amenable to large cohorts both in terms of time and allele inclusiveness.

The development of high through-put, sequence-based MHC typing technologies has been driven by the need to define allelic diversity in HLA class I and class II genes due to their central role in allogeneic tissue transplantation.[18] In this study, we aim to expand on recent development of this technology for cattle[19–21] to analyse MHC haplotype diversity in a large cohort of Scottish Blackface sheep.

2 | METHODS AND MATERIALS

2.1 | Animals

Four geographically distinct Scottish Blackface sheep flocks were selected from Western Scotland (Flock A),

Northeast Scotland (Flock B) Southern Scotland (Flock C), and Central Scotland (Flock D). Venous blood samples from 50 lambs per flock were collected by jugular cannulation into vacutainers containing EDTA as an anticoagulant. The generation of MHC homozygous Scottish Blackface and the inbred Prealpe sheep lines has been described previously.[16,22]

2.2 | Sample preparation

From 10 mL of whole blood the buffy coat layer containing white bloods cells was isolated by centrifugation for 10 min at 2500*g* at room temperature. Buffy coat cells were collected, and 5 mL of red blood cell lysis buffer (Qiagen, catalogue no. 79217) added, and cells incubated on ice for 15 min with occasional mixing. The white blood cells were pelleted by centrifugation at 1200*g* at 4°C for 5 min, washed in 5 mL cold PBS and suspended to a final volume of 180 μL in PBS. Two 30 μL aliquots were archived for genomic DNA preparation and four 30 μL aliquots archived for total RNA preparation. The archive was stored in 1.5 mL Eppendorf tubes at −80°C until required.

2.3 | Genomic DNA preparation

Genomic DNA was extracted from a single archived Eppendorf tube using the Qiagen DNeasy kit according to the manufacturer's instructions. Genomic DNA was quantified using a Nanodrop spectrometer and archived at −20°C until required.

2.4 | Total RNA preparation

Three hundred microlitre of TRIzol reagent (Invitrogen) was added to an archived sample and pipetted several times to lyse the cells. RNA was extracted from the lysate by adding 100 μL of chloroform followed by shaking for 15 s. The tube was centrifuged at full speed (14,000*g*) for 15 min at 4°C and the upper aqueous phase collected. RNA was precipitated by addition of 200 μL of isopropanol, and incubated at room temperature for 10 min. Following centrifugation, the RNA pellet was washed in 500 μL of 70% ethanol, air-dried and suspend in 20 μL of RNAse free water prior to nanodrop quantification.

2.5 | Preparation of first strand cDNA

First strand cDNA synthesis was prepared using the ImProm-II Reverse Transcription System (Promega) according to the manufacturer's instructions. Briefly, the Reverse Transcriptase, 5× Reaction Buffer and dNTP's were thawed and kept on ice prior to use. In sterile, thin-walled 200 μL PCR tubes we combined 200 ng of total RNA with the oligo dT primer (0.5 μg/20 μL reaction) and nuclease free water to a final volume of 5 μL. The tube was heated at 70°C for 5 min and immediately chilled on ice water for 5 min. In a second 200 μL PCR tube 4 μL 5× reaction buffer, 2 μL MgCl, 1 μL dNTP mix, 1 μL RNase inhibitor, and 1 μL reverse transcriptase enzyme was combined with 6 μL nuclease free water to a final volume of 15 μL. The 5 μL RNA primer mix was then added, mixed with the end of a pipette tip and cycled at 25°C for 5 min at 42°C for 1 h before cooling to 4°C. The first strand cDNA was stored at −80°C. Prior to use the stock cDNA was diluted 1:5 in nuclease free water.

2.6 | MHC class II *DRB1* genotyping

Sanger sequence-based genotyping of the highly polymorphic MHC class II *DRB1* gene using genomic DNA template was performed as described by Ballingall and Tassi[23] with the primer modifications detailed in Reference 24 Validation of *DRB1* homozygous animals was carried out using cDNA as the template and primers 222 and 223 (Table 1), located within exons one and three of the *DRB1* transcript. RT-PCR transcripts of 408 bp were gel purified and sequenced in both directions. Individual *DRB1* alleles were defined as described for genomic DNA typing and each allele named according to IPD-MHC nomenclature.[16]

2.7 | Selection of animals for MHC class I genotyping

A total of 96 animals were selected for MHC class I genotyping. Eighty-six Scottish Blackface animals were selected based on the following criteria: (i) five animals with each of the five most common *DRB1* alleles from each of the four flocks, to provide an indication of the stability of these haplotypes within and between flocks, and (ii) at least one representative of each of the other *DRB1* alleles, to provide a global picture of haplotype diversity across the breed. Ten control animals representing a range of MHC haplotypes were included to validate the typing methods. These included MHC homozygous Scottish Blackface sheep, purpose bred by sire daughter mating along with the two paternal rams providing each of these haplotypes.[14] RNA from three lines of Prealpe sheep, homozygous at the MHC through multiple rounds

TABLE 1 List of PCR primer sequences used in this study.

PCR primer sets	Target gene	Amplification length	Forward primer sequence	Reverse primer sequence
Primer set 1	MHC class I	410 bp	GTCGGCTAYGTGGACGAC	CTCCAGGTRTCTGMGSAGC
Primer set 2	MHC class I	339 bp	GGGCCGGAGWAWWGGGA	GCAGCGTGTCCTTCCCG
Primer set 3	MHC class I	459 bp	CGGCTACGTGGACGACAYG	ATGGGTCACATGTGYCTTTG
Primers 222 and 223	MHC class II-DRB1	408 bp	GCTCTGATAGTGATGCTGATGG	ATTCACAGAGCAGACCAGGAG

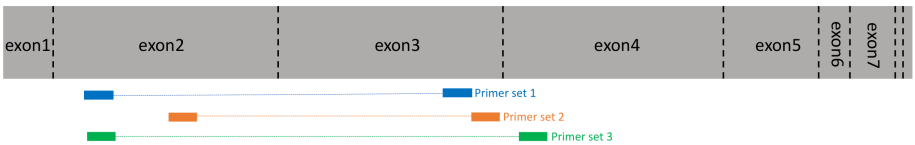

FIGURE 1 Primer binding sites. The positions of the forward and reverse primer binding sites, which encompass exon 2 and exon 3 of a typical Ovine MHC Class I transcript are displayed. The binding sites for all three PCR primer sets are indicated with different colours: blue, orange and green indicates primer set 1, 2 and 3, respectively.

of inbreeding were also included for breed comparative purposes.

2.8 | MHC class I amplification

Three sets of primers (Table 1) were used for MHC class I genotyping to reduce the likelihood of sequence dropout, control for primer bias and to allow relative quantification of MHC class I sequences. Briefly, two primer sets (primer set 1 and primer set 2) amplify overlapping 410 and 339 bp fragments respectively; combining to cover 431 bp of the highly polymorphic regions of the second and third exons of MHC class I transcripts. The third primer set (primer set 3) amplifies a slightly larger 459 bp fragment covering the second to fourth exons. Each primer incorporates Illumina adaptors and a unique molecular identifier (MID) sequence (i5 and i7 index) specific for each animal. The location of each of the three primer sets is shown in Figure 1. Amplification conditions, proof reading Taq polymerase, library construction and paired-end sequencing using Illumina MiSeq v3 chemistry are as described in Reference 19.

2.9 | Bioinformatics pipeline

Sequencing reads were segregated based on MID combinations into 96 samples. The raw data was assessed for quality (trimmed at threshold score of >Q28) and paired-end sequences overlapped using FLASH.[25] Data were processed using the bioinformatics pipeline developed for bovine MHC class I as described in Reference 19 with the modifications described in Reference 20. As part of the bioinformatics pipeline, it was necessary to introduce a read count threshold filter to remove low frequency read sequences generated as a consequence of nonspecific PCR and/or sequencing errors. To empirically define this threshold, we presumed 0.2% as defined in the bovine MHC class I analysis pipeline. We assessed all transcript sequences that are observed in multiple animals with >0.2% read frequency with each PCR primer set or sequences that are observed in a single animal with >1% read frequency in one of the PCR primers. Downstream analysis of haplotypes and individual transcripts including calculating frequency, associations and linkage were carried out using scripts deposited at (https://github.com/deepalivasoya/MHCtyping).

2.10 | Nomenclature

Novel MHC sequences identified in this study were named according to guidelines provided by the IUIS/ISAG Comparative MHC Nomenclature Committee.[11] The sequences from the three primer sets were overlapped and extended to generate a 460 bp reference sequence for each transcript. All nucleotide sequences were translated to amino acid sequences and compared with the data held in the IPD-MHC Database.[17] If transcripts showed ≤4 amino acid differences, within the region encoded by exons 2 and 3, from an allele with an official nomenclature, it was named within the same allelic group and given an alphabetical suffix such as *Ovar-N*12:AA* to identify it as a novel allele in that subgroup and to avoid confusion with official nomenclature which

uses numerical suffixes. Sequences with >4 amino acid differences from sequences in the IPD-MHC Database were considered to represent novel allelic groups. Due to amplicon size, we could not assign such MHC class I sequences to specific groups. These were given the prefix Ovar-MHCI* and the allele group assigned by two alphabetical letters such as *Ovar-MHCI*AA:01*. Alleles with only synonymous substitutions from a sequence within IPD-MHC were assigned a letter to the allelic nomenclature (e.g., *Ovar-N*12:01:AA*) while a two-digit number was assigned to sequences that differed by synonymous substitution from a novel allele group (e.g., *Ovar-MHCI*AA:01:01*).

2.11 | MHC class I haplotype classification

Haplotypes were assigned and validated by identification of identical combinations of class I sequences in a minimum of two heterozygous animals. Haplotypes may also be derived when they appear in a single heterozygous animal when the second haplotype is already validated. Such haplotypes remain classified as 'unvalidated'. Haplotypes were each given the prefix HP and assigned a unique number corresponding to their order of identification, for example, HP:08 was the eighth haplotype identified. Haplotypes that occurred in only a single heterozygous animal were assigned the prefix unHP, for unvalidated haplotype. Haplotypes that occurred in a single purpose bred MHC homozygous animal such as the Prealpe haplotypes were also considered as validated.

2.12 | Estimates of transcript abundance

Transcript abundance was estimated by comparison of the relative read frequency data from all three PCR reactions. It involved calculating the relative expression frequencies of each transcript within each haplotype. A line of best fit described by the equation $y = b + a{\cdot}x$, where a is the slope and b is the y-intercept was calculated for each of the three combinations of primers (Data S1). The calculated slope (between 0.60 and 1.50) and the correlation coefficient (>0.80) were used to predict if the sequence and corresponding haplotype has strong correlation between each of the three PCR reactions and the hierarchical frequency of transcripts within each haplotype. Read frequencies were generally consistent across all three sets of primers and haplotypes which allowed transcripts to be assigned to high (range 25%–97%), medium (10%–24%), low (2%–9%) or very low (0.1%–1.9%) expression level groups.

2.13 | MHC class I and class II linkage

To determine the frequency at which MHC class I haplotypes and class II DRB1 alleles occur together, a pairwise comparison was used to calculate the normalised frequency of each DRB1 allele with each MHC class I haplotype. MHC class I haplotype and DRB1 in each animal were linked using the highest normalised frequency between them. The frequency at which each MHC class I haplotype and DRB1 allele occurred together was plotted in an alluvial diagram.

