Genome-Wide Association Identifies Multiple Genomic Regions Associated with Susceptibility to and Control of Ovine Lentivirus

Stephen N. White1,2*, Michelle R. Mousel3, Lynn M. Herrmann-Hoesing2, James O. Reynolds1, Kreg A. Leymaster4, Holly L. Neibergs5, Gregory S. Lewis3, Donald P. Knowles1,2

Abstract

Background: Like human immunodeficiency virus (HIV), ovine lentivirus (OvLV) is macrophage-tropic and causes lifelong infection. OvLV infects one quarter of U.S. sheep and induces pneumonia and body condition wasting. There is no vaccine to prevent OvLV infection and no cost-effective treatment for infected animals. However, breed differences in prevalence and proviral concentration have indicated a genetic basis for susceptibility to OvLV. A recent study identified TMEM154 variants in OvLV susceptibility. The objective here was to identify additional loci associated with odds and/or control of OvLV infection.

Methodology/Principal Findings: This genome-wide association study (GWAS) included 964 sheep from Rambouillet, Polypay, and Columbia breeds with serological status and proviral concentration phenotypes. Analytic models accounted for breed and age, as well as genotype. This approach identified TMEM154 (nominal P = 9.2 × 10⁻⁷; empirical P = 0.13), provided 12 additional genomic regions associated with odds of infection, and provided 13 regions associated with control of infection (all nominal P < 1 × 10⁻⁵). Rapid decline of linkage disequilibrium with distance suggested many regions included few genes each. Genes in regions associated with odds of infection included DPPA2/DPPA4 (empirical P = 0.006), and SYTL3 (P = 0.051). Genes in regions associated with control of infection included a zinc finger cluster (ZNF192, ZSCAN16, ZNF389, and ZNF165; P = 0.001), C19orf42/TMEM38A (P = 0.047), and DLGAP1 (P = 0.092).

Conclusions/Significance: These associations provide targets for mutation discovery in sheep susceptibility to OvLV. Aside from TMEM154, these genes have not been associated previously with lentiviral infection in any species, to our knowledge. Further, data from other species suggest functional hypotheses for future testing of these genes in OvLV and other lentiviral infections. Specifically, SYTL3 binds and may regulate RAB27A, which is required for enveloped virus assembly of human cytomegalovirus. Zinc finger transcription factors have been associated with positive selection for repression of retroviral replication. DLGAP1 binds and may regulate DLG1, a known regulator of HIV infectivity.


Editor: Baochuan Lin, Naval Research Laboratory, United States of America

Received June 19, 2012; Accepted September 17, 2012; Published October 17, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This research was funded by USDA-ARS CWA No. 5348-32000-031-00D. (http://www.ars.usda.gov/main/main.htm). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Stephen.White@ars.usda.gov

Introduction

Ovine lentivirus (OvLV), like the human immunodeficiency virus (HIV), is a macrophage-tropic lentivirus that leads to persistent, lifelong infection of the host [1,2]. Seroprevalence of OvLV is 24–26% in U.S. domestic sheep, with positive animals in approximately half of all U.S. flocks [3,4]. Ovine lentivirus infection leads to varying degrees of dyspnea (respiratory distress), cachexia (body condition wasting), mastitis, arthritis, and/or encephalitis [5,6]. One of the most commonly observed symptoms is an interstitial pneumonia [5,6,7,8] that has led to additional names for OvLV including ovine progressive pneumonia virus and maedi-visna virus, from the Icelandic “maedi” for dyspnea. As with all known lentiviruses, there is currently no known way to eliminate OvLV infection, and no vaccine to completely prevent infection [9].

However, breed differences in susceptibility to OvLV suggest that it may be possible to breed animals with reduced susceptibility to the virus. It has been shown that reproducible breed differences exist in seroprevalence of OvLV across several studies done in different locations over many decades [10,11,12,13,14]. For example, Rambouillet sheep had lower odds of infection, Columbia sheep had higher odds of infection, and the recent

The authors have declared that no competing interests exist.
Results

Genotypes and Phenotypes

From the original set of 997 animals, sample quality control criteria eliminated 18 samples from additional analysis. Multidimensional scaling (MDS) identified 3 clusters corresponding to breed, and 15 animals were eliminated for outlier status (Figure S1). The remaining animals included Rambouillet (N = 399), Polypay (N = 423), and Columbia (N = 142). Pairwise population concordance (PPC) clusters matched breed for Columbia and Polypay (N = 423), and Columbia (N = 142). Pairwise population stratification was eliminated by accounting for SNP or genes stratification (Figure S3), but most of this apparent population stratification was eliminated by accounting for SNP or genes stratification in the Columbia analysis.

