# VARIANT DISCOVERY IN GENES IDENTIFIED AS DIFFERENTIALLY EXPRESSED GENES BETWEEN THE ABOMASAL LYMPH NODE TRANSCRIPTOME OF RESISTANT AND SUSCEPTIBLE ADULT SHEEP TO *TELADORSAGIA CIRCUMCINCTA* INFECTION

Chitneedi[1], P.K., Suárez-Vega[1], A., Martínez Valladares[2,3], M., Arranz[1] J.J. y Gutiérrez-Gil[1], B.

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. [2]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain. [3] Instituto de Ganadería de Montaña, CSIC-Universidad de León, 24346, Grulleros, León, Spain pchi@unileon.es

## INTRODUCTION

Gastrointestinal nematode infections are one of the major health issues facing grazing sheep populations and it incurs on major economic losses for sheep breeders. The resistance/susceptibility trait appears to be a highly complex trait (Behnke et al. 2003; Dominik 2005). In sheep resistance to nematode infection shows a moderate level of heritability (range 0.3-0.6) (Stear et al. 2001). Several QTL mapping studies have tried to identify genomic regions and mutations that influence resistance to nematode infection (Atlija et al. 2016; Coltman et al. 2001; Gutiérrez-Gil et al. 2009; Sayers et al. 2005), although the detection of causal mutations for this trait is still a challenge for the research community. The recently available RNA-seq technology provides the opportunity to extract high-throughput transcriptome data from a specific tissue to perform gene quantification, differential gene expression and detection of variants (SNPs and indels), which could be assessed as potential causal mutations (Hudson, Dalrymple, and Reverter 2012). A previous study of our research group has identified a list of 106 differential expression genes (DEGs) based on RNA-Seq dataset obtained from the abomasal lymph nodes of 12 adult sheep, previously classified as resistant or susceptible to GIN infection based on an artificial infection with *T. circumcincta* larvae Chitneedi et al. (2018). In the present study we present a detailed study of the variants mapping within the list of DEGs previously reported in that study. Thus, the present study provides a list of functionally relevant variants that could underlie the genetic control of resistance/susceptibility to *T. circumcincta* in adult sheep.

## MATERIALS AND METHODS

**Experimental infection:** The animals included in the present study were 12 adult ewes of Churra sheep from a commercial flock of Churra dairy breed reared under a semi-intensive management and belonging to the National Association of Spanish Churra sheep breeders (ANCHE). These animals were subjected to two experimental infections with *T. circumcincta* larvae, as described in detail by Chitneedi et al. (2018).

**RNA sequencing:** From the six resistant and the six susceptible sheep included in this study, mRNA was extracted from abomasal lymph nodes, using the Absolutely RNA miRNA Kit from Agilent (La Jolla, CA, USA). RNA integrity (RIN value) was analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The average RIN values of the RNA samples was 7.2 ranging between 6.7 and 7.5. After library preparation, the subsequent sequencing was performed with an Illumina HiSeq sequencer 2000, generating stranded paired-end reads of 75 base pairs with a depth of 30 million reads.

**Bioinformatics analysis**: The read quality of the 12 abomasal lymph node samples was determined using FastQC_v0.11.5 (Bioinformatics 2011). All the samples were aligned against the ovine genome assembly v.3.1 (Oar_v3.1) using the STAR_v2.5.3a aligner (Dobin et al. 2013). After alignment, editing read groups and marking duplicate reads were carried out with Picard_v2.9.0 (Broad Institute 2016). Then, the pre-processing of aligned reads, including read trimming, realignment and recalibration was performed with GATK_v3.7 (Van der Auwera et al. 2013). The variant calling analysis was individually performed using two different software, GATK_v3.7 and SAMTOOLS_v1.4 (Wysoker et al. 2009). Later, quality filters were applied independently to each of the resulting VCF files. The variants detected by Samtools were filtered with SnpSift (QUAL > 30) whereas for the GATK-detected variants the GATK specific filtering recommendations were followed (DP > 5 & QUAL > 30 & MQ > 40.0 & QD > 5.0 & FS < 60.0). After that, variants commonly identified by the two independent software were extracted with BCFtools_v1.4 ("isec" option) (Wysoker et al. 2009) and were considered as high-quality variants.

With the aim of characterizing transcriptome variants that may underlie sheep resistance to GIN, we used SnpEff ("–fi" option) followed by a bed file with the coordinates of the 106 DEGs reported by Chitneedi et al. (2018) to select the variants included in the studied genes. The variants extracted from the DEG coordinates were individually annotated for functional consequences with the software VEP_v90 (McLaren et al. 2016) and SnpEff_v4.3 (Cingolani et al. 2012). Later, we selected, for each subset, those predicted by the two annotation analyses to have relevant functional consequences (classified as HIGH or MODERATE impact). The selected variants were annotated with the online web tool VEP (https://www.ensembl.org/Tools/VEP) and the amino acid substitution effects on protein function were predicted using the SIFT algorithm (Kumar, Henikoff, and Ng 2009).