2.14 | Evolutionary analysis

Multiple alignments of the MHC class I transcripts generated here were produced using CLUSTAL Omega available on the EMBL-EBI website http://www.ebi.ac.uk/Tools/msa/clustalo/. The multiple alignment which included published sequences held in the sheep section of the IPD-MHC Database (downloaded August 2023) was used to estimate maximum likelihood trees. Prior to tree estimation, the optimum nucleotide substitution model was selected using the model selection feature[26] launched in IQ-TREE.[27] The optimum substitution model selected for the MHC class I sequences was the transversion model TVM + F + R4, which reflects variable base frequencies, variable transversion rates and equal transition rates. The topology of the tree was tested with 2000 bootstrap replicates using the ultrafast boost strap method of Reference 28.

3 | RESULTS

3.1 | MHC class II *DRB1* analysis and selection of animals for class I haplotype analysis

Sanger sequence-based typing defined allelic diversity at the MHC class II-*DRB1* locus in 200 Scottish Blackface sheep from four geographically distinct flocks. In total, 18 *Ovar-DRB1* alleles were identified with allelic frequencies ranging from 0.3 to 0.01 (Table 2). All 18 alleles had previously been submitted to the IPD-MHC Database and assigned official nomenclature. Between 11 and 14 *DRB1* alleles were identified in each of the four flocks with many alleles shared between flocks. The *Ovar-DRB1*01:01* allele was found at the highest frequency in each of the four flocks. Eighty-six sheep were selected for MHC class I haplotype analysis based on their DRB1 allelic diversity to ensure that the repertoire of class I genes linked to each DRB1 allele could be assigned and

TABLE 2 Ovar-MHC class II-DRB1 allele frequencies in four different farms.

Allele *Ovar*	Farm A	Farm B	Farm C	Farm D
*DRB1*01:01*	0.3	0.19	0.25	0.21
*DRB1*01:02*	0.01		0.02	0.01
*DRB1*03:01*	0.15	0.07	0.16	0.13
*DRB1*03:02*		0.03		0.07
*DRB1*03:04*	0.01	0.02		
*DRB1*03:08*	0.16	0.03	0.01	0.08
*DRB1*04:04*	0.01	0.13		
*DRB1*05:01*	0.06	0.17	0.05	0.04
*DRB1*07:01*	0.02		0.01	0.06
*DRB1*08:02*		0.09	0.06	0.01
*DRB1*09:01*	0.09		0.09	0.08
*DRB1*10:01*		0.03	0.03	
*DRB1*10:02*			0.16	0.04
*DRB1*11:01*		0.08	0.07	
*DRB1*12:01*	0.17	0.03	0.02	0.21
*DRB1*13:01*		0.04	0.01	0.04
*DRB1*18:02*				0.02
*DRB1*22:01*	0.02	0.09	0.06	

validated in multiple animals. Ten additional MHC homozygous Blackface and Prealpe animals were also included to validate the methodology and to provide breed comparison.

3.2 | MHC class I sequence analysis

High quality MHC class I sequence data was generated from 94 of the 96 samples using all three primer sets. Two animals were removed due to cross-contamination (BF119 and PA9). Each of the three primer sets provided high quality MHC class I sequence data from the three MHC homozygous Blackface animals. Each provided a single MHC class I haplotype consistent with previous analysis of their heterozygous sires and homozygous offspring. The inbred French Prealpe animals also provided single MHC class I haplotypes consistent with previous analysis.

Analysis of sequence data from the full cohort identified a total of 134 distinct MHC class I sequences. Of the 134 class I transcripts, 14 were identical to sequences currently held in the IPD-MHC Database and to which official names have been assigned. One-hundred and twenty were considered novel as no match was found with a sequence in the IPD-MHC Database. The nucleotide and predicted amino acid sequences of the novel sequences are shown in Data S2 and S3, respectively. All novel sequences have been submitted to the Genbank Database with accessions listed in Table 3. Based on the nomenclature system defined by the IUIS/ISAG Comparative MHC Nomenclature Committee, the 120 novel sequences were assigned to 62 new allelic groups, 24 subgroups of these new allelic groups and 5 alleles with only synonymous substitutions within a new allelic group. Also, 25 new subgroups and 3 alleles with only synonymous substitutions compared with a sequence already deposited within the IPD-MHC Database, were identified.

3.3 | PCR correlation analysis

Some amplification bias was associated with each of the three primer sets however, paired correlation analysis identified strong correlation coefficients ($r = 0.853$, $r = 0.945$ and $r = 0.811$) in comparisons of the three primer sets in the amplification of class I transcripts from each animal (Figure 2). Of the 134 sequences identified, 12 were identified using only primer 1 or primer 2 sets and 27 were identified using combinations of two of the three primer sets (Table 3). All remaining sequences were identified with all three primer sets. The correlation analysis was carried out for each combination of sequences inherited together within a haplotype. The data for each haplotype are shown in Data S1. Strong correlation between all three primers sets was observed in 17 of 28 validated haplotypes, for example HP:19 (Figure 3). Most of the PCR bias in the remaining haplotypes was associated with low or very low expressed transcripts, however, in two haplotypes, HP:16 and HP:10, PCR bias was identified in high and medium expressed genes (Figure 3). The only example of drop out of a high or medium transcribed sequence occurred in HP:10. In 14 HP:10 positive animals the *Ovar-MHCI*BX:01* transcript was amplified and sequenced with high read frequencies using primer sets 1 and 3 but it failed to amplify with primer set 2.

3.4 | MHC class I haplotype assignment and distribution

MHC class I haplotypes were successfully assigned in 93 of the 94 animals. A total of 38 haplotypes were identified of which 10 remained unvalidated as they occurred in only a single heterozygous animal. The haplotypes in one animal could not be characterised as two possible options were identified. Including the seven purpose-bred MHC homozygous animals, 12 animals were homozygous at the MHC class I locus. Eleven of these were also

TABLE 3 List of 134 MHC class I alleles identified in this study, including details on number of assigned haplotypes, total observation, corresponding PCR primer sets, read frequency in each PCR primer set and Genbank accession ids for novel alleles. The novel alleles are highlighted in bold. The 31 unassigned alleles form seven allelic groups, *MHCI*AA* to *MHCI*AG*, are highlighted with underline.

Alleles *Ovar*	Number of assigned haplotypes	Number of observations	Average read frequency for primer set1	Average read frequency for primer set2	Average read frequency for primer set3	Associated PCR primers	Genbank accession ID
MHCI*AA:01:01	0	87	2.12	0.21	0.88	1, 2, 3	OP957131
MHCI*AA:01:02	0	19	1.08	0.11	0.36	1, 2, 3	OP957238
MHCI*AA:01:03	0	23	1.41	0.23	0.77	1, 2, 3	OP957175
MHCI*AA:02	0	8	0.91	0.06	0.63	1,2,3	OP957121
MHCI*AA:03	0	2	0.64	0.18	0.49	1, 2, 3	OP957123
MHCI*AB:01:01	0	6	0.32			1	OP957179
MHCI*AB:01:02	0	17	0.31			1	OP957164
MHCI*AB:03	0	20	0.25			1	OP957141
MHCI*AB:04	0	11	0.21			1	OP957161
MHCI*AB:05	0	4	0.34			1	OP957192
MHCI*AB:06	0	2	0.42			1	OP957204
MHCI*AC:01:01	0	12	0.31	0.29		1, 2	OP957184
MHCI*AC:01:02	0	18	0.41	0.4		1, 2	OP957143
MHCI*AC:01:03	0	16	0.21	0.43		1, 2	OP957152
MHCI*AC:04	0	5	0.32	0.3		1, 2	OP957197
MHCI*AC:05	0	1	0.63	1.11		1, 2	OP957230
MHCI*AC:06	0	19	0.27	0.38		1, 2	OP957160
MHCI*AD:01	0	28	0.07	0.72		1, 2	OP957120
MHCI*AD:02	0	7		0.23		2	OP957126
MHCI*AD:03	0	10		0.29		2	OP957125
MHCI*AE:01	0	16	0.25	0.19		1, 2	OP957183
MHCI*AE:02	0	49	0.26	0.3		1, 2	OP957130
MHCI*AE:03	0	8	0.3	0.28		1, 2	OP957155
MHCI*AF:01	0	16	0.16	0.18		1,2	OP957208
MHCI*AF:02	0	37	0.19	0.39		1, 2	OP957165
MHCI*AF:03	0	8	0.26			1	OP957122
MHCI*AF:04	0	12	0.16	0.46		1, 2	OP957210
MHCI*AG:01	0	3	0.47	0.84	0.44	1, 2, 3	OP957186
MHCI*AG:02	0	1	0.31	0.57	0.21	1, 2, 3	OP957119
MHCI*AG:03	0	18	0.27	0.5		1, 2	OP957150
MHCI*AH:01	2	10		0.18	0.06	2, 3	OP957127
MHCI*AI:01	1	16	0.75	2.23		1, 2	OP957134
MHCI*AJ:01	0	2	0.49			1	OP957185
MHCI*AK:01	3	26	1.55			1	OP957129
MHCI*AL:01	1	10	11.73	13.09	10.18	1, 2, 3	OP957146
MHCI*AM:01	1	3	20.23	16.37	3.52	1, 2, 3	OP957118
MHCI*AN:01	1	7	10.83	10.81	6.51	1, 2, 3	OP957151
MHCI*AO:01	1	2	20.27	22.81	12.42	1, 2, 3	OP957201
MHCI*AP:01	1	14	8.58	4.11	8.25	1, 2, 3	OP957140
MHCI*AP:02	1	3	7.97	3.05	4.46	1, 2, 3	OP957193

(Continues)

TABLE 3 (Continued)

Alleles *Ovar*	Number of assigned haplotypes	Number of observations	Average read frequency for primer set1	Average read frequency for primer set2	Average read frequency for primer set3	Associated PCR primers	Genbank accession ID
MHCI*AQ:01	1	4	10.84		5.4	1, 3	OP957182
MHCI*AQ:02	0	2	12.09		7.08	1, 3	OP957221
MHCI*AR:02	1	5	5.06	5.85	4.32	1, 2, 3	OP957174
MHCI*AR:03	1	4	5.17	6.36	4.06	1, 2, 3	OP957191
MHCI*AR:04	0	2	6.93	6.41	4.27	1, 2, 3	OP957213
MHCI*AS:01	0	1	8.94	8.29	4.95	1, 2, 3	OP957226
MHCI*AT:01	0	2	27.48	24.98	28.06	1, 2, 3	OP957224
MHCI*AU:01	0	1	16.28	20.3	13.48	1, 2, 3	OP957236
MHCI*AV:01	0	1	15.33	17.98	11.96	1, 2, 3	OP957220
MHCI*AV:02	0	1	23.79	31.35	17.37	1, 2, 3	OP957216
MHCI*AW:01	0	1	21.88	34.78	14.35	1, 2, 3	OP957229
MHCI*AX:01	0	1	16.88	19.11	11.61	1, 2, 3	OP957232
MHCI*AY:01	1	7	0.19		0.04	1, 3	OP957211
MHCI*AY:02	1	4	0.24			1	OP957207
MHCI*AZ:01	0	1	1.46	2.72	1.33	1, 2, 3	OP957225
MHCI*BA:01	0	1	25.85	37.33	24.31	1, 2, 3	OP957212
MHCI*BB:01	0	1	13.07	8.83	14.63	1, 2, 3	OP957222
MHCI*BC:01	0	1	5.44	7.83	4.41	1, 2, 3	OP957237
MHCI*BD:01	2	11	1.31	3.5	1.13	1, 2, 3	OP957147
MHCI*BE:01	1	7		1.84	0.58	2, 3	OP957124
MHCI*BF:01	1	7	9.08	8.21	5.18	1, 2, 3	OP957162
MHCI*BG:01	1	5	30	40.19	25.56	1, 2, 3	OP957176
MHCI*BH:01	1	8	0.49		0.47	1, 3	OP957156
MHCI*BI:01	1	5	12.61	15.21	15.87	1, 2, 3	OP957180
MHCI*BJ:01	1	5	16.97	23.6	15.34	1, 2, 3	OP957181
MHCI*BK:01	1	7	0.22		0.41	1, 3	OP957189
MHCI*BL:01	1	5	19.22	27.87	14.5	1, 2, 3	OP957177
MHCI*BM:01	1	6	17.22	2.04	10.98	1, 2, 3	OP957178
MHCI*BN:01	1	6	14.59	26.09	18.28	1, 2, 3	OP957170
MHCI*BO:01	1	4	14.77	28.46	12.63	1, 2, 3	OP957187
MHCI*BP:01	1	3	9.78		11.83	1, 3	OP957190
MHCI*BQ:01	1	15	22.83	28.65	24.17	1, 2, 3	OP957136
MHCI*BR:01	1	3	5.17	8.5	9.2	1, 2, 3	OP957188
MHCI*BS:01	2	17	3.37	8.9	0.62	1, 2, 3	OP957132
MHCI*BT:01	1	2	18.27	24.7	17.64	1, 2, 3	OP957198
MHCI*BU:01	0	2	9.92	3.91	12.19	1, 2, 3	OP957200
MHCI*BV:01	1	2	5.8		9.03	1, 3	OP957205
MHCI*BW:01	1	2	14	3.96	11.04	1, 2, 3	OP957202
MHCI*BX:01	1	14	7.51		7.81	1, 3	OP957138
MHCI*BY:01	1	6	0.23	0.65	0.19	1, 2, 3	OP957209
MHCI*BZ:01	1	7	14.87	18.86	19.16	1, 2, 3	OP957157