A Manhattan plot showing P-values arranged by chromosome position is shown in Figure 1. A series of Quantile-Quantile (Q-Q) plots showing observed versus expected P-value distributions are shown in Figures S3,S4,S5. Initial analysis showed population stratification (Figure S3), but most of this apparent population stratification was eliminated by accounting for SNP or genes identified in Figure 1 (Figures S4,S5).

Decline of Linkage Disequilibrium with Distance

To estimate the length of useful LD and thus provide genome-wide average expectation for the approximate distance from identified SNP within which underlying mutations might be found, LD was calculated for pairs of observed SNP. Decline of LD with physical distance in each breed is shown in Figure S2. The first distance bin exceeding the threshold of 1000 observations was 5–10 kilobase pairs (Kb), and this bin is reported in Figure S2 as 10 Kb. All individual breeds had $r^2 \leq 0.40$ at 10 Kb and $r^2 \leq 0.25$ at 35 Kb (Figure S2). The combined all-breeds set had lower LD of $r^2 \leq 0.35$ at 10 Kb, $r^2 \leq 0.25$ at 20 Kb, and $r^2 \leq 0.20$ at 35 Kb (Figure S2).

Genome-wide Association with Serological Status

Because OvLV infection is lifelong, concordance between serological status and direct viral measures of infection is high and serological status is a proxy for odds of infection in exposed populations [19]. Genome-wide association with serological status yielded 13 unique SNP, including 3 genome-wide significant and 10 genome-wide suggestive SNP (Table 1). For each SNP, results from the best-fitting mode-of-inheritance association model were reported in Table 1. Only one SNP (OAR1_185938380) was identified in separate analyses of both the all-breeds animal set and in an individual breed (Polypay). After accounting for TMEM154 risk status as previously reported [16], the Table 1 SNP were associated with serologic status to varying degrees except OAR17_5388531, the SNP located within TMEM154 (Table 2). Further, no additional SNP became genome-wide suggestive after accounting for TMEM154 risk status.

For each analysis, a second run was performed that dropped screening criteria including genotype call rate by individual and by SNP, minor allele frequency, and Hardy-Weinberg equilibrium. This technique had the potential to identify associated SNP that failed one or more of the screening criteria, for example common homozygous lethal genomic regions, but no additional genome-wide suggestive SNP were identified. No SNP reached genome-wide suggestive association in the Columbia analysis.

A Manhattan plot showing P-values arranged by chromosome position is shown in Figure 1. A series of Quantile-Quantile (Q-Q) plots showing observed versus expected P-value distributions are shown in Figures S3,S4,S5. Initial analysis showed population stratification (Figure S3), but most of this apparent population stratification was eliminated by accounting for SNP or genes identified in Figure 1 (Figures S4,S5).

Genome-wide Association with Proviral Concentration

The concentration of provirus in peripheral blood was used as a measure of control of infection, and it has been shown to be correlated with OvLV pathology [20]. Genome-wide association with proviral concentration identified 13 SNP, including 2 genome-wide significant and 11 genome-wide suggestive SNP (Table 3). For each SNP, results from the best-fitting mode-of-inheritance association model were reported in Table 3. The genotypic mean proviral concentrations, adjusted for age and breed, were reported in Table S1.

For each analysis, a second run was performed that dropped screening criteria including genotype call rate by individual and by SNP, minor allele frequency, and Hardy-Weinberg equilibrium. This technique had the potential to identify associated SNP that failed one or more of the screening criteria, but no additional genome-wide suggestive SNP were identified. The Columbia animal set included fewer animals with many fewer polymorphic SNP, and results from association with proviral concentration in
Columbia sheep were reported in Table S2. Genotype frequencies of reported SNP were included in Table S3.

A Manhattan plot showing P-values arranged by chromosome position is shown in Figure 2. A series of Quantile-Quantile (Q-Q) plots showing observed versus expected P-value distributions are shown in Figures S6, S7. The initial analysis showed apparent population stratification (Figure S6), but most of this apparent population stratification was eliminated by accounting for SNP in Table 3 (Figure S7).

