Proceedings, 10th World Congress on Genetics Applied to Livestock Production

Association of *TMEM154* missense mutations with lentiviral infection and virus subtypes in sheep

M.P. Heaton, K.A. Leymaster, and M.L. Clawson
USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, Nebraska

ABSTRACT: Small ruminant lentivirus (SRLV) infections are a major cause of production losses in many sheep industries. Genetic susceptibility to SRLV infection in sheep is associated with the transmembrane protein 154 gene (*TMEM154*). A lysine mutation affecting the extracellular domain (K35, variant 1) is associated with significant reduction in infection rate. In production environments, the incidence of SRLV in ewes homozygous with variant 1 can be one eighth that of ewes with the variants encoding glutamate (E35, variants 2 or 3). The highly-susceptible, ancestral variant 3 allele shows complete dominance compared to variant 1. Twelve variants encoding different amino acid sequences have been identified in sheep, including frameshift deletions predicted to obliterate *TMEM154* function. Distinct SRLV genetic subgroups are associated with E35 and K35 variants, suggesting a direct interaction. Managing genetic variation in the *TMEM154* gene may help reduce, and then eradicate, SRLV in affected flocks.
Keywords: Sheep; SRLV; *TMEM154*

Introduction

Small ruminant lentiviruses (SRLV) are members of the *Retroviridae* family of enveloped viruses that infect sheep and goats causing chronic progressive diseases. SRLV infections in sheep cause an incurable, slow-acting, wasting disease that affects millions of sheep worldwide. Ovine progressive pneumonia (OPP) virus in North America and Visna/Maedi virus elsewhere are ovine SRLV strains that target the immune system causing persistent infections (Thormar (2005)). The disease affects multiple tissues, including those of the respiratory and central nervous systems. In North America, OPP is one of the most costly diseases affecting sheep due to decreased productivity, lameness, "hard bag", and early culling of ewes. In sheep with clinical OPP, interstitial pneumonia is readily apparent at necropsy. In research flocks, infected ewes were 20% less productive than uninfected ewes (Keen et al. (1997)). It has been estimated that 36% of sheep operations and 24% of all animals tested in the U.S. were infected with SRLV (USDA, APHIS (2003). Once infected with the virus, sheep are carriers throughout life, as there is no effective treatment or vaccine.

Recently, we reported that amino acid sequence variation encoded by ovine transmembrane protein 154 gene (*TMEM154*) was associated with susceptibility to SRLV infection (Heaton et al. (2012)). Additional reports have documented the distribution of variants encoded by *TMEM154*, the mode of inheritance for common variants, the efficacy of vertical transmission, and the association of SRLV genetic subtypes with *TMEM154* variants (Heaton et al. (2013); Leymaster et al. (2013); (Sider et al. (2013)). The present article summarizes our current understanding of amino acid variation encoded by *TMEM154* variants and their influence on the risk of infection when exposed to SRLV genetic subtypes endemic at USMARC.

Results and Discussion

A host genetic approach to understanding variation in susceptibility to SRLV infection in sheep. Prevalence studies in Idaho and Nebraska showed that breed type was a risk factor for SRLV infection in sheep (Gates et al. (1978); Keen et al. (1997)). In the 1990's at USMARC, the SRLV prevalence among comingled Finnsheep and Suffolk breeds was 80% and 15%, respectively. A 2003 prevalence survey at USMARC with more than 3,500 sheep showed similar results (Heaton et al. (2012)). Some breeds had a high prevalence in relatively young flocks, while other breeds had low prevalence in older flocks (Figure 1). Thus, in the USMARC production environment, there were apparent genetic differences in susceptibility to SRLV. The combination of diverse germplasm, persistent exposure to an endemic pathogen, and intermediate frequencies of infected sheep within a breed, indicated that this disease might be amenable to a genome-wide association study (GWAS).