TABLE 3 (Continued)

Alleles *Ovar*	Number of assigned haplotypes	Number of observations	Average read frequency for primer set1	Average read frequency for primer set2	Average read frequency for primer set3	Associated PCR primers	Genbank accession ID
MHCI*CA:01	1	2	19.29	28.34	13.97	1, 2, 3	OP957206
MHCI*CB:01	0	1	23.11	18.96	15.86	1, 2, 3	OP957234
MHCI*CC:01	0	1	17.7	19.55	13.9	1, 2, 3	OP957218
MHCI*CD:01	0	1		15.93	7	2, 3	OP957128
MHCI*CE:01	0	1	17.2	17.38	21.37	1, 2, 3	OP957214
MHCI*CF:01	1	7	12.97	11.49	7.65	1, 2, 3	OP957149
MHCI*CG:01	1	11	17.86	10.65	18.12	1, 2, 3	OP957148
MHCI*CH:01	1	10	14.87	17.76	14.75	1, 2, 3	OP957145
MHCI*CI:01	1	5	2.65	4	2.19	1, 2, 3	OP957168
MHCI*CJ:01	0	1	3.41	2.93	1.81	1, 2, 3	OP957233
MHCI*CK:01	0	1	8.33	12.35	7.39	1, 2, 3	OP957219
*N*01:01*	1	7	20.13	23.93	17.49	1, 2, 3	
*N*02:01*	1	4	14.44	19.68	11.29	1, 2, 3	
N*02:AA	1	7	14.53	22.53	12.48	1, 2, 3	OP957159
*N*03:01*	1	3	23.98	28.31	22.12	1, 2, 3	
*N*04:01*	1	15	21.13	12.9	25.1	1, 2, 3	
*N*05:01*	1	24	28.89	36.85	32.34	1, 2, 3	
*N*06:01*	1	23	1.32	2.2	1.04	1, 2, 3	
*N*07:01*	1	3	4.62	7.12	1.33	1, 2, 3	
N*07:AA	1	4	9.01	10.8	2.13	1, 2, 3	OP957194
*N*08:01*	2	12	4.97	5.8	2.88	1, 2, 3	
N*08:01:AA	1	6	3	6.6	2.3	1, 2, 3	OP957163
*N*12:01*	0	2	11.49	8.11	4.58	1, 2, 3	
N*12:01:AA	2	8	7.16	7.23	3.75	1, 2, 3	OP957153
N*12:AB	1	7	3.45	5.14	3.06	1, 2, 3	OP957154
N*12:AC	1	6	3.22	8.41	1.49	1, 2, 3	OP957166
N*12:AD	1	5	4.98	5.32	2.61	1, 2, 3	OP957169
N*12:AE	1	5	6.12	8.81	2.53	1, 2, 3	OP957173
N*12:AF	1	15	4.13	3.81	2.88	1, 2, 3	OP957135
N*12:AG	1	4	7.33	6.12	3.07	1, 2, 3	OP957195
N*12:AI	1	7	2.88	2.93	1.54	1, 2, 3	OP957158
N*12:AJ	0	1	5.07	4.88	4.19	1, 2, 3	OP957215
N*12:AK	2	14	3.04	4.79	2.47	1, 2, 3	OP957137
N*12:AL	1	2	3.4	3.96	6.08	1, 2, 3	OP957203
N*12:AM	1	5	4.46	4.41	4.84	1, 2, 3	OP957171
N*12:AN	0	1	4.51	6.58	2.51	1, 2, 3	OP957227
N*12:AO	0	1	5.36	5.66	2.97	1, 2, 3	OP957231
*N*24:01*	0	2	1.77	9.89	3.18	1, 2, 3	
N*26:AA	1	6	13.64	9.23	17.11	1, 2, 3	OP957167
N*27:AA	1	5	20.43	32.02	32.2	1, 2, 3	OP957196
*N*50:00*	1	23	5	5.43	1.84	1, 2, 3	
*N*50:01*	2	12	1.03	1.85	1.64	1, 2, 3	

(Continues)

TABLE 3 (Continued)

Alleles *Ovar*	Number of assigned haplotypes	Number of observations	Average read frequency for primer set1	Average read frequency for primer set2	Average read frequency for primer set3	Associated PCR primers	Genbank accession ID
N*50:01AA	0	1	3.62	6.55	3.25	1, 2, 3	OP957223
N*50:01AB	0	1	1.65	2.64	1.57	1, 2, 3	OP957235
*N*50:02*	1	3	2.52	0.02	2.31	1, 2, 3	
N*50:02:AA	1	5	1.91		1.82	1, 3	OP957172
N*50:02AB	0	1	2.19	2.81	3.12	1, 2, 3	OP957228
*N*50:03*	1	15	10.98	18.33	9.79	1, 2, 3	
N*50:03AB	2	13	1.17	1.01	1.38	1, 2, 3	OP957142
N*50:03AC	1	10	2.49	3.1	2.72	1, 2, 3	OP957144
N*50:03AD	2	17	2.05	2.67	2.49	1, 2, 3	OP957133
N*50:03AE	1	2	0.82	1.16	0.43	1, 2, 3	OP957199
N*50:03AF	5	17	1.42	1.84	1.34	1, 2, 3	OP957139

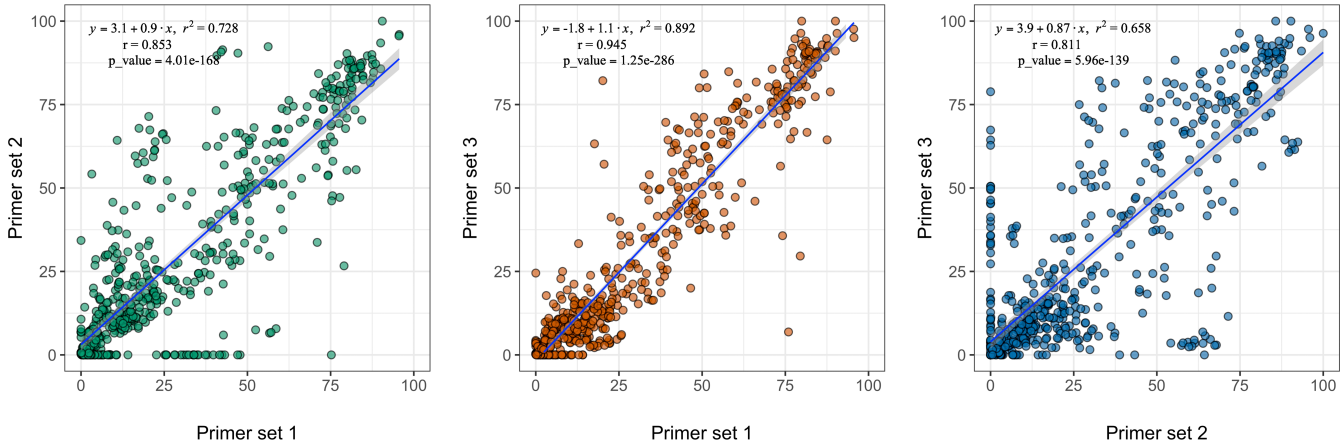

FIGURE 2 Correlation of class I amplification from each animal across the three primer sets. For each haplotype, the normalised relative read frequency for each transcript was calculated, for example, if haplotype 1 contains three transcripts, A, B and C, with read frequencies of 17%, 3% and 2%, the relative frequencies are 77%, 14% and 9%, respectively. A linear regression line was plotted for the relative frequencies in pairwise comparisons between primer sets represented by green for primer set 1 versus 2, orange for primer set 1 versus 3 and blue for primer set 2 versus 3. For each primer combination, a line of best fit, the correlation coefficient and *p*-value were calculated.

homozygous at the MHC class II *DRB1* locus providing an indicative frequency of homozygosity (4/85) across the MHC region in this sample of Blackface sheep of 0.047.

The distribution of the 38 haplotypes across the Blackface cohort identified HP:05 at the highest frequency, occurring on 23 occasions or 12% of observations (Figure 4). The combination of the four highest frequency haplotypes; HPs, 05, 10, 19 and 04 represent 35% of the total haplotypes identified. Not all haplotypes are distributed equally across the four Blackface flocks (Figure 4). Only five of the 38 haplotypes were identified in all four Blackface flocks (HP: 05, 04, 03, 25, 14). The number of haplotypes identified in the four flocks also showed considerable variation. Most diversity was identified in Flock A which included 21 MHC class I haplotypes, while flock B had the least with 14 haplotypes. Excluding the three Prealpe specific haplotypes, 17 haplotypes were found in only one of the four Blackface flocks (Figure 4).