Table 1. Genomic regions associated with susceptibility to OvLV.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position (bp)</th>
<th>Animal Set</th>
<th>Best fitting model</th>
<th>Nominal P-value</th>
<th>Empirical P-value</th>
<th>Odds Ratio</th>
<th>Genes within 100 Kb on either side</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAR1_185953850</td>
<td>1</td>
<td>172,600,491</td>
<td>All</td>
<td>additive</td>
<td>3.3 x 10^-6</td>
<td>0.006</td>
<td>1.98</td>
<td>DPPA2**, DPPA4*</td>
</tr>
<tr>
<td>OAR1_185953850</td>
<td>1</td>
<td>172,600,491</td>
<td>Polypay</td>
<td>additive</td>
<td>6.5 x 10^-7</td>
<td>0.048</td>
<td>2.60</td>
<td>DPPA2**, DPPA4*</td>
</tr>
<tr>
<td>OAR1_186779231</td>
<td>1</td>
<td>173,437,685</td>
<td>Polypay</td>
<td>dominant</td>
<td>1.7 x 10^-7</td>
<td>0.012</td>
<td>4.45</td>
<td></td>
</tr>
<tr>
<td>OAR4_38205790</td>
<td>4</td>
<td>35,398,410</td>
<td>All</td>
<td>dominant</td>
<td>3.1 x 10^-6</td>
<td></td>
<td>2.84</td>
<td></td>
</tr>
<tr>
<td>s54511</td>
<td>6</td>
<td>12,045,872</td>
<td>Polypay</td>
<td>additive</td>
<td>3.8 x 10^-6</td>
<td></td>
<td>2.36</td>
<td>CAMK2D**</td>
</tr>
<tr>
<td>OAR7_82644472</td>
<td>7</td>
<td>75,607,567</td>
<td>Polypay</td>
<td>additive</td>
<td>5.7 x 10^-6</td>
<td></td>
<td>2.53</td>
<td>GPHN**</td>
</tr>
<tr>
<td>OAR8_7355614</td>
<td>8</td>
<td>68,927,958</td>
<td>All</td>
<td>recessive</td>
<td>8.7 x 10^-6</td>
<td></td>
<td>2.61</td>
<td>UTRN*, STX11</td>
</tr>
<tr>
<td>OAR8_88021348</td>
<td>8</td>
<td>82,158,519</td>
<td>Rambouillet</td>
<td>recessive</td>
<td>4.2 x 10^-6</td>
<td>0.051</td>
<td>4.46</td>
<td>SYTL3**, GTF2H5**, DYNLT1*, TMEM181, EZR</td>
</tr>
</tbody>
</table>

Table 2. Genomic regions associated with OvLV serologic status comparing association with and without accounting for TMEM154 mutations.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Animal Set</th>
<th>Nominal P-value (Without TMEM154)</th>
<th>Nominal P-value (Accounting TMEM154)</th>
<th>Empirical P-value (Without TMEM154)</th>
<th>Empirical P-value (Accounting TMEM154)</th>
<th>Genes within 100 Kb on either side</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAR1_185953850</td>
<td>All</td>
<td>3.3 x 10^-6</td>
<td>5.2 x 10^-6</td>
<td>0.006</td>
<td></td>
<td>DPPA2**, DPPA4*</td>
</tr>
<tr>
<td>OAR1_185953850</td>
<td>Polypay</td>
<td>6.5 x 10^-7</td>
<td>3.8 x 10^-5</td>
<td>0.048</td>
<td></td>
<td>DPPA2**, DPPA4*</td>
</tr>
<tr>
<td>OAR1_186779231</td>
<td>Polypay</td>
<td>1.7 x 10^-7</td>
<td>4.5 x 10^-6</td>
<td>0.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OAR4_38205790</td>
<td>All</td>
<td>3.1 x 10^-6</td>
<td>1.0 x 10^-4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>s54511</td>
<td>Polypay</td>
<td>3.8 x 10^-6</td>
<td>6.7 x 10^-4</td>
<td>1</td>
<td>1</td>
<td>CAMK2D**</td>
</tr>
<tr>
<td>OAR7_82644472</td>
<td>Polypay</td>
<td>5.7 x 10^-6</td>
<td>3.0 x 10^-5</td>
<td>1</td>
<td>1</td>
<td>GPHN**</td>
</tr>
<tr>
<td>OAR8_7355614</td>
<td>All</td>
<td>8.7 x 10^-6</td>
<td>5.8 x 10^-5</td>
<td>1</td>
<td>1</td>
<td>UTRN*, STX11</td>
</tr>
<tr>
<td>OAR8_88021348</td>
<td>Rambouillet</td>
<td>4.2 x 10^-6</td>
<td>3.2 x 10^-6</td>
<td>0.051</td>
<td>0.052</td>
<td>SYTL3**, GTF2H5**, DYNLT1*, TMEM181, EZR</td>
</tr>
</tbody>
</table>

5): P > 0.15
**: SNP located within gene
*: SNP located within 35 Kb of gene

doi:10.1371/journal.pone.0047829.t001
doi:10.1371/journal.pone.0047829.t002
Discussion