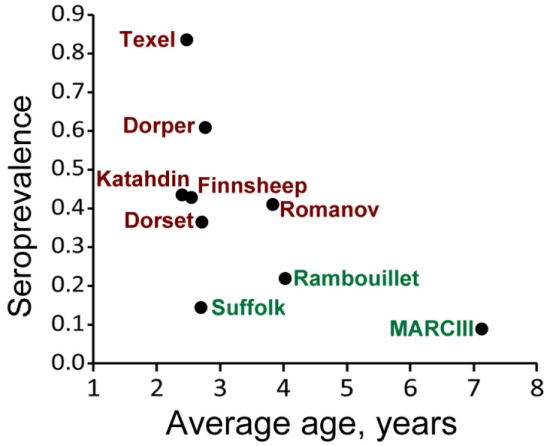

Figure 1: SRLV seroprevalence in USMARC sheep.

A study was designed to detect genetic variation influencing susceptibility to SRLV infection. Genotype tests were applied to matched pairs of ewes that had received a lifetime of natural SRLV exposure at USMARC. Each pair contained one infected ewe and an uninfected ewe of the same age, breed, and flock. Stringent matches were made using extensive production records. Another key feature of the design was the identification of two sets of pairs sampled: one for "discovery" (69 pairs of 5- to 9-years-

old), and another for "validation" (61 matched pairs of 4-year-old ewes (Figure 2).

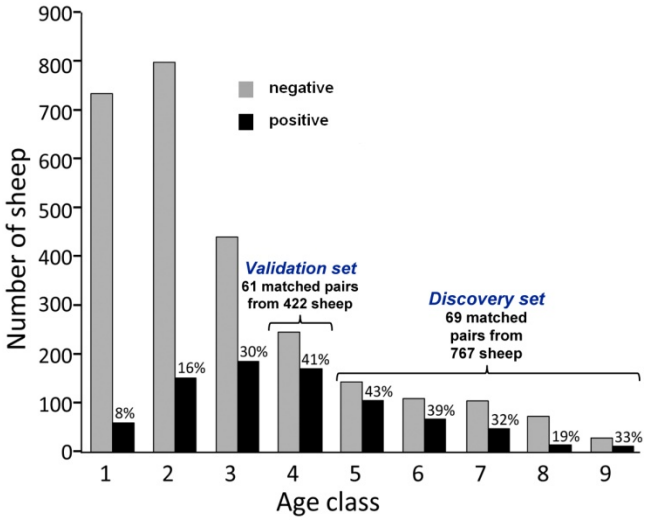

Figure 2: Matched case-control design for susceptibility to SRLV infection.

Approximately 50,000 single nucleotide polymorphisms (SNPs) in the Ovine SNP50 BeadChip array were scored in the discovery set of case-control pairs and tested for association with SRLV infection. Although the McNemar's test for correlated proportions was appropriate for paired samples, software that could analyze 50k data sets with McNemar's test was not available. However, the less sensitive chi-squared test identified a single SNP on ovine chromosome 17 that was highly significant and no evidence of an inflated test statistic was detected (Figure 3). The unadjusted *p*-value for the association was 3.19×10^{-9} and compared favorably to the significance threshold of 1×10^{-6}. The c/t SNP (OAR17_5388531) was in intron 5 of an ovine gene homologous to the human *TMEM154* gene on human chromosome 4, and the "c" allele was associated with infection.

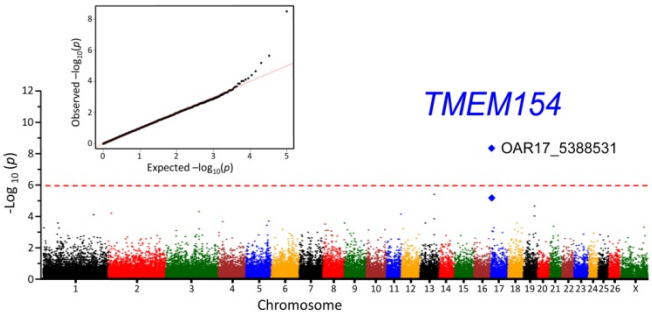

Figure 3: GWAS results using the discovery set.

Having identified a region of the genome associated with SRLV infection, the McNemar's test was applied to Ovine SNP50 BeadChip array SNPs in that region. The most significant association from the McNemar's test was the same SNP identified by the chi-squared test (SNP OAR17_5388531). The dichotomous variable for this McNemar's test was defined as having zero or one copy of the *TMEM154* genetic risk factor, i.e., a "c" nucleotide allele at SNP OAR17_5388531. In the discovery set, the McNemar's test showed that the odds of being infected were 18-fold higher in ewes with one copy of this "c" allele. This was confirmed in the validation set (Table 1). Genotype analysis of Sanger sequences from a 78kb region of chromosome 17 in the discovery set demonstrated that *TMEM154*, and not flanking genes, was the likely source of the association.