3.5 | Haplotype diversity

A high degree of haplotype diversity in both the number of sequences within a haplotype and the level of transcription associated with each sequence was identified. The full range of haplotype diversity is detailed in Table 4, Figure 5 and Figure S5. Fifteen haplotypes with two high expressed transcripts (>25% of reads), were

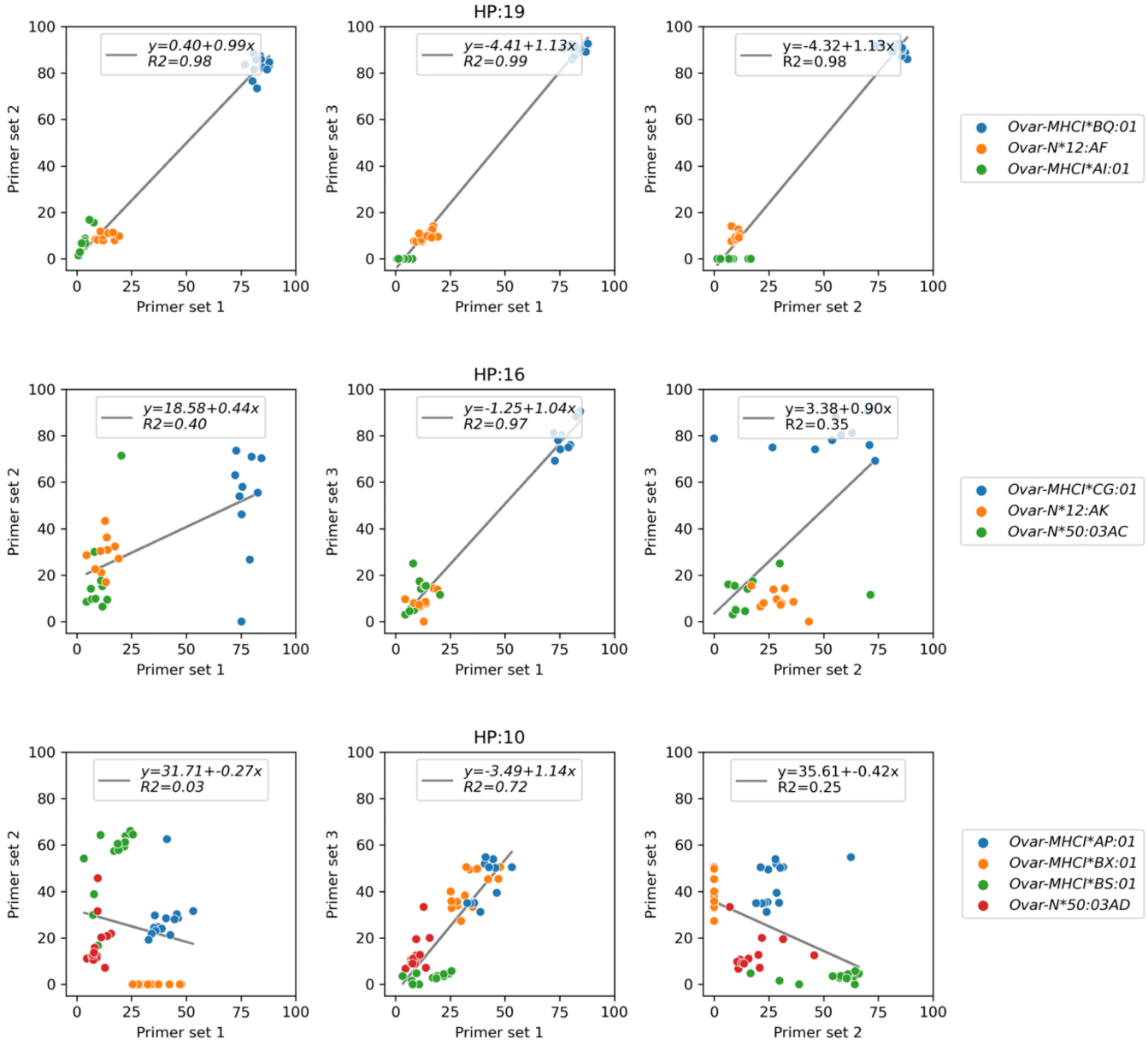

FIGURE 3 Example correlation analysis of expressed transcripts in individual haplotypes. The normalised relative frequency is plotted for each transcript in HP:19, HP:16 and HP:10 in pairwise comparisons between the three PCR primer sets. Different transcripts are represented in different colours. For each primer combination, the correlation coefficient and a line of best fit was calculated. HP:19 is an example of strong correlation between all three primers sets and HP:16 and HP:10 represent examples of PCR bias. The full set of haplotypes is shown in Data S1.

identified. Only a single haplotype (HP:15) was identified with more than two highly expressed transcripts. In haplotypes with a single high expressing transcript, it remains a possibility that a second transcript failed to amplify although, as we employed three sets of primers, we consider this unlikely. All but two of the 23 haplotypes with a single highly expressed transcript also include one or two medium expressed transcripts. We have no evidence to suggest that medium expressed class I genes are not also classical in function. Five haplotypes representative of those with between two and six transcripts are described in more detail below. Each of these haplotype combinations with associated expression levels for each transcript is presented in Figure 5.

- HP:11 was identified in 10 Blackface sheep from flocks A, B and C. Two MHC class I transcripts, *MHC1*CH:01* and *MHC1*AL:01* were associated with this haplotype in all 10 animals using all three primer sets. Both transcripts are transcribed at high levels

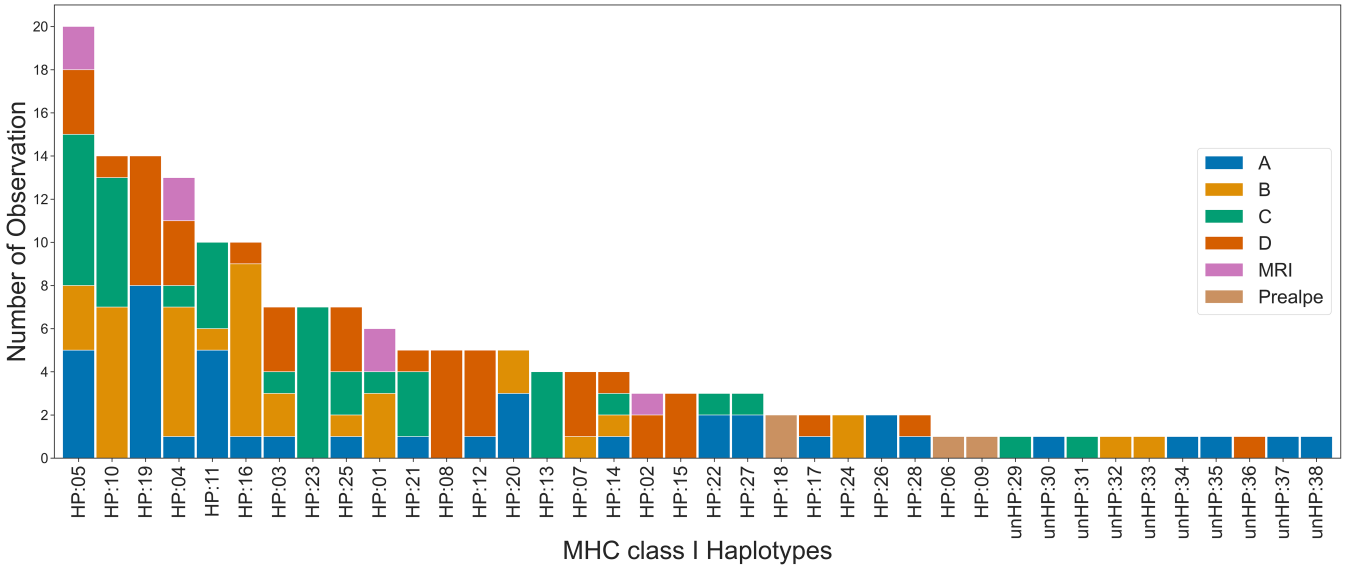

FIGURE 4 Histogram of MHC class I haplotypes observed in this study. Haplotypes are sorted in descending order of frequency along the horizontal axis. The distribution of haplotypes between flocks A-D is indicated by colour coding. MRI includes four haplotypes from three MHC homozygous animals and their two heterozygous sires from the Moredun Research Institute flock. Three haplotypes from the four Prealpe animals are included for comparative purposes. Unvalidated haplotypes observed on a single occasion are prefixed 'unHP'.

with *MHC1*AL:01* associated with 42.96% of reads while *MHC1*CH:01* is associated with 57.03% of reads. Unusually, no medium or low expressing transcripts were associated with HP:11 (Figure 5).

- HP:05 was identified in 23 animals including MHC homozygous animal BF284, its heterozygous sire, Ram 504, and Blackface sheep from all four flocks (Figure 4). Three sequences *N*05:01*, *N*06:01* and *N*50:00* were observed in each animal. *N*05:01* was the dominant transcript with a mean of 86.53% of sequence reads. The remaining two transcripts were expressed at medium and low levels, representing 10.46% and 3.89 of reads for *N*50:00* and *N*06:01*, respectively (Figure 5).
- HP:01 was identified in six Blackface sheep including MHC homozygous animal BF209, the heterozygous sire of this animal, Ram 504 and animals from flocks B and C (Figure 4). Four MHC class I transcripts were identified in all six animals, three of which had previously been identified, *N*01:01*, *N*08:01* and *N*50:01* (Figure 5). *N*01:01* was the dominant transcript accounting for 78.1% of sequence reads, *N*08:01* was medium in its level of transcription, accounting for 15.5% of reads while *N*50:01* was transcribed at low levels, accounting for 6.14% of reads. A fourth transcript, *MHCI*AY:01* was expressed at very low levels in HP:01, accounting for 0.45% of reads.
- HP:08 was identified in five animals within flock D (Figure 4). Five transcripts, *MHCI*BL:01*, *MHCI*AK:01*, *MHCI*AH:01*, *N*50:02:AA*, *N*12:AE* were identified in each of these five animals. *MHCI*BL:01* was the dominant transcript in this haplotype corresponding to 72.72% of all reads. *N*12:AE* was medium in its level of transcription corresponding to 19.74% of reads while *N*50:02:AA, MHCI*AK:01* and *MHCI*AH:01* were each transcribed at low or very low levels (Figure 5).
- HP:25 was identified in seven sheep including animals from all four flocks (Figure 4). Six transcripts, *MHCI*CF:01, MHCI*AN:01, MHCI*BE:01, MHCI*BH:01, MHCI*BK:01, N*50:03AB* were identified in all seven sheep (Figure 5). *MHCI*CF:01* and *MHCI*AN:01* are the two dominant transcripts in this haplotype corresponding to 47.04% and 41.96% of the total reads. The other four transcripts are all expressed at low or very low levels.

Of the 26 validated haplotypes, 17 included members of the *Ovar-N*50* allelic family, thought to represent non-classical MHC class I genes. The sequences in this family show considerable variation in expression level with a normalised average read frequency of 38.6 for *Ovar-N*50:03* compared with 3.84 for *Ovar-N*50:03AB* (Table 4). Except for the low expressing *N*50:03AF* sequence associated with five validated haplotypes from both Blackface and Prealpe breeds, the remaining N*50 family members are associated with only one or two haplotypes (Table 3). Similar diversity was identified in

TABLE 4 Description of 38 MHC class I haplotypes identified in this study.