Breed differences in susceptibility to OvLV have long suggested genetic involvement, and a recent study identified TMEM154 mutations as consistently associated with OvLV prevalence [16]. However, those mutations did not confer complete resistance, and some flocks with a high frequency of the less susceptible genotypes nonetheless have high OvLV prevalence. This GWAS identified multiple additional genomic regions as associated with OvLV susceptibility. Further, this was the first GWAS to examine control of OvLV replication to our knowledge, and multiple genomic regions were identified as associated with control of OvLV replication. The short LD length observed (Figure S2) was consistent with other reports in sheep [17] and implicated a smaller number of genes within each associated region than in many other mammal species [21,22,23,24,25]. For each region associated with either susceptibility or control of OvLV, the association suggested one or more underlying variants within the region had a functional relationship with some aspect of OvLV infection, replication, or transmission. Some regions contained promising candidate genes that suggested further biological hypotheses for specific gene product involvement. Below, we summarize major results as defined by empirical P-values and discuss future hypotheses of function to test for genes identified in connection with odds of infection and then for genes identified as associated with OvLV control of infection.

Genes involved in odds of infection could be used to select sheep with lower probabilities of natural infection, and the top GWAS markers provide important evidence for the involvement of several interesting genes. Only one gene had been identified previously as consistently associated with susceptibility to OvLV infection [16], and this GWAS also confirmed the association with TMEM154 (empirical P = 0.13; Table 1). Further, the odds ratio was estimated at 7.57 (Table 1), which is consistent with prior data [16]. A recent study identified TMEM154 in a human GWAS for asthma severity [26], suggesting that TMEM154 may play a conserved role in airway immune responses. Only markers observed at empirical significance exceeding the TMEM154 region harboring known mutations (empirical P<0.10) will be discussed further.

The locus most highly associated with odds of infection was a SNP in the DPPA2 (Developmental Pluripotency Associated 2) gene which was associated in both the all-breeds analysis (empirical P = 0.006; Table 1) and in Polypays analyzed separately (empirical P = 0.048; Table 1). DPPA2 and the closely related DPPA4 (Developmental Pluripotency Associated 4) gene are expressed in embryo and germ line cells [27], and there are many potential mechanisms by which DPPA2 or DPPA4 may be involved in OvLV infection. Both DPPA2 and the nearby DPPA4 play essential roles in lung development and formation [28,29], suggesting differential lung development in sheep may play a role in OvLV susceptibility. However, OvLV has been found in semen and in the female reproductive tract of sheep [30,31], and the possibility of sexual transmission and/or paternal transmission of the virus are important unanswered questions. If OvLV is sexually or paternally transmitted, germline expressed DPPA2 may play a role in some aspect of transmission. A third possibility is differential development of the immune system. Since DPPA2 is expressed during embryonic development, it is possible that DPPA2 variants may influence the development or maintenance of the immune system. The SheepQTLdb [32,33] shows a QTL for Haemonchus contortus...
<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position (bp)</th>
<th>Animal Set</th>
<th>Best fitting model</th>
<th>Nominal P-value</th>
<th>Empirical P-value</th>
<th>Genotypic Log(_{10}) Conc. Diff.</th>
<th>Genes within 100 Kb on either side</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU231007_156</td>
<td>3</td>
<td>58,955,947</td>
<td>Polypay</td>
<td>dominant</td>
<td>3.5×10(^{-6})</td>
<td>0.78</td>
<td>PAX8, IGK</td>
<td></td>
</tr>
<tr>
<td>OAR3_144283427</td>
<td>3</td>
<td>135,036,950</td>
<td>Polypay</td>
<td>genotypic</td>
<td>2.0×10(^{-6})</td>
<td>0.84</td>
<td>SLC11A2*</td>
<td></td>
</tr>
<tr>
<td>OAR3_144414855</td>
<td>3</td>
<td>135,043,018</td>
<td>Polypay</td>
<td>genotypic</td>
<td>2.0×10(^{-6})</td>
<td>0.84</td>
<td>SLC11A2**</td>
<td></td>
</tr>
<tr>
<td>s27054</td>
<td>5</td>
<td>6,250,424</td>
<td>Polypay</td>
<td>recessive</td>
<td>1.3×10(^{-6})</td>
<td>0.047</td>
<td>0.84 C19orf42**, TMEM38A*, NWD1, MED26, SLC35E1, CHERP</td>
<td></td>
</tr>
<tr>
<td>OAR9_10735564</td>
<td>9</td>
<td>10,702,461</td>
<td>Polypay</td>
<td>dominant</td>
<td>1.6×10(^{-6})</td>
<td>0.073</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>OAR9_10749779</td>
<td>9</td>
<td>10,723,321</td>
<td>Polypay</td>
<td>additive</td>
<td>1.5×10(^{-6})</td>
<td>0.069</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>s48118</td>
<td>9</td>
<td>14,589,931</td>
<td>Polypay</td>
<td>dominant</td>
<td>9.4×10(^{-6})</td>
<td>0.62</td>
<td>BAI1**, LOC529919*, ARC, RPL38</td>
<td></td>
</tr>
<tr>
<td>OAR13_56607666</td>
<td>13</td>
<td>52,062,577</td>
<td>All</td>
<td>additive</td>
<td>4.3×10(^{-6})</td>
<td>0.58</td>
<td>TGM6*</td>
<td></td>
</tr>
<tr>
<td>OAR18_5646940</td>
<td>18</td>
<td>5,926,257</td>
<td>Polypay</td>
<td>genotypic</td>
<td>2.5×10(^{-6})</td>
<td>0.68</td>
<td>MEF2A**</td>
<td></td>
</tr>
<tr>
<td>OAR18_5701234</td>
<td>18</td>
<td>5,984,107</td>
<td>Polypay</td>
<td>genotypic</td>
<td>2.6×10(^{-6})</td>
<td>0.68</td>
<td>MEF2A**</td>
<td></td>
</tr>
<tr>
<td>s65956</td>
<td>20</td>
<td>29,213,047</td>
<td>Rambouillet</td>
<td>dominant</td>
<td>5.9×10(^{-8})</td>
<td>0.001</td>
<td>ZNF192*, ZSCAN16*, ZNF165*, ZNF389*</td>
<td></td>
</tr>
<tr>
<td>OAR22_43742889</td>
<td>22</td>
<td>39,013,397</td>
<td>Polypay</td>
<td>additive</td>
<td>9.4×10(^{-6})</td>
<td>1.32</td>
<td>INPP5F**, MCM8P, BAG3</td>
<td></td>
</tr>
<tr>
<td>OAR23_40410527</td>
<td>23</td>
<td>38,376,040</td>
<td>All</td>
<td>additive</td>
<td>1.5×10(^{-6})</td>
<td>0.092</td>
<td>0.66 DLGAP1**</td>
<td></td>
</tr>
</tbody>
</table>