Table 1: McNemar's test for correlated proportions with SNP OAR17_5388531.

Discordant pair[a] (risk allele)	Discovery set	Validation set
Case (1) Control (0)	36 pairs	30 pairs
Case (0) Control (1)	2 pairs	2 pairs
Odds ratio	10	15
p-value	<0.0001	<0.0001
CI_{95}	5 - 150	4 - 130

[a]Only discordant pairs are informative in McNemar's test. A number 1 in brackets indicates the presence of one copy of a risk allele.

The function of *TMEM154* has not been reported in any species and remains largely unknown. It has the structure of a Type 1 membrane protein: a signal peptide sequence, one membrane spanning domain, an N-terminus on the extracellular side of the membrane, and the C-terminus on the cytoplasmic side (Figure 4). Mammals, birds, and fish appear to have one copy of a *TMEM154* gene. The bovine, human, and murine proteins are 93%, 67%, and 54% identical with the mature ovine protein, respectively. In humans, *TMEM154* mRNA is most abundant in CD19+ B cells and CD14+ monocytes. Cells of the monocyte lineage are the target cells for SRLV infection in sheep. Thus, it is biologically plausible that the extracellular domain of the TMEM154 protein interacts with SRLV during infection.

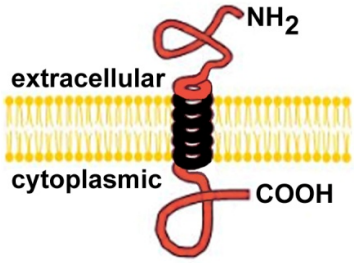

Figure 4: Typical secondary structure and orientation of a Type 1 membrane protein.

The identification of *TMEM154* as a major gene influencing SRLV infection prompted a search for causative polymorphisms within the gene. Sequencing *TMEM154* transcripts from cases and controls showed that mRNA splice variants were not common or associated with SRLV infection. However, genomic DNA sequencing from more than 300 sheep and 40 breeds from around the world (Heaton et al. (2013)) revealed 10 missense and two frameshift mutations (Figure 5A). Haplotype phase for these mutations was determined by analyzing their segregation in families and by scoring individuals with zero or one heterozygous sites. The phylogenetic relationship of distinct protein variants encoded by *TMEM154* is illustrated in a rooted median-joining network (Figure 5B). The ancestral root for

this network (variant 3) was established by comparing *TMEM154* sequences from more than a dozen ruminant species. Variant 3 has amino acid residues R4, A13, L14, T25, E31, D33, E35, T44, N70, I74, E82, and I102 at the sites encoding missense or frameshift mutations. Each node in the network represents a single mutation that affects the protein sequence. In more than 8,000 sheep tested, 97% had some combination of variants 1, 2, and 3 shown in Figure 5B.

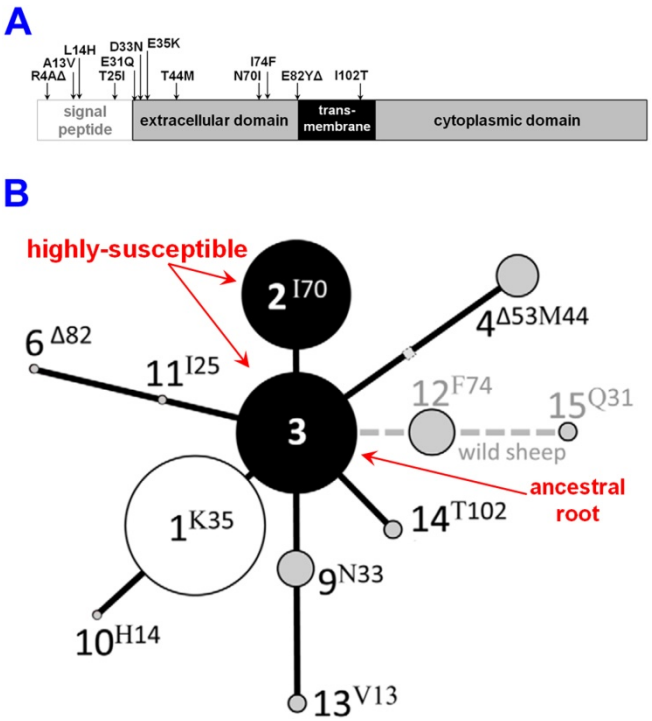

Figure 5: *TMEM154* mutations and median-joining network of protein variants.