Haplotype	Number of observations	Percentage of observations	Number of alleles	Allele1 *Ovar-*	Allele1: Normalised average read frequency	Allele2 *Oavr-*	Allele2: Normalised average read frequency	Allele3 *Ovar-*	Allele3: Normalised average read frequency	Allele4 *Ovar-*	Allele4: Normalised average read frequency	Allele5 *Ovar-*	Allele5: Normalised average read frequency	Allele6 *Ovar-*	Allele6: Normalised average read frequency	Allele7 *Ovar-*	Allele7: Normalised average read frequency
HP:01	7	3.76	4	*N*01:01*	78.07	*N*08:01*	15.53	*N*50:01*	6.14	*MHCI*AY:01*	0.45						
HP:02	3	1.61	3	*N*03:01*	81.04	*N*07:01*	13.3	*N*50:02*	5.65								
HP:03	7	3.76	3	*N*02:AA*	73.31	*N*12:AB*	17.22	*MHCI*BD:01*	9.47								
HP:04	15	8.06	3	*N*04:01*	59.56	*N*50:03*	38.6	*MHCI*AK:01*	5.54								
HP:05	23	12.37	3	*N*05:01*	86.53	*N*50:00*	10.46	*N*06:01*	3.89								
HP:06	2	1.08	5	*N*26:AA*	50.93	*MHCI*AQ:02*	26.21	*N*08:01:AA*	14.59	*N*12:AC*	14.03	*N*50:03AF*	2.98				
HP:07	4	2.15	5	*N*26:AA*	45.86	*MHCI*AQ:01*	27.04	*N*12:AC*	16.48	*N*08:01:AA*	13.82	*N*50:03AF*	5.81				
HP:08	5	2.69	5	*MHCI*BL:01*	72.72	*N*12:AE*	19.75	*N*50:02:AA*	8.3	*MHCI*AK:01*	5.1	*MHCI*AH:01*	0.45				
HP:09	2	1.08	4	*MHCI*AT:01*	59.41	*N*12:01*	17.32	*MHCI*AR:04*	12.72	*N*24:01*	10.56						
HP:10	14	7.53	4	*MHCI*AP:01*	37.35	*MHCI*BX:01*	37.09	*MHCI*BS:01*	24.76	*N*50:03 AD*	13.35						
HP:11	10	5.38	2	*MHCI*CH:01*	57.04	*MHCI*AL:01*	42.96										
HP:12	5	2.69	4	*MHCI*BI:01*	60.78	*MHCI*AR:02*	21.28	*N*12:AD*	17.68	*MHCI*AH:01*	0.39						
HP:13	5	2.69	4	*MHCI*BG:01*	82.86	*N*12:AM*	12.39	*N*50:03AB*	3.84	*MHCI*AK:01*	2.71						
HP:14	4	2.15	3	*N*02:01*	61.5	*N*12:01:AA*	27.24	*MHCI*CI:01*	11.27								
HP:15	3	1.61	4	*MHCI*BP:01*	53.18	*MHCI*BS:01*	28.61	*MHCI*AP:02*	25.69	*N*50:03 AD*	10.25						
HP:16	10	5.38	3	*MHCI*CG:01*	71.3	*N*12:AK*	17.21	*N*50:03 AC*	13.75								
HP:17	2	1.08	2	*MHCI*CA:01*	93.54	*N*50:03AF*	6.47										
HP:18	4	2.15	5	*N*27:AA*	67.41	*N*07:AA*	16.84	*N*12:AG*	13.15	*MHCI*BD:01*	2.37	*MHCI*AY:02*	0.66				
HP:19	14	7.53	3	*MHCI*BQ:01*	85.52	*N*12:AF*	11.26	*MHCI*AI:01*	4.83								
HP:20	5	2.69	3	*MHCI*BJ:01*	74.88	*N*08:01*	19.05	*N*50:01*	6.07								
HP:21	6	3.23	3	*MHCI*BN:01*	66.3	*MHCI*BM:01*	32.53	*MHCI*BY:01*	1.18								
HP:22	3	1.61	3	*MHCI*BO:01*	64.69	*MHCI*AR:03*	18.55	*N*12:01:AA*	16.76								
HP:23	7	3.76	3	*MHCI*BZ:01*	63.55	*MHCI*BF:01*	27.51	*N*12:AI*	8.95								
HP:24	2	1.08	3	*MHCI*BT:01*	75.8	*N*12:AL*	17	*N*50:03AF*	7.2								
HP:25	7	3.76	6	*MHCI*CF:01*	47.05	*MHCI*AN:01*	41.96	*N*50:03AB*	4.71	*MHCI*BE:01*	5.42	*MHCI*BH:01*	2.7	*MHCI*BK:01*	1.73		
HP:26	2	1.08	3	*MHCI*AO:01*	78.59	*N*12:AK*	16.01	*N*50:03AF*	5.4								
HP:27	3	1.61	3	*MHCI*AM:01*	54.12	*MHCI*BR:01*	39.7	*N*50:03AF*	6.17								
HP:28	2	1.08	3	*MHCI*BW:01*	66.38	*MHCI*BV:01*	36.07	*N*50:03AE*	9.58								
unHP:29	1	0.54	3	*MHCI*BB:01*	52.14	*N*12:01:AA*	32.05	*MHCI*CI:01*	15.82								
unHP:30	1	0.54	3	*MHCI*CK:01*	55.17	*MHCI*AS:01*	43.32	*MHCI*BH:01*	2.26								
unHP:31	1	0.54	2	*MHCI*CE:01*	96.23	*N*50:03AF*	3.77										
unHP:32	1	0.54	3	*MHCI*AV:02*	78.93	*N*12:AO*	15.16	*MHCI*AZ:01*	5.92								
unHP:33	1	0.54	2	*MHCI*BA:01*	83.31	*MHCI*BC:01*	16.69										
unHP:34	1	0.54	3	*MHCI*CC:01*	84.23	*MHCI*CJ:01*	13.16	*MHCI*AK:01*	7.83								
unHP:35	1	0.54	7	*MHCI*AU:01*	50.29	*MHCI*CD:01*	29.4	*N*12:AJ*	14.61	*N*50:03AF*	11.96	*MHCI*BQ:01*	1.81	*MHCI*BU:01*	1.51	*MHCI*AI:01*	0.35
unHP:36	1	0.54	2	*MHCI*AXE:01*	84.84	*N*50:02AB*	15.16										
unHP:37	1	0.54	3	*MHCI*CB:01*	69.95	*N*50:01AA*	16.12	*N*12:AK*	13.94								
unHP:38	1	0.54	3	*MHCI*AV:01*	79.4	*N*12:AK*	10.39	*N*50:01AB*	10.21								

Note: The table provides information on the constituent alleles within each haplotype, including the number of observations, their respective frequencies and the number of alleles. Additionally, the relative frequency of alleles within a haplotype, averaged across three PCR sets, is presented for each allele. The alleles are colour-coded according to their read frequencies: Pink indicates high (25%–97%), Blue indicates medium (10%–24%), Yellow indicates low (2%–9%), and Orange represents very low (0.1%–1.9%).

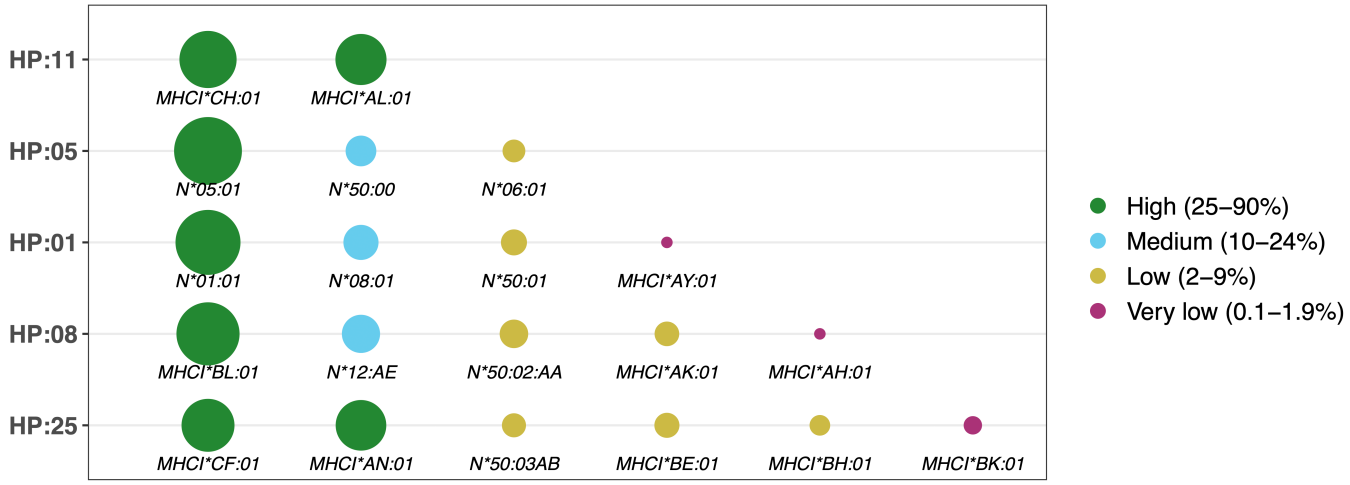

FIGURE 5 Selected haplotypes representing the range of haplotype diversity observed in this study. Each transcript identified in Haplotype HP:11, HP:05, HP:01, HP:08 and HP:25 is represented by a coloured circle and arranged according to expression level—green for high expressed, cyan for medium expressed, yellow for low expressed and pink for very low expressed. The size of each circle reflects its expression level, with transcript name displayed below the corresponding circle.

TABLE 5 Haplotype structure of unassigned genes in MHC homozygous animals.

Animal	Haplotype	*MHCI*AA*	*MHCI*AB*	*MHCI*AC*	*MHCI*AD*	*MHCI*AE*	*MHCI*AF*	*MHCI*AG*
BF121	HP:05	*AA:01:01*				*AE:02:01*		*AG:03:01*
BF34	HP:05	*AA:01:01*				*AE:02:01*		*AG:03:01*
BF284	HP:05	*AA:01:01*				*AE:02:01*		*AG:03:01*
BF414	HP:04	*AA:01:03*		*AC:06:01*	*AD:01:01*		*AF:02:01*	
BF66	HP:04	*AA:01:03*		*AC:06:01*	*AD:01:01*		*AF:02:01*	
PA3	HP:18	*AA:02:01*	*AB:06:01*	*AC:01:02*	*AD:03:01*	*AE:01:01*		
PA5	HP:18	*AA:02:01*	*AB:06:01*	*AC:01:02*	*AD:03:01*	*AE:01:01*		
BF126	HP:13	*AA:01:03*	*AB:04:01*	*AC:06:01*	*AD:01:01*	*AE:03:01*		
BF127	HP:21	*AA:01:01*				*AE:02:01*		
BF209	HP:01	*AA:01:01*	*AB:01:02*				*AF:01:01*	
PA1	HP:09	*AA:01:01*	*AB:04:01*	*AC:01:01*	*AD:02:01*	*AF:01:01*		

members of the *Ovar-N*12* allelic family present in 13 haplotypes with expression levels ranging from low (*Ovar-N*12:AI*) to high (*Ovar-N*12:01:AA*, Tables 3 and 4).