1: P>0.15
**: SNP located within gene
*: SNP located within 35 Kb of gene
doi:10.1371/journal.pone.0047829.t003
fecal egg count [34] that overlaps this region. An intriguing possibility would involve DPPA2 influencing both OvLV and parasitic infection through alterations in immune system development, since alterations in embryonic and/or white blood cell development could influence formation and/or maintenance of immune responses.

A second locus associated with OvLV susceptibility after correction for multiple testing was identified in Polypays (OAR1_186779231; empirical P = 0.012) and was located 837 Kb distal on ovine chromosome 1 from the SNP in DPPA2 (Table 1). This distance is considerably further than the genome-wide average decline of LD would suggest (Figure S2), but there is nonetheless appreciable LD between the markers (Table S4). In Polypays, these two SNP had r² = 0.40, and in all breeds examined these SNP had r² = 0.23 (Table S4). It is possible that the associations of both SNP reflect a single underlying gene through long-range LD due to admixture in the composite breed formation of the Polypay population [35,36]. Alternatively, a separate associated genetic factor could be present, though the current version of the sheep genome contains no annotated genes near OAR1_186779231. Nonetheless, the example of the callipyge muscular hypertrophy in sheep demonstrates that even a single nucleotide change in a gene desert can have a dramatic impact [37,38].

A separate region associated with odds of infection included SYTL3 (Synaptotagmin-Like 3) on ovine chromosome 8, and this region had the second largest odds ratio after TMEM154 (Table 1). SYTL3 is a peripheral membrane protein that interacts with RAB27A (RAB27A, member RAS oncogene family) and is thought to play a role in vesicular trafficking [39,40]. To our knowledge there has been no report of SYTL3 associated with any lentivirus infection in any mammal, but RAB27A is required for enveloped virus assembly of human cytomegalovirus [41]. SYTL3 may regulate RAB27A and thereby interfere with viral assembly. Further, RAB27A is a known negative regulator of phagocytosis [42], and is involved in exosome synthesis [43]. As such, SYTL3 could also influence either of those processes. While the current version of the sheep genome also shows SNP OAR8_88021348 within transcriptional regulator GTF2H5 (General Transcription Factor IIIH, polypeptide 5), the gene order does not agree with other mammal genomes including cow, human, dog, and mouse, and it may well be a mis-assembly in the current version of the sheep genome.