The median-joining network provided a framework for evaluating the influence of *TMEM154*-encoded polypeptide variants on SRLV infection. The SNP encoding the E35 allele was in strong linkage disequilibrium with the "c" nucleotide allele at SNP OAR17_5388531 (r^2 = 0.98). However, E35 was present on 10 of the 12 predicted protein variants. Thus, McNemar's test was used to evaluate the common *TMEM154* variants as risk factors for infection in the matched pairs. The most informative result was with variants 2 and 3. In the discovery set, these risk factors were present in every case of the discordant pairs (Table 2). The odds of infection were not calculated for the discovery set as there were no discordant pairs where the control had the risk factor. In the validation set, McNemar's test showed the odds of being infected were 28-fold higher in older ewes with one copy of variant 2 or 3.

In less controlled, retrospective cohort studies with sheep in Idaho, Montana, Nebraska, and Iowa, the infection rate for ewes with one copy of *TMEM154* variant 2 or 3 was approximately three times that of sheep without these genetic risk factors. The cohort studies had a combination of uncontrolled variables including: age, breed, environmental conditions, management type, *TMEM154* allele frequencies, SRLV strains, and pathogen exposure.

Table 2: McNemar's test repeated with *TMEM154* variants 2 or 3 (E35).

Discordant pair[a] (risk allele)	Discovery set	Validation set
Case (1) Control (0)	41 pairs	28 pairs
Case (0) Control (1)	0 pairs	1 pairs
Odds ratio	undefined	28
p-value	<0.0001	<0.0001
CI_{95}	undefined	5 - 1100

[a]Only discordant pairs are informative in McNemar's test. A number 1 in brackets indicates the presence of one copy of a risk allele.

The effects of *TMEM154* variants 1 and 3 on SRLV infection in lambs. A prospective cohort study was designed to estimate additive and dominance effects of *TMEM154* variants 1 and 3 on susceptibility to SRLV infection (Leymaster et al. (2013)). The study design focused on SRLV transmission in the first eight weeks of life (i.e., the pre-weaning period) in a drylot production setting. A key feature of the design was the exclusive use of mature, infected ewes to maximize the trial lambs' natural exposure to SRLV. Parents of 187 trial lambs were heterozygous for *TMEM154* variants 1 and 3, and produced lambs with all three haplotype combinations (i.e., diplotypes). A group of 20 sentinel lambs contained individuals that were either homozygous for variant 1 or heterozygous with variants 1 and 3. The sentinel lambs were raised by mature, uninfected ewes but comingled with the trial lambs and their infected ewes during the experiment. Only one sentinel lamb became infected, indicating that little horizontal transmission occurred in spite of close contact with many infected ewes.

Lambs were isolated from their dams at weaning, and monitored to nine months of age. During this time, the lamb's passively-acquired maternal SRLV antibody titers decayed, and infected lambs developed new SRLV antibody titers. At nine months of age, the probability of infection for lambs with diplotypes 1, 3 or 3, 3 averaged 3.3 times that of lambs with diplotype 1, 1 (Table 3). Thus, the lamb's infection status was affected by its *TMEM154* diplotype (*p*-value <0.005) and was consistent with complete dominance of haplotype 3 relative to haplotype 1.

The greatest risk for SRLV transmission during the pre-weaning period was from mature, infected ewes raising lambs that were genetically most-susceptible to infection (1, 3 or 3, 3). Notably, only about 35% of such lambs were infected due to vertical transmission, implying that 65% of the genetically-susceptible lambs were not infected at nine months of age despite constant exposure to their infected dams (Table 3).

Table 3: SRLV infection rates at 9 and 35 months of age by *TMEM154* diplotype.