3.6 | Unassigned MHC class I sequences

A group of 31 sequences are not included in the haplotype structures described above as they are transcribed at very low levels and associated with multiple haplotypes which makes assignment to haplotypes difficult in heterozygous animals. These 31 sequences form seven allelic groups, *MHCI*AA* to *MHCI*AG* (Table 3) each of which includes a family of between 3 and 6 closely related sequences which include both non-synonymous and synonymous nucleotide substitutions. The nucleotide sequences of these transcripts are aligned in Data S6. Their distribution across the seven haplotypes represented in the 11 MHC homozygous animals indicates that between two and five are present on each haplotype with all seven haplotypes including an *Ovar-MHCI*AA* family member with variable numbers of the other six gene families (Table 5). Only one member of a family is found on each haplotype suggesting that each represents a different gene with haplotype variation in gene numbers. In homozygous animals that share haplotypes such as HP:05, HP:04 and HP:18, the identification of each

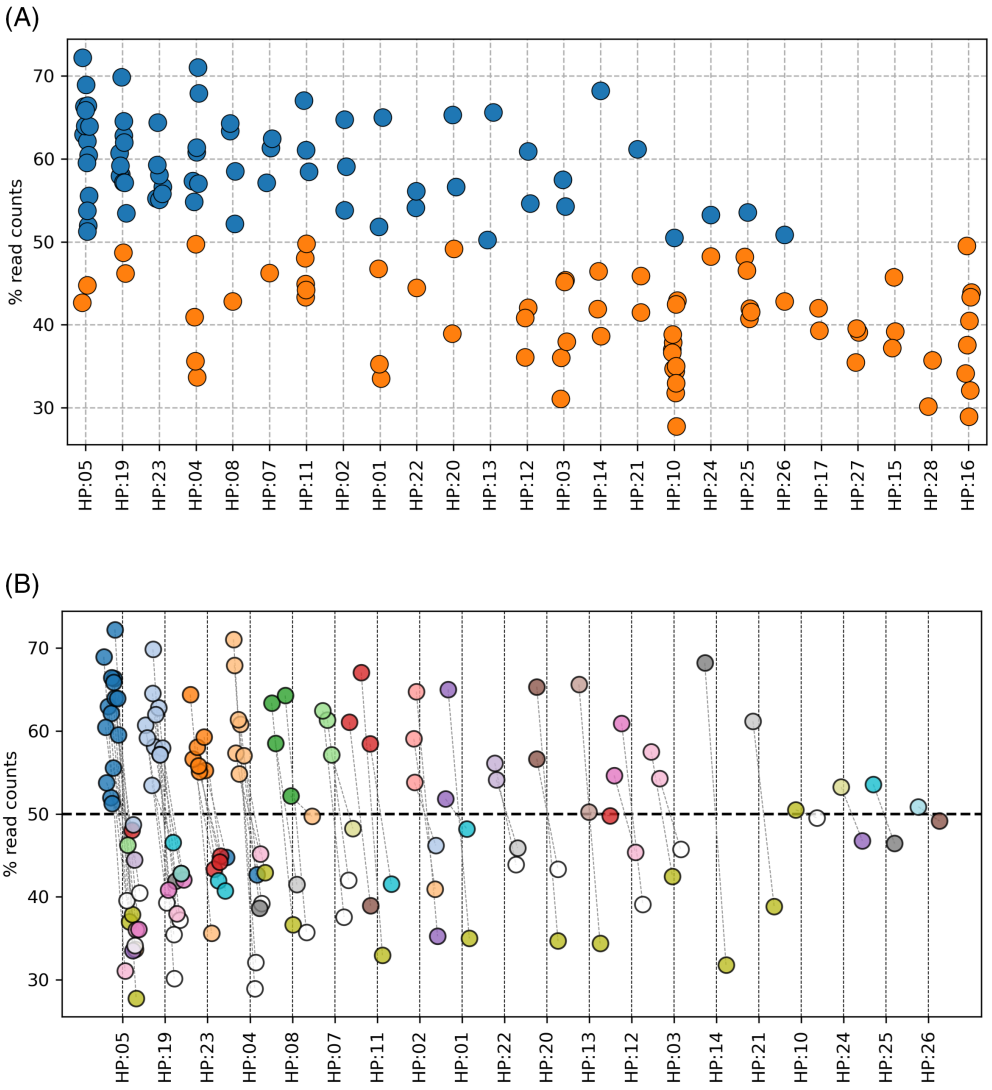

FIGURE 6 Comparison of read counts across all heterozygous animals. (A) Individual haplotypes with read counts >50% of the total highlighted in blue, while those <50% are shown in orange. (B) This plot illustrates the relationship between the two haplotypes in heterozygous animals. The connecting lines link haplotype combinations observed together in heterozygous animals. Data points above 50% read frequency (black dashed horizontal line) correspond to the haplotype labels on the *x*-axis, while the points below 50% represent their co-expressed haplotypes with colours matching those above 50%. The white points, represent the merger of the five haplotypes displayed on the right in plot A which do not occur above 50% read count.

sequence is consistent across animals. As shown in Table 5, each haplotype includes a unique combination of these sequences. A BLAST search with the first family member (*AA:01:01*, *AB:01:01*, *AC:01:01*) from each of the seven allelic groups against the sequences held within the IPD-MHC Database identified between 90% and 92% identify with the BoLA-NC (non-classical) gene *NC-9*. The exception was *Ovar-MHCI*AB:01* which showed the highest level of sequence identity (89%) with *BoLA-NC3*. Other sequences which are not members of these seven multiple allelic families, but cluster with them in the phylogenetic analysis described in section 3.10, are also expressed at low (*MHCI*AK:01*) and very low levels (*MHCI*AH:01*, *AJ:01* and *AY:01*). These sequences which are included in the haplotype structures as they are not spread across multiple haplotypes, also show high levels of sequence identity with *BoLA-NC9* (AH, 91% identity) and *BoLA-NC4* (AJ 85% identity, AK 87% identity and AY 93% identity).

3.7 | Haplotype dominance

To test for evidence of haplotype dominance we analysed read counts from all three primer sets associated with each of the two haplotypes in all heterozygous animals. Where no haplotype dominance exists, we expected the read counts for each set of sequences in a haplotype to be approximately equal. Figure 6A displays each haplotype plotted with corresponding read counts. Some haplotypes appear always below 50% of the total read counts (HP:17, 27, 15, 28 and 16), while others, such as HP:23, 02 and 13 are always above 50%. Others such as HP:05, 19, 04 and 08 are also largely dominant but with some exceptions. The relationship between dominant haplotypes was explored further by linking the read counts of both haplotypes in heterozygous animals. In Figure 6A, HP:05 appears largely dominant except for two animals where it is not the dominant haplotype. Figure 6B shows that in these two animals HP:05 occurs with HP:23 and HP:04. As

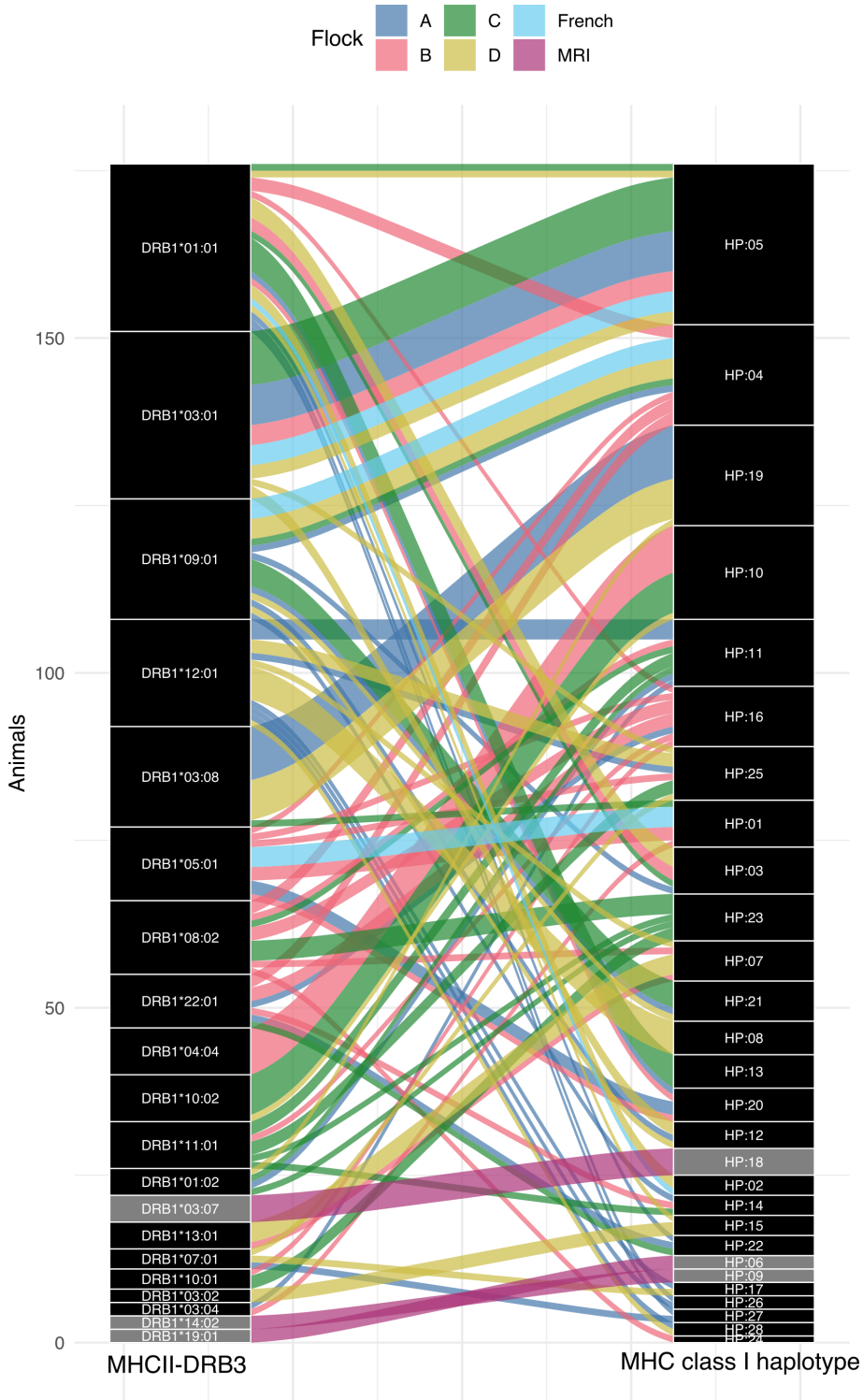

FIGURE 7 An alluvial diagram showing haplotype linkage between MHC class I and MHC class II regions. Only validated MHC class I haplotypes are included, and homozygous haplotypes are counted twice. Linkage was determined by, (i) pairwise comparison to calculate the normalised frequency of each DRB1 allele for each MHC class I haplotype, (ii) the highest calculated normalised frequency of each DRB1 allele and MHC class I haplotype, and (iii) the frequency at which each MHC class I-DRB1 combination occurred. Linkage is colour coded by flock origin.

both HP:23 and HP:04 are also dominant it suggests that HP:23 and HP:04 may have dominance over HP:05. Likewise, HP:19 occurs 14 times in this cohort of animals and in 12 of these it is the dominant haplotype. The two exceptions are when it occurs with HP:02 and HP:05 suggesting that HP:02 and HP:05 may be dominant over HP:19. HP:19 also occurs with HP:08 another largely dominant haplotype. In this case, H:19 remains dominant.