A further analysis of genomic regions associated with serological status incorporated data on risk status defined by known TMEM154 diplotypes [16]. In this analysis, SNP near DPPA2/DPPA4 dropped from genome-wide significant to genome-wide suggestive, but the SNP in SYTL3 retained its level of significance (Table 2). This strengthened the case for association between serological status and an underlying variant near SYTL3. It is possible that part of the association between serological status and SNP in or near DPPA2/DPPA4 was due to random association between genotypes at TMEM154 and DPPA2/DPPA4. However, this type of analysis tested association of SNP in or near the DPPA2/DPPA4 region averaged over all genotypes of TMEM154. Thus, it could underrepresent association in the presence of epistasis, where differential association exists at one locus by genotype at another. However, the relatively small overlap between genotype frequencies of SNP near DPPA2/DPPA4 and risk diplotypes of TMEM154 suggested a larger sample including patients with different clinical outcomes.
additional animals would be required to perform a thorough test for genetic interaction.

As well as adding more genomic regions associated with odds of infection, this was the first genome-wide study to examine association with the control of viral replication as measured by proviral concentration, to our knowledge. The most significant SNP (s#5936; empirical P = 0.001) was located in a cluster of zinc finger genes on ovine chromosome 20 (Table 3). Zinc finger proteins are known to have undergone duplication and divergent positive selection in the DNA-binding zinc finger domain following challenge by new retroviruses during mammalian evolution [44,45]. This suggests the hypothesis that certain zinc finger proteins might interfere with lentiviral replication, and indeed that has recently been shown with the Zinc Finger Antiviral Protein (ZAP) and HIV-1 [46]. However, to our knowledge none of the zinc finger proteins in the cluster identified here has ever been shown to be associated with any mammalian lentiviral infection.

Another empirically significant SNP (s#7053; P = 0.047) was located within both C19orf42 ([human] Chromosome 19 Open Reading Frame 19) and TMEM33a (Transmembrane Protein 33a) (Table 3). Not much is known about either C19orf42 or TMEM33a at this time. C19orf42 is a short (75 amino acid) open reading frame with similarity to Yos1, a yeast protein required for transport between endoplasmic reticulum and Golgi complex [47]. As such, C19orf42 could be involved in efficiency of viral packaging, but data to demonstrate such involvement are currently lacking. TMEM33a encodes a protein associated with vascular smooth muscle control of blood pressure [48]. It is possible that TMEM33a could be involved by an as-yet undetermined mechanism. Much work remains to be done to elucidate the potential roles of these genes in limiting OvLV replication.

Finally, there were two additional regions with suggestive empirical association (P < 0.10) with proviral concentration. The first included two SNP on ovine chromosome 9 in Polypays (Table 3), but no genes are annotated in that region in the current pre-release version of the sheep genome. The other is SNP OAR23_4010527 located within the DLGAP1 (Discs Large (Drosophila) Homolog-Associated Protein 1) gene (Table 3). The product of DLGAP1 is a primary interacting protein of DLG1 and is believed to function by holding DLG1 at place in cell junctions [49,50,51]. DLG1 is a strong negative regulator of HIV-1 infectivity, such that depletion of DLG1 enhances HIV infectivity 5–6 fold [52]. The allelic variant(s) of DLGAP1 in sheep may physically interfere with DLG1 positioning and/or function. Though DLGAP1 was associated with mouse survival following influenza H5N1 infection [53], possibly mediated through its interaction with DLG1 [54], to our knowledge DLGAP1 has not been reported to have a role in lentiviral infection in any mammal to date.

In conclusion, this GWAS found many additional genomic regions besides TMEM154 associated with odds of infection, and this was the first genome-wide study to provide regions associated with OvLV control as measured by proviral concentration. These results confirmed that TMEM154 is an important component of host susceptibility to OvLV, but there were many additional factors that contributed to genetic differences in susceptibility to and control of OvLV infection. These included many genes never previously associated with lentiviral infection, which may extend structural and/or regulatory networks implicated in lentiviral transmission, infection, and/or replication. These genes provide targets for additional investigation into lentiviral infection that may generalize beyond OvLV to other members of the lentiviral family. Further, the associated genes provide the basis for additional work to identify genetic markers associated with odds of OvLV infection and proviral concentration. As has already been done with TMEM154, such markers may be used to select sheep with lower odds and/or improved control of OvLV infection. Finally, they may also be of great interest for further study in goats, which are host to the closely related caprine encephalitis arthritis virus.

**Materials and Methods**

**Ethics Statement**

All animal care and handling procedures were reviewed and approved by the Washington State University Institutional Animal Care and Use Committee (Permit Number: 3171) and/or by the U.S. Sheep Experiment Station Animal Care and Use Committee (Permit Numbers: 10-06, 10-07). All efforts were made to minimize any discomfort during blood sampling.