Age	*TMEM154* diplotype		
	1, 1	1, 3	3, 3
9 months	10.7	33.8	36.7
35 months	10.7	81.8	89.3

The ewe lambs from this trial were combined with an infected flock of mature ewes in a natural production environment. After two lambing cycles and 35 months of age, the probability of infection for ewes with either diplotype 1, 3 or 3, 3 averaged eight times that of ewes with diplotype 1, 1 (Table 3). This demonstrated that genetic susceptibility to SRLV infection can be reduced by selection to increase the frequency of haplotype 1, resulting in a greater proportion of lambs with diplotype 1, 1.

The increase in SRLV prevalence at 35 months of age documented the impact of horizontal transmission. The major cause of lifetime infection is likely due to horizontal transmission that occurs after uninfected ewes join a breeding flock of infected ewes. Therefore, a key management strategy to control SRLV infection is isolation of young ewes to prevent subsequent horizontal exposure to SRLV.

The influence of SRLV genetic subgroups on susceptibility to infection. Dissecting the genetics of host-pathogen interactions requires knowledge of both players. This is particularly relevant for retroviruses like SRLV which evolve at an accelerated rate compared to other viruses. SRLVs are genetically diverse with subtypes (i.e., "strains") that appear to be geographically stratified throughout many locations of the world, including the United States. Consequently, strains of SRLVs that differ in their ability to infect sheep may have evolved in some locations. To address this possibility, SRLV strains were characterized in naturally infected sheep and tested for associations with *TMEM154* variants 1, 2, and 3 (Sider et al. (2013)). SRLV strains were characterized by sequence analyses of two gene regions (*gag* and *env*), which reside on opposite sides of the SRLV genome (Figure 6).

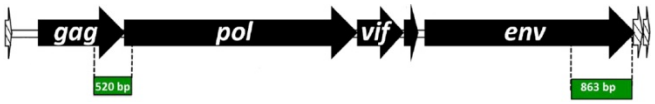

Figure 6: Physical map of the SRLV genome showing regions used for genetic typing (green).

Two predominant genetic subgroups of SRLVs were found in USMARC sheep (subgroups 1 and 2, Figure 7). Both subgroups were distinct from SRLVs that infected sheep in other regions of North America, or elsewhere in the world. Importantly, subgroup 2 associated with sheep having *TMEM154* variants 2 or 3, and subgroup 1 associated with sheep having only variant 1 (Figure 7). This indicated that SRLVs in the U.S. have adapted to infect sheep with specific *TMEM154* variants, and genotypes from both the host and pathogen affect the relative risk of infection. Thus, efforts to reduce the prevalence of SRLV by increasing the frequency of *TMEM154* variant 1 may be affected by virus strains that have adapted to the host genotype.

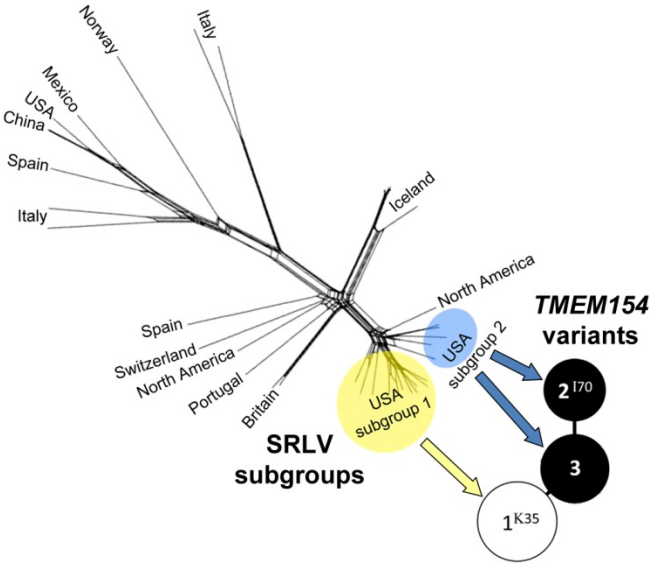

Figure 7: Neighbor-joining network of SRLV *env* subgroups and their association with *TMEM154* variants 1, 2, and 3.