## RESULTS AND DISCUSSION

The pipeline followed for variant discovery with the transcriptome data of abomasal lymph node is shown in Figure 1. Based on the FastQC estimates, all the 12 samples analyzed were of high quality; thus, no trimming was performed. After aligning against the ovine reference genome (Oar_v3.1) around 80.27% paired reads per sample were uniquely mapped against the reference genome and around 8.5 % paired reads were mapped to multiple loci. After performing the variant calling and variant filtering, we found 1,326,960 common variants considered as high quality variants across the whole genome. From these high-quality variants, we extracted 6,168 variants (6,104 SNPs, 30 insertions and 34 deletions) mapping within the 106 DEGs. Among these, 332 variants (329 SNPs, 3 insertions and 2 deletions) were predicted to be of relevant functional impact ("High" and "Moderate"). The functional annotation of these variants with the VEP, which included 109 novel variants revealed a total of 471 functional consequences distributed across 60 genes. From these, the SIFT algorithm detected 50 deleterious variants most were missense consequence and two were splice region variants. These deleterious variants were included in a total of 15 genes. Four of these genes, *BPIFB1* (Zhou et al. 2017), *KRT20* (Sen et al. 2012), *SLC38A2* (Carter 2012) and FNDC1 (Sigdel et al. 2015) have been found to be associated with the immune response. Another gene harboring a missense deleterious variant was *MMP28*, which has been reported as responsible for cell proliferation in response to skin injury (Saarialho-Kere et al. 2002). Other genes harboring "High" impact variants were *LGALS4* (splice acceptor variant), *SLC38A2* (stop lost), *ASIC3* (stop gained and splice acceptor variant), and *SULF1* (frameshift variant). Future studies focusing on the

variations in these gene regions and pathway analysis combining these genes may provide information about potential causal mutations related to resistance against GIN infection in sheep. In addition, the deleterious variants which are located in the non-annotated gene regions were equally important for further investigation.

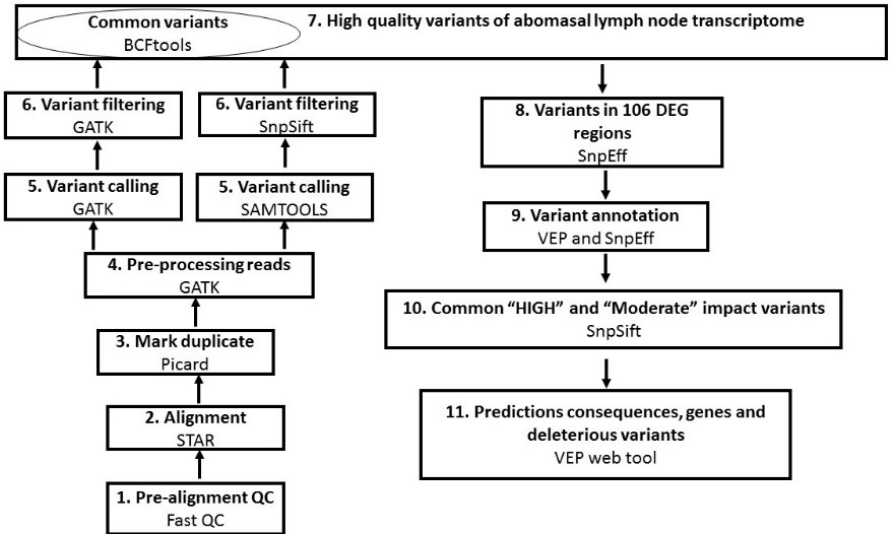


***Figure 1.*** *The pipeline shows the steps followed and tools used to detect variants and the consequences related to nematode infection in sheep found in DEG regions.*

## REFERENCES


•Atlija, M, et al. 2016. Genet. Sel. Evol 48: 4. • Van der A et al. 2013. Cur Prot in Bioinfo (SUPL.43). • Behnke, J M et al. 2003. J. helmi 77: 99–110. • Bioinformatics, Babraham. 2011. UK: Babraham Institute. • Broad Institute. 2016. Picard Tools. • Carter, A. M. 2012. Phys Rev 92: 1543–76. • Chitneedi, P K et al. 2018. Vet. Res 49: 39. • Churko et al. 2013. Circu. Res 112: 1613–23. • Cingolani, P et al. 2012. Fly 6: 80–92. • Coltman, D.W. et al. 2001. Paras 122: 571–82. • Dobin, Alexander et al. 2013. Bioinfo 29: 15–21. • Dominik, S. 2005. Genet. Sel. Evol 37 Sup 1: S83–96. • Gutiérrez-Gil, B et al. 2009. Genet. Sel. Evol 41: 46. • Hudson, N J. et al. 2012. BMC Geno 13. • Kumar, P. et al. 2009. N. Prot 4: 1073–82. • McLaren, W. et al. 2016. G. Biology 17. • Saarialho-Kere, U et al. 2002. J. I. Derm 119: 14–21. • Sayers, G. et al. 2005. Paras 131: 403–9. • Sen, A. et al. 2012. P. N. A. Sci 109: 20667–72. • Sigdel, K R et al. 2015. J. Imm. Res 2015. • Stear, M. J. et al. 2001. R Vet Sci 71: 1–7. • Wysoker, A et al. 2009 Bioinfo 25: 2078–79. • Zhou et al. 2017. Open Medicine (Poland) 12: 299–307. •



**Acknowledgements:** Financial support for this project was received from the LE248U14 project of Junta de Castilla and León Government. P. K. Chitneedi is funded by a predoctoral fellowship from the Junta de Castilla and León Government and the European Social Fund. B Gutiérrez-Gil is funded by the "Ramón y Cajal" Programme (RYC-2012-10230) from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO). M. Martínez-Valladares is also funded by the "Ramón y Cajal" Programme (RYC-2015-18368) from MINECO.