3.8 | Haplotype comparison

The three haplotypes identified in the homozygous Prealpe sheep (HP:06, HP:09 and HP:18) were compared with those from the Scottish Blackface breed. While the overall haplotype structure was the same, no Prealpe haplotype was shared with the Blackface animals. All transcripts in HP:09 are unique to this haplotype and the

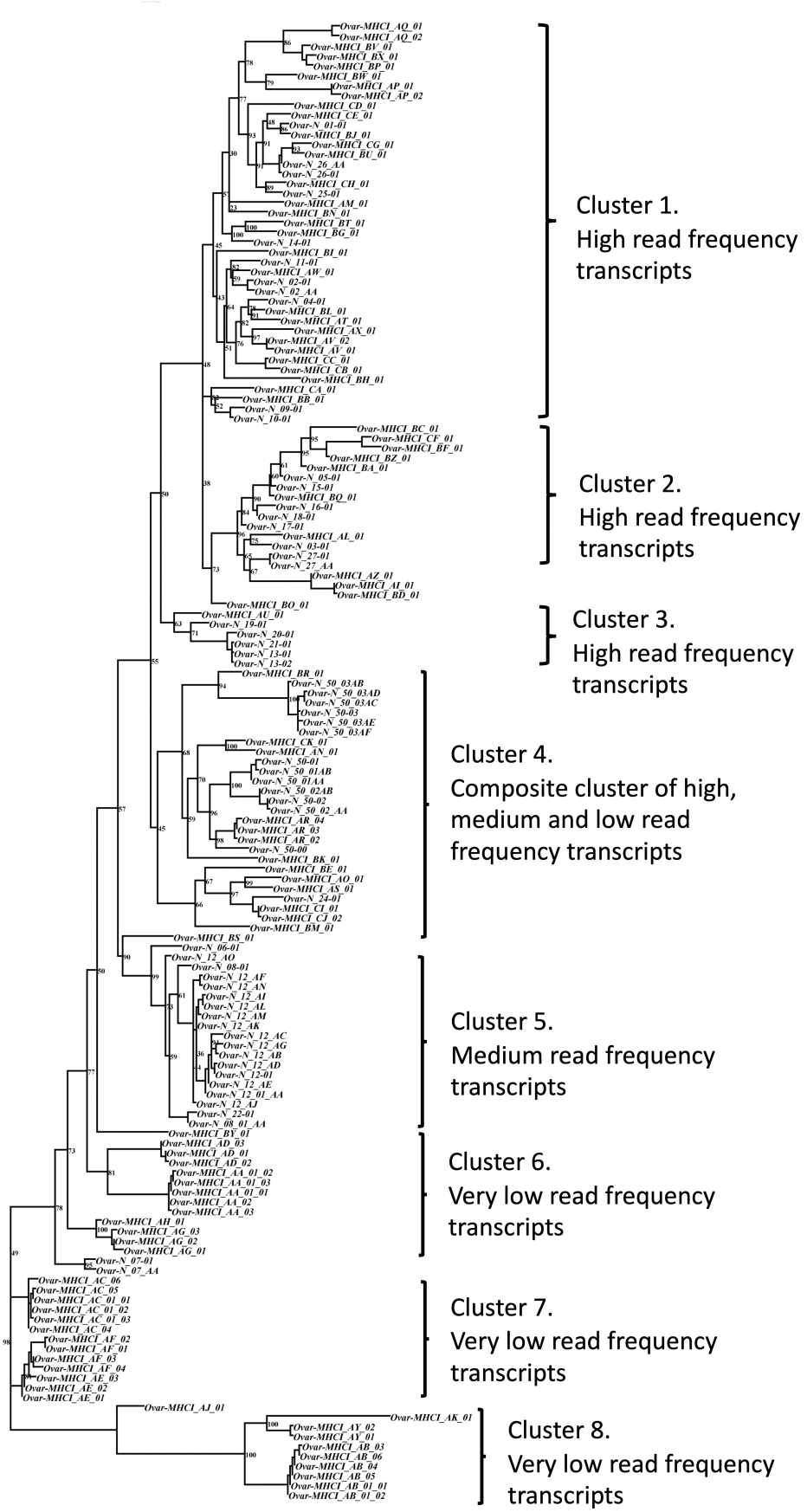

FIGURE 8 Midpoint rooted, maximum likelihood phylogenetic tree estimating the relationship between the 120 novel MHC class I sequences identified in this study aligned with 32 sequences downloaded from the sheep section of the IPD-MHC Database.

high expressed transcripts in HP:18 and HP:06 are unique to these haplotypes.

3.9 | Haplotype conservation

To determine the degree of extended haplotype conservation across the four Blackface sheep flocks, sheep were selected for MHC class I analysis based on shared MHC class II DRB1 allelic diversity. Linkage between MHC class II and MHC class I haplotypes is shown in Figure 7. This identified a wide range in extended MHC haplotype conservation. Four Blackface DRB1 alleles (*DRB1*03:02, 13:01, 04:04* and *10:02*) were associated with single class I haplotypes (HP:15, HP:07, HP:10 and HP:10, respectively). *DRB1*03:02* was identified twice in flocks D, *13:01* was identified on four occasions in both flocks D and B while *10:02* was identified seven times in flocks C and D. Of the more abundant DRB1 alleles, most animals with *DRB1*03:01* (88% animals) and *03:08* (93% animals) were associated with class I haplotypes, HP:05 and HP:19, respectively. In contrast, *DRB1*01:01, 09:01* and *12:01* were associated with 11, 5 and 7 different validated class I haplotypes, respectively.

3.10 | Phylogenetic analysis

Phylogenetic reconstruction allowed the relationship between the MHC class I sequences identified in this study along with additional sequences downloaded from the IPD-MHC Database, to be analysed in greater detail. The tree topology estimates that most high expressed transcripts locate to three clusters (Figure 8, clusters 1–3). In some haplotypes with two highly expressed transcripts (HP:11, 25 and 27), each locates to a different cluster while in others (HP:7, 28 and 15) both locate to cluster 1 indicating that each cluster does not represent allelic diversity at a single MHC class I locus. This is also true for cluster 4 which is a composite cluster composed of high, medium and low read frequency transcripts. Cluster 5 is composed of transcripts with mostly medium read frequency, many of which are closely related sequences within the *Ovar-N*12* allelic family. Clusters 6–8 are distinct clusters of closely related, mostly very low frequency transcripts within the *Ovar-MHCI*AA* to *AG* allelic families described in 3.6.

4 | DISCUSSION

Using three independent sets of primers, the highly variable second and third exons of sheep MHC class I genes were sequenced from 96 animals. Our use of cDNA template allows a focus on diversity in MHC class I genes that are likely to be translated into functional MHC class I proteins. Of the 134 distinct transcripts identified, 120 were novel with only 14 previously submitted to the sheep section of the IPD-MHC Database (https://www.ebi.ac.uk/ipd/mhc/) reflecting the absence of studies that have focused on sheep MHC class I diversity at scale. By using normalised read frequency data from each of the three sets of primers across multiple animals, we were able to identify strong correlations in read frequency between each of the three sets of primers. This allowed an estimation of expression levels for each transcript within each haplotype.

By analysing multiple animals with the same combinations of MHC class I sequences we were able to assign each to 38 haplotypes, 28 of which occurred in two or more animals and were considered validated. The 35 Scottish Blackface MHC class I haplotypes were linked with 18 MHC class II *DRB1* alleles indicating that MHC class I haplotype diversity exceeds MHC class II diversity and reflects similar observations in Holstein and Brazilian cattle breeds.[20] We had wished to target specific class I haplotypes for generation of panels of MHC tetramer reagents by sequence-based typing of the linked class II DRB1gene, however, while this may be possible for certain haplotypes with linkage to a single DRB1 allele it may not extend to other flocks or breeds. However, as the complement of class I allelic diversity associated with each haplotype appears unique, alternative haplotype specific screening targeting specific alleles in a target haplotype could be developed.

As this study represents the first large scale analysis of functional MHC class I haplotype diversity in sheep it extends the number of class I haplotypes with a well-defined set of class I genes from four to thirty-eight. To validate the methodology, three of the four homozygous animals that provided the first four haplotypes (BF414, BF284 and BF209) and their heterozygous sires (BF501 and BF504) were included here. These four haplotypes described by Miltiadou et al[14] and Ballingall et al[29] included a range of full-length and partial MHC class I transcripts identified by RT-PCR, amplicon cloning and Sanger sequencing of large numbers of individual clones. These transcripts were also identified in this study, however, the increased sensitivity provided by the NGS approach described here identified transcripts transcribed at very low levels which were not identified in the earlier study. The levels of transcript expression described in the earlier analysis[14] is also consistent with that defined here. Transcripts previously identified with high and medium levels of transcription express proteins at the cell surface,[29] so we can be confident that those identified

here will also be translated into proteins capable of presenting antigen to the T cell receptor and/or act as ligands for NK cell receptors.[30]

Validated haplotypes with between two and six transcripts were identified (Table 4). This mirrors the haplotype variation described in cattle where six classical class I genes have been identified but only between one and four are present in any given haplotype.[31,32] The number of transcripts and transcription levels were generally consistent across animals considering that RNA was prepared from blood samples collected at a single time point.

Each haplotype includes at least one dominant class I transcript. In the 20 animals with HP:05, the normalised average read frequency of *Ovar-N*05:01* was 86.5% while, *Ovar-MHC1*CA:01* represented 93.5% of the reads in the two animals with haplotype HP:29. Variation in transcription levels of MHC class I genes is not unusual, and has been described in many other species including human (HLA), murine (H-2), bovine (BoLA) and galline MHC regions.[20,33–35] Such variation reflects differences in function ranging from highly expressed classical class I genes with ubiquitous tissue expression to low expressing non-classical class I genes often with a more targeted tissue distribution and function.[8] The dominant class I transcripts described here in sheep are likely to represent alleles of classical MHC class I genes, with one or two identified in each haplotype. The phylogenetic clustering of high expressing transcripts provides an indication of allelic diversity however, assigning alleles to individual genes based purely on highly variable second and third exon sequence is not straightforward in ruminants.[36] The haplotype data provided by this study will allow selection of sheep with specific haplotypes for analysis of full-length transcripts to improve locus assignment and allow an official nomenclature to be assigned to each transcript deposited within the IPD-MHC Database.

In contrast to the haplotype variation observed for sheep and recently for cattle,[19–21] the HLA and chicken MHC regions have haplotypes with a fixed number of MHC class I genes. The HLA class I region includes three classical class I genes, HLA-A, B and C. HLA-A and B are the dominant class-I genes while HLA-C is transcribed at slightly lower levels.[37] While all three present peptide antigens to T cells, the principal function of HLA-C may be more related to NK-cell activation.[38] The chicken has only two MHC class I genes, one of which is highly expressed (BF2) and dominates antigen presentation to T-cells whereas the second very low expressed gene (BF1) may act as a ligand for NK cell receptors.[39] The wide distribution in expression levels described here in sheep may also reflect a similar distribution of functions.

Additional functional complexity associated with variation in expression levels for alleles at the dominant chicken class I gene BF2 has recently been reported. BF2 alleles expressing at higher levels are associated with the presentation of a more restricted range of peptides while lower expressing alleles present a broader range of peptides to T-cells. Similar correlations between expression level and peptide repertoire have also been described for the highly expressed classical HLA-A and HLA-B class I genes.[40] In contrast to the stable HLA and chicken haplotype structures, it is not known how the combination of haplotype and transcriptional variation described here in sheep influences the repertoire of peptides available to T-cells. The underpinning haplotype data presented here will allow us to address these questions in future studies.

A feature of the data presented here is the large number of very low ($n = 37$) expressed class I transcripts identified in multiple haplotypes. These include the 31 sequences within allelic groups, *MHCI*AA* to *MHCI*AG* which are found in haplotype specific combinations of between two and five in each of the seven haplotypes present in MHC homozygous animals. These sequences show between 85% and 93% identity with the proposed non-classical class I genes NC3, NC4 and NC9 described in the bovine MHC.[41,42]

As described for human and murine non-classical class I genes, the predicted bovine and ovine non-classical genes have distinct features when compared with the classical genes. These include truncated transmembrane regions with characteristic nucleotide and amino acid motifs, distinct 3-prime UTR regions, tissue specific expression and low levels of allelic polymorphism when compared with classical class I genes as described for the *Ovar-N*50* allelic family.[4,14,29] As the data presented here derives from transcripts covering exons 2–4 sequence, these diagnostic regions are missing from the *MHCI*AA* to *MHCI*AG* allelic families. However, all are likely to be derived from full-length transcripts as the cDNA was generated using an oligo-dT primer targeting the Poly-A tail of the RNA molecule and based on the available transcript data, all are predicted to translate into functional proteins (Data S3). Future analysis using tissue-specific transcript resources available for sheep[43] and the latest genome resources will improve our understanding of the evolution, structure and ultimately the function of these genes.