The apparent coevolution of SRLVs and ovine *TMEM154* variants raises questions about the possible function of *TMEM154* and its origin. For example, the acidic E35 residue encoded by *TMEM154* is strictly conserved from ruminants through marsupials. Sheep are the only known species with the basic K35 residue. Why is K35 (variant 1) distributed in breeds worldwide, including those near the center of domestication in present day Iran and Turkey (Figure 8)? Could the wide distribution of variant 1 be the result of selective pressure from SRLV infections? Also, the amino acid variation encoded by *TMEM154* appears to be localized to its extracellular domain (Figure 5A). Is this phenomenon the result of diversifying selection of a co-receptor? Moreover, two distinct frameshift mutations are predicted to obliterate *TMEM154* function and have risen to a frequency that occasionally results in homozygous individuals. Could the frameshift mutations also be the result of selective pressure?

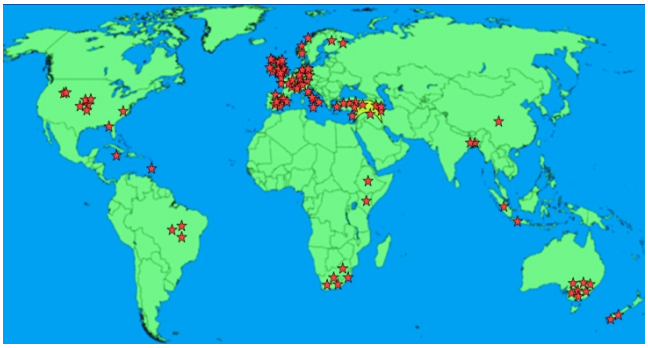

Figure 8: Geographic origin of sheep with *TMEM154* variant 1 (K35).

Is complete genetic resistance to SRLVs possible? Complete genetic resistance to lentivirus infection has been documented with human immunodeficiency virus 1 (HIV-1) and the human cytokine receptor *CCR5*. *CCR5* is a

coreceptor for HIV-1 and naturally occurring protein variants of CCR5 influence susceptibility to HIV-1 infection (Niaf (2013)). People born with two copies of the defective variant, *CCR5Δ32*, are resistant to HIV-1 infection. Also, replacing an infected person's stem cells with those from a homozygous *CCR5Δ32* donor has resulted in a cure for HIV-1. If *TMEM154* in sheep is analogous to CCR5 in humans, it may explain the selection for *TMEM154* frameshift variants. The median-joining network in Figure 5B predicts that the direction of evolution is from the center outward. Thus, each distal node is younger that it's proximal ancestor. This suggests that *TMEM154* deletion variants arose on separate haplotype lineages. Variant 4 has been found in breeds originating in Europe and the Middle East. Variant 6 has only been reported in Suffolk (Figure 9).

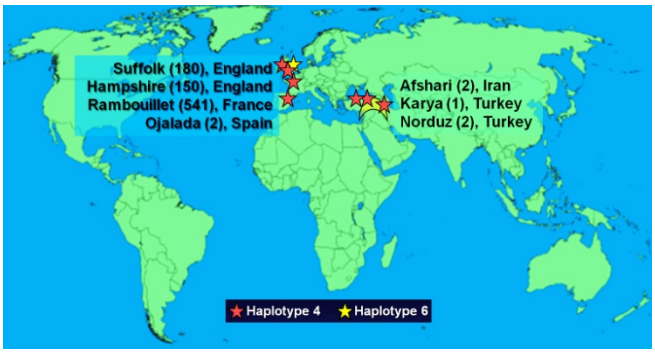

Figure 9: Geographic origin of sheep with *TMEM154* frameshift deletion variants 4 and 6 (number of animals genotyped).

The first homozygous genotype for *TMEM154* deletion variant 4 was discovered retrospectively in samples previously collected at USMARC. The 10-year-old Suffolk ewe was in good health, and still productive. The ewe tested negative for SRLV infection at 3 and 10 years of age, lived to 11 years, and produced five sets of twins in the last five years of her life. This demonstrated that variant 4 was not lethal in the homozygous form. Subsequently, five sheep with homozygous *TMEM154* deletions were identified retrospectively. They had an average lifespan of 6.7 years and all were uninfected at their last testing despite significant SRLV exposure. The first *TMEM154* "knockout" lambs (4, 4) were purposely produced at USMARC in 2011 and have appeared normal throughout their growth and development. Because information on *TMEM154* knockout sheep is limited, additional research is needed to determine the effects of variants 4 and 6 before recommendations can be made on using these haplotypes to lower SRLV infection.