**Populations and Phenotypes**

Flocks were chosen for study with high historical OvLV seroprevalence among mature ewes [15]. Specifically, whole blood was collected from ewes of Rambouillet (N = 414), Polypay (N = 430), and Columbia (N = 145) breeds, ages 1–5 years, from the U.S. Sheep Experiment Station. These animals were managed similarly but bred separately in pure breed groups. Blood was processed for serum and peripheral blood leukocytes as previously described [19]. A subset of 365 older ewes (ages 3–5) had been used as validation animals in a previous study on TMEM154 [16]. Serologic data were collected using a competitive ELISA assay (VMRD Inc., Pullman, WA) to detect anti-OvLV antibodies in sheep [55]. DNA was extracted from peripheral blood leukocytes and OvLV proviral concentrations were determined by a validated qPCR method [19].

**Genotyping**

Blood was collected by jugular venipuncture into EDTA-coated vacutainer tubes. DNA was isolated using the Invitrogen GeneCatcher™ gDNA 3–10 ml Blood Kit as per manufacturers' instructions (Life Technologies, Carlsbad, CA). The DNAs were checked for quality and quantity using an ND-1000 spectrophotometer (Nanodrop, Wilmington, DE) and equilibrated to 50 ng/ml. DNA samples from a subset of animals were processed for serum and peripheral blood leukocytes as previously described [19]. A subset of 365 older ewes (ages 3–5) had been used as validation animals in a previous study on TMEM154 [16]. Statistical analysis was performed using the PLINK software package (http://pngu.mgh.harvard.edu/purcell/plink/) written by Shaun Purcell [56] to identify genotypic outliers. Genotyping services were provided by Geneseek Inc. (Lincoln, NE) using the OvineSNP50 Infinium BeadChip (Illumina Inc., San Diego, CA) with a set of 54,977 SNP designed by the International Sheep Genome Consortium [17].

**Statistical analysis**

Genotypic quality control and clustering. Initial quality control criteria included low genotype call rates and high genotype identity to any other sample. Samples with low genotype call rates (<97%) were removed from additional analysis. PLINK identity-by-state analysis identified samples that could have been involved in label errors (as determined by IBS distances >0.95) for removal from additional analysis. Multidimensional scaling (MDS) was performed in the PLINK software package (http://pngu.mgh.harvard.edu/purcell/plink/) written by Shaun Purcell [56] to check for genotypic outliers. MDS was performed on a reduced set of markers in approximate linkage equilibrium from animals of each breed separately. Outlier animals were removed to reduce population stratification. Then, a pairwise population concordance test (--ppc option) of the hypothesis that each animal was from a separate breeding population was performed using a P-value of 10^-5, which is P = 0.05 after Bonferroni-correction for the total number of pairwise comparisons between animals. Where the resulting clusters contained additional information compared to
recorded breed data, the clusters were included in addition to breed as stratification categories for later association analyses.

**Linkage Disequilibrium Versus Genomic Distance.** Analysis of high density SNP data was performed using PLINK v1.06 [56]. A preliminary screen was performed to eliminate SNP with minor allele frequencies less than 0.10 from comparisons used to calculate average LD. LD within each breed was calculated with settings to include comparisons of all SNP within 5 Mbp of each other and minimum reported $r^2$ adjusted to 0.0 to include even low LD comparisons. These $r^2$ values were analyzed by distance bins with SAS 9.2 (SAS Institute, Cary, NC). Distance bins were calculated for each 5 Kbp interval beginning with 0–5 Kbp. Mean $r^2$ was only included in analysis if a bin contained at least 1000 values to assure each reported mean was well-estimated.

**Association analysis.** The OvLV positive or negative status as measured by cELISA was analyzed using separate logistic models for the minor allele of each SNP in PLINK to account for breed and pairwise population concordance clusters, for animal age as a covariate, and for the SNP minor allele. Log$_{10}$-transformed proviral concentrations were analyzed using similar separate general linear models for the minor allele of each SNP in PLINK to account for breed and pairwise population concordance clusters, for animal age as a covariate, and for the SNP minor allele. For both logistic and linear analyses, PLINK screening criteria were employed including missingness by individual (0.1), missingness by marker (0.03), minor allele frequency (0.01), and Hardy-Weinberg equilibrium (0.000001, which is $P = 0.05$ Bonferroni-corrected for 50,000 SNP tests). Genome-wide significance was defined by empirical $P \leq 0.05$. Genome-wide suggestive results were defined by nominal $P$-values $< 1 \times 10^{-5}$ [57]. Both logistic and linear analyses were performed using association models including additive allelic, genotypic 2 degree-of-freedom, dominant, and recessive. Family structure was addressed for both logistic and general linear models using permutation within sire families of 10 or more genotyped offspring; the remaining sire families with fewer than 10 offspring were grouped together for permutation purposes. One thousand permutations were used to obtain each empirical $P$-value. Additional analyses of serologic data included the presence or absence of risk haplotypes as defined by Heaton et al. [16] as covariates in the respective association models. Visualization of association data in manhattan and quantile-quantile plots was performed using a script generously provided by Dr. Stephen Turner (http://gettinggeneticsdone.blogspot.com/2011/04/annotated-manhattan-plots-and-qq-plots.html, viewed on 11-15-11) using the R environment [58]. Since PLINK only reports regression coefficients as a measure of effect size for linear regression, SAS 9.2 (SAS Institute, Cary, NC) was used to run similar genotypic models in the general linear models procedure to obtain largest adjusted genotypic mean differences in log$_{10}$-transformed proviral concentration as a measure of effect size for control of OvLV.

**Supporting Information**

**Figure S1** Multidimensional scaling of genotypes showing 3 breed clusters of animals included in analysis. Colombias are included in the top cluster, Polypays in the bottom right cluster, and Rambouillets in the bottom left cluster. The clustering of individuals by breed is clear even from these related breeds. The Columbia breed was developed pre-1920 with ½ Rambouillet composition [59]. The Polypay breed was developed in the 1970s with ¼ Rambouillet composition [36].

**Figure S2** Decline of linkage disequilibrium with distance by breed.

**Figure S3** Quantile-Quantile plot for odds of infection. Quantile-quantile plots from association with serological status, where the red line shows the expected distribution. Representative data from the all-breeds, additive mode of inheritance analysis are shown. The results show deviation from the expected distribution indicating population stratification by factors unaccounted in the analytic model, which could include frequencies of underlying mutations for susceptibility loci that differ between seropositive and seronegative individuals.

**Figure S4** Quantile-Quantile plot for odds of infection conditioned on TMEM154 risk status. A second analysis conditioned on TMEM154 risk status shows an observed distribution closer to expected than the primary analysis, but still does not account for the majority of apparent population stratification.

**Figure S5** Quantile-Quantile plot for odds of infection conditioned on Table 1 SNP. A third analysis conditioned on all the SNP in Table 1 shows a distribution much closer to expected, demonstrating that host genetic factors tracked by these SNP account for the majority of apparent population stratification.

**Figure S6** Quantile-Quantile plot for control of viral replication. Quantile-quantile plot from association with proviral concentration, where the red line shows the expected distribution. Representative data from the Rambouillet, dominant mode of inheritance analysis are shown. The results show deviation from the expected distribution indicating population stratification by factors unaccounted in the analytic model, which could include frequencies of underlying mutations for susceptibility loci that differ by proviral concentration.

**Figure S7** Quantile-Quantile plot for control of viral replication conditioned on Table 3 SNP. A second analysis was performed by conditioning on all the SNP in Table 3, minus close equivalents on the same chromosome ($r^2 > 0.8$; removed to prevent inestimable multicollinearity) for which only the best $P$-value SNP was retained from each pair. This analysis shows a distribution much closer to expected, demonstrating that host genetic factors tracked by these SNP account for the majority of apparent population stratification.
Acknowledgments

The authors gratefully acknowledge Codie Durfee, Nic Durfee, Liam Broughton-Neiswanger, W. Carl Johnson, Scott Bullers, Tish O’Rourke, Caylee Birge, Michael Lee, David Liu, Ralph Horn, James Allison, Duane Chandler, Tom Kellorn, and the farm crews at Washington State University, ADRU, and USSES for technical assistance. The genotyping of TMEM154/24 haplotypes was performed at USMARC. Special thanks to Dr. James Kijas for helpful review and input that improved the quality of the manuscript. Thanks to Dr. Stephen Turner of the University of Virginia Bioinformatics Core for providing R code for Manhattan and QQ plots.

Author Contributions

Conceived and designed the experiments: SNW JRK JAL KAL HLN GSL DPK. Performed the experiments: SNW LMH JAL KAL. Analyzed the data: SNW MRJ LMH. Contributed reagents/materials/analysis tools: SNW MRJ LMH HLN GSL DPK. Wrote the paper: SNW MRJ LMH JAL HLN GSL DPK.

References

10. Badke YM, Bates RO, Ernst CW, Schwab C, Steibel JP (2012) Estimation of diplotype was performed at USMARC. Special thanks to Dr. James Kijas for helpful review and input that improved the quality of the manuscript. Thanks to Dr. Stephen Turner of the University of Virginia Bioinformatics Core for providing R code for Manhattan and QQ plots.

GWAS Susceptibility/Control of Ovine Lentivirus

PLOS ONE | www.plosone.org 9 October 2012 | Volume 7 | Issue 10 | e47829