By including the very low expressing transcripts, the total number of class I genes in individual haplotypes will vary between 5 for HP:21 and 10 for HP:18. The higher range is consistent with the numbers described for cattle.[41] The functional consequence of such variation remains to be determined.

MHC alleles are co-dominantly expressed but it appears that certain haplotype combinations are consistently expressed at higher or lower levels than those in other haplotypes. In addition, a hierarchy in haplotype expression may exist in sheep. While a larger data set is required to confirm these observations at the haplotype level in sheep; hierarchy in allelic expression associated with differences in peptide repertoires and disease progression has been reported in chicken and human MHC class I genes.[40,44] Indeed, evidence of haplotype dominance in expression of classical HLA class I genes has recently been reported following infection of cell lines with SARS-Cov-2.[45]

In conclusion, this study has substantially increased our knowledge of MHC class I haplotype structure, diversity and gene expression in the Scottish Blackface breed. It will underpin future studies focusing on the functional consequence of haplotype variation, possible haplotype dominance and support the development of MHC tetramer reagents for screening T cell responses to vaccine antigens in sheep. It will allow more in-depth analyses of the tissue specific expression and function of the large numbers of low and very low expressing transcripts identified in variable numbers in each haplotype. As these data derive from a single northern European sheep breed it is likely to represent only a small fraction of the haplotype diversity that exits across sheep breeds worldwide. The small number of Prealpe haplotypes provides an indication of the diversity present in other breeds. A wide scale analysis of MHC class I haplotye diversity across multiple breeds is required to provide a fuller picture of allelic and haplotype diversity in sheep at the population level which will ultimately support comparative immunity, vaccine development and conservation efforts.

AUTHOR CONTRIBUTIONS

The study was designed by KB and TC. All bioinformatic analysis was carried out by DV. TZ and HT provided technical advice and support. KB and DV drafted the manuscript and all authors agreed on the final version.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of Scottish Blackface sheep farmers who offered animals for blood sampling. The cDNA from the French Prealpe lines was supplied by Frederic Lantier (INRA, France).

FUNDING INFORMATION

The study was funded through the European Union's Horizon 2020 VetBioNet project, 731014. Additional support was provided through the Innovation Fund of the Moredun Research Institute. For the purpose of open access, the author has applied a Creative Commons Attribution CC-BY licence to any Author Accepted Manuscript version arising from this submission.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw sequencing data has been submitted to ENA with the accession number PRJEB55926. Bioinformatics analysis pipeline and associated scripts are available on GitHub at this link: https://github.com/deepalivasoya/MHCtyping.

ETHICS STATEMENT

All animal samples were collected under UK Government Home Office licence following ethical review by the Moredun Animal Welfare and Ethical Review Body.

ORCID

Deepali Vasoya https://orcid.org/0000-0001-6764-5418
Keith T. Ballingall https://orcid.org/0000-0001-6920-8369

REFERENCES

1. Hall SJG. Livestock biodiversity as Interface between people, landscapes and nature. *People Nat*. 2019;1:284-290.
2. Klein J. *The Natural History of the Major Histocompatibility Complex*. John Wiley; 1986.
3. Gao J, Liu K, Liu H, et al. A complete DNA sequence map of the ovine major histocompatibility complex. *BMC Genomics*. 2010;11:466.
4. Subramaniam NS, Morgan EF, Wetherall JD, Steer MJ, Groth DM. A comprehensive mapping of the structure and gene organisation in the sheep MHC class I region. *BMC Genomics*. 2015;16:810.
5. Germain RN. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell*. 1994;76:287-299.
6. Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science*. 1996;272(5258):67-74. doi: 10.1126/science.272.5258.67
7. Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*. 1987;329:512-518.
8. D'Souza MP, Adams E, Altman JD, Birnbaum ME, Boggiano C, Casorati G. Casting a wider net: Immunosurveillance by non-classical MHC molecules. *PLoS Pathog*. 2019;15(2).
9. Lin A, Yan W-H. The emerging roles of human leukocyte antigen-F in immune modulation and viral infection. *Front Immunol*. 2019;10:964. doi:10.3389/fimmu.2019.00964
10. Ellis SA, Ballingall KT. Cattle MHC: evolution in action? *Immunol Rev*. 1999;167:159-168.
11. Ballingall KT, Bontrop RE, Ellis SA, et al. Comparative MHC nomenclature: report from the ISAG/IUIS-VIC committee. *Immunogenetics*. 2018;70:625-632.

12. Altman JD, Moss PA, Goulder PJ, et al. Phenotypic analysis of antigen-specific T lymphocytes. *Science*. 1996;274:94-96.
13. Sims S, Willberg C, Klenerman P. MHC–peptide tetramers for the analysis of antigen-specific T cells. *Expert Rev Vaccines*. 2010;9(7):765-774. doi:10.1586/erv.10.66
14. Miltiadou D, Ballingall KT, Ellis SA, Russell GC, McKeever DJ. Haplotype characterization of transcribed ovine major histocompatibility complex (MHC) class I genes. *Immunogenetics*. 2005;57(7):499-509. doi:10.1007/s00251-005-0008-y
15. Ballingall KT, Herrmann-Hoesing L, Robinson J, Marsh SG, Stear MJ. A single nomenclature and associated database for alleles at the major histocompatibility complex class II DRB1 locus of sheep. *Tissue Antigens*. 2011;77:546-553.
16. Ballingall KT, Dicks K, Kyriazopoulou P, Herrmann-Hoesing L. Allelic nomenclature for the duplicated MHC class II DQ genes in sheep. *Immunogenetics*. 2019;71:347-351.
17. Maccari G, Robinson J, Ballingall K, et al. IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Res*. 2017;45(D1):D860-D864. doi:10.1093/nar/gkw1050
18. Barker DJ, Maccari G, Georgiou X, et al. The IPD-IMGT/HLA database. *Nucleic Acids Res*. 2023;51(D1):D1053-D1060. doi:10.1093/nar/gkac1011
19. Vasoya D, Law A, Motta P, et al. Rapid identification of bovine MHCI haplotypes in genetically divergent cattle populations using next-generation sequencing. *Immunogenetics*. 2016;68(10):765-781. doi:10.1007/s00251-016-0945-7
20. Vasoya D, Oliveira PS, Muriel LA, et al. High throughput analysis of MHC-I and MHC-DR diversity of Brazilian cattle populations. *HLA*. 2021;98:93-113.
21. Silwamba I, Vasoya D, Simuunza M, et al. High throughput analysis of MHC class I and class II diversity of Zambian indigenous cattle populations. *HLA*. 2023;101(5):458-483. doi:10.1111/tan.14976
22. Ballingall KT, Lantier I, Todd H, Lantier F, Rocchi M. Structural and functional diversity arising from intra and inter haplotype combinations of duplicated DQA and B loci within the ovine MHC. *Immunogenetics*. 2018;70:257-269.
23. Ballingall KT, Tassi R. Sequence-based genotyping of the sheep MHC class II DRB1 locus. *Immunogenetics*. 2010;62:31-39.
24. Dicks KL, Pemberton JM, Ballingall KT. Characterisation of major histocompatibility complex class IIa haplotypes in an Island sheep population. *Immunogenetics*. 2019;71:383-393.
25. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011;27(21):2957-2963. doi:10.1093/bioinformatics/btr507
26. Kalyaanamoorthy S, Minh BQ, Wong TFK, Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587-589.
27. Nguyen L-T, Schmidt HA, Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol*. 2015;32:268-274.
28. Hoang DT, Chernomor O, Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35:518-522.
29. Ballingall KT, Miltiadou D, Chai Z-W, et al. Genetic and proteomic analysis of the MHC class I repertoire from four ovine haplotypes. *Immunogenetics*. 2008;60:177-184.
30. Braud VM, Allan DSJ, O'Callaghan CA, et al. HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature*. 1998;391(6669):795-799. doi:10.1038/35869
31. Holmes EC, Roberts AFC, Staines KA, Ellis SA. Evolution of major histocompatibility complex class I genes in Cetartiodactyls. *Immunogenetics*. 2003;55(4):193-202. doi:10.1007/s00251-003-0560-2
32. Ellis SA, Hammond JA. The functional significance of cattle major histocompatibility complex class I genetic diversity. *Annu Rev Anim Biosci*. 2014;2(1):285-306. doi:10.1146/annurev-animal-022513-114234
33. Wallny HJ, Avila D, Hunt LG, et al. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proc Natl Acad Sci U S A*. 2006;103(1434):1434-1439.
34. Ohtsuka M, Inoko H, Kulski JK, Yoshimura S. Major histocompatibility complex (Mhc) class Ib gene duplications, organization and expression patterns in mouse strain C57BL/6. *BMC Genomics*. 2008;9:178.
35. René C, Lozano C, Eliaou J-F. Expression of classical HLA class I molecules: regulation and clinical impacts. *HLA*. 2016;87:338-349.
36. Hammond JA, Marsh SGE, Robinson J, Davies CJ, Stear MJ, Ellis SA. Cattle MHC nomenclature: is it possible to assign sequences to discrete class I genes? *Immunogenetics*. 2012;64(6):475-480. doi:10.1007/s00251-012-0611-7
37. Apps R, Meng Z, Del Prete GQ, Lifson JD, Zhou M, Carrington M. Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. *J Immunol*. 2015;194(8):3594-3600. doi:10.4049/jimmunol.1403234
38. Aguilar AMO, Guethlein LA, Adams EJ, Abi-Rached L, Moesta AK, Parham P. Coevolution of killer cell Ig-like receptors with HLA-C to become the major variable regulators of human NK cells. *J Immunol*. 2010;185:4238-4251.
39. Kim T, Hunt HD, Parcells MS, van Santen V, Ewald SJ. Two class I genes of the chicken MHC have different functions: BF1 is recognized by NK cells while BF2 is recognized by CTLs. *Immunogenetics*. 2018;70(9):599-611. doi:10.1007/s00251-018-1066-2
40. Chappell PE, Meziane EK, Harrison M, et al. Expression Levels of MHC Class I Molecules Are Inversely Correlated with Promiscuity of Peptide Binding. 2015.
41. Birch J, Codner G, Guzman E, Ellis SA. Genomic location and characterisation of nonclassical MHC class I genes in cattle. *Immunogenetics*. 2008;60(5):267-273. doi:10.1007/s00251-008-0294-2
42. Schwartz JC, Maccari G, Heimeier D, Hammond JA. Highly-contiguous bovine genomes underpin accurate functional analyses and updated nomenclature of MHC class I. *HLA*. 2022;99:167-182.
43. Clark EL, Bush SJ, McCulloch MEB, et al. A high resolution atlas of gene expression in the domestic sheep. *PLoS Genet*. 2017;15(13):1006997.
44. Kaufman J. Generalists and specialists: a new view of how MHC class I molecules fight infectious pathogens. *Trends Immunol*. 2018;39(5):367-379. doi:10.1016/j.it.2018.01.001
45. Francisco Junior RDS, Temerozo JR, Ferreira CDS, et al. Differential haplotype expression in class I MHC genes during

SARS-CoV-2 infection of human lung cell lines. *Front Immunol*. 2023;13:1101526. doi:10.3389/fimmu.2022.1101526

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Vasoya D, Connelley T, Tzelos T, Todd H, Ballingall KT. Large scale transcriptional analysis of MHC class I haplotype diversity in sheep. *HLA*. 2024;103(2):e15356. doi:10.1111/tan.15356