Accordingly, a natural challenge cohort study was designed to compare the long-term SRLV susceptibility of ewes with combinations of *TMEM154* variants 1 and 4. The study is currently underway. Exposure to SRLV is being controlled by introducing ewe lambs to a highly infected ewe flock at 7 months of age. The ewe lambs will remain in the infected flock and are being tested for infection three times a year through 4 years of age. Results from this study will help provide selection guidelines to producers for using these variants.

Opportunity for *TMEM154* genetic testing. A genotype test for *TMEM154* has been commercially available from at least one laboratory since May 2012. The aim is to determine *TMEM154* variants for approximately $10 to $12 US/test. However, for sheep producers to manage *TMEM154* variants they first need to know if their flock is infected, and to what extent. Blood tests for anti-SRLV antibodies are available from regional diagnostic laboratories for approximately $5-$10 US per animal.

In areas where SRLV is prevalent, sampling the oldest ewes provides a sensitive measure of seroprevalence within a flock. For infected flocks, a combination of both serological and genetic testing, as well as management strategies, may be appropriate to rapidly decrease the prevalence of infection. Protocols will likely need to be customized to account for conditions existing in individual flocks. Some producers have implemented *TMEM154* testing procedures to reduce or eradicate SRLV. In flocks that are SRLV-negative, producers may select for *TMEM154* variant 1 to increase the potential revenue from sale of seed stock or to provide genetic protection against accidental exposure to SRLV. Additional options include retaining any sheep with deletion variants 4 and 6. Also, variant 10 (K35) has been reported in Rambouillet sheep and may offer an advantages similar to variant 1.

A caveat for using *TMEM154* variation to reduce SRLV prevalence. It is important to consider that some sheep homozygous for variant 1 may still become infected with sufficient exposure. Adverse production conditions like high animal density, indoor housing with poor ventilation, moist climates, and the presence of certain SRLV strains, may enhance transmission and overcome genetic resistance provided by some *TMEM154* variants. Furthermore, the effects of variants 4, 6, and 9 through 15 are unknown. However, as strategies for *TMEM154* genetic testing are evaluated under field conditions, additional genetic guidelines for reducing the incidence of SRLV infection will emerge. Ultimately, information and products of the research will be used to select for animals less likely to be infected by SRLV.

Acknowledgements. We thank J. Carnahan and G. Schuller-Chavez for technical assistance; Mike Wallace, Erwin Heiden, Lynette Anderson and the USMARC sheep crew for sheep production, sample collection and management of sheep; J. Watts for secretarial assistance; coauthors from the four articles reviewed in this proceedings manuscript, the members of ISGC for producing and making publicly available sheep genomic sequences and the OvineSNP50 BeadChip. Mention of trade name, proprietary product, or specified equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that may be suitable. USDA is an equal opportunity provider and employer.

Literature Cited

Gates, N. L., Winward, L. D., Gorham J. R., et al. (1978) J. Am. Vet. Med. Assoc., 173:1575.

Heaton, M. P., Clawson, M. L., Chitko-Mckown, C. G. et al. (2012) PLoS Genet., 8: e1002467.

Heaton, M. P., Kalbfleisch, T. S., Petrik, D. T., et al. (2013) PLoS ONE, 8: e55490.
Keen, J. E., Hungerford, L. L., Wittum, T. E., et al. (1997) Prev. Vet. Med., 30:81.
Leymaster, K. A., Chitko-McKown C .G., Clawson M. L., et al. (2013) J. Anim. Sci., 91:5114.
Naif, H. M. (2013) Infect Dis Rep. 5:e6.
Sider, L. H., Heaton, M. P., Chitko-McKown, C. G. et al. (2013) Vet. Res., 44:64.
Thormar H. (2005) AIDS Rev 7:233
USDA, APHIS (2003) www.aphis.usda.gov/animal_health/nahms/sheep/downloads/sheep01/Sheep01_is_OPP.pdf .