Meta-analysis of Genome Wide Association and Gene Expression Studies to Identify Candidate Genes for Tick Burden in Cattle

L. R. Porto Neto[1,2,3], E. K. Piper[1,4], N. N. Jonsson[1,4,5], W. Barendse[1,3] and C. Gondro[1,6]

Introduction

The cattle tick is endemic throughout some of the most productive cattle growing regions in the world, with a higher incidence in tropical and subtropical areas (Barker & Murrell 2004). Ticks and tick borne diseases are responsible for large losses in livestock productivity (Jongejan & Uilenberg 2004), estimated to be around 18 billion dollars per year globally (de Castro 1997).

Genome wide association studies (GWAS) and gene expression studies (GES) are two approaches being applied to study the tick-host interaction to understand the biological processes involved with host immune responses to tick infestation, and to identify genomic regions associated with tick burdens.

Very few markers in single nucleotide polymorphism (SNP) GWAS are functional mutations. Primarily the markers serve to indicate genomic regions that have a significant association with the trait in question. In GES the relative abundance of specific mRNA transcripts are estimated to gain insights into the biological/molecular processes taking place *in vivo*. A meta-analysis combining both approaches could potentially give more robust predictions adding biological information into the GWAS and reducing the background noise of expression arrays used in GES. Here we present a meta-analysis combining both approaches to the case study of tick resistance.

Material and methods

Genome wide association studies (GWAS). We used data from two GWAS studies:
1) Dairy cattle of a variety of breeds. The dairy study (Turner *et al*. 2010) comprised the analysis of 189 animals from the North-east of Australia, in which multiple tick counts were taken and animals genotyped with the GeneChip® Bovine Mapping 10K SNP (Affymetrix).
2) Brahman animals. The Brahman study used 482 animals from the Cooperative Research Centre for Beef Genetic Technologies (Barwick *et al*. 2009), in which serial scores for tick

[1] Cooperative Research Centre for Beef Genetic Technologies, Armidale, 2351, Australia
[2] The University of Queensland, School of Animal Studies, Gatton, 4343, Australia
[3] CSIRO Livestock Industries, Queensland Bioscience Precinct, St Lucia, 4067, Australia
[4] The University of Queensland, School of Veterinary Science, Gatton, 4343, Australia
[5] University of Glasgow, Faculty of Veterinary Medicine, Glasgow, G61 1QH, UK
[6] University of New England, The Centre for Genetic Analysis and Applications, Armidale, 2351, Australia

infestation were recorded (Prayaga *et al*. 2009). These Brahman samples were genotyped with the BovineSNP50® (Illumina) chip. The association between each SNP and tick burden was assessed by regression analysis of residual phenotype on numbers of copies of a particular allele (Table 1).

Gene expression studies (GES). Three Holstein-Friesian and three Brahman heifers were artificially infested with ticks weekly for seven weeks and tick side counts were undertaken (Piper *et al*. 2008). Skin biopsies were obtained from sites where a tick larva was attached to the skin and total RNA extracted. Skin biopsies were also obtained from three Holstein-Friesian and three Brahman heifers that had never before been exposed to ticks. Transcription profiling of the RNA was conducted using the Affymetrix GeneChip Bovine Genome Array platform. Transcription intensities were estimated in log2 scale from the probe level data using RMA summarization. Only significantly differentially expressed ($P <$ 0.001) probes with a log2 fold change of at least 1, average intensity above 8 and flagged as present in at least 50% of the samples were considered to be significant (table 1).

Table 1. Summary of the genome wide association and gene expression studies

***GWAS*[1]**	***n_animals***	***Phenotype***	***n_SNP*[2]**	***Significant*[3]**
Dairy	189	tick counts	7,458	32
Brahman	482	tick scores	53,797	142
***GES*[1]**	***n_animals***	***Phenotype***	***n_probes*[2]**	***Significant*[3]**
Holstein	6	naïve vs infested	18,058	153
Brahman	6	naïve vs infested	18,058	395

[1] GWAS – genome wide association studies, GES – gene expression studies. [2] Number of SNP or probes tested after filtering. [3] Number of markers/probes significant at 0.001 (dairy GWAS and GES) and 0.0001 (Brahman GWAS).

Meta-analysis. We mapped the significant SNP from the GWAS onto cattle chromosome positions and considered the genes located within a window of 1Mb either side of the SNP. For the GES we mapped the genes (probes) significantly differentially expressed between naïve and infested cattle of each breed. The overlaps were identified using FunctSNP (Goodswen *et al*. 2009) . The candidate genes were assigned to their gene ontology (GO) terms and gene clusters of the same ontology were observed and tested for overrepresentation.

Results and discussion

The positional mapping analyses for GWAS and GES is shown in Figure 1. Most of the chromosomes have significant regions for one or more experiments. The overlap regions, however, are present in only a reduced number of chromosomes. The strongest overlaps were found in chromosome 2, 10, 13 and 19 and are significantly overrepresented at $p<0.01$. In total 20 genes overlap across projects, which is more than expected by chance ($p<0.05$). These chromosomes are also significantly enriched in the number of differentially expressed genes outside the SNP overlap regions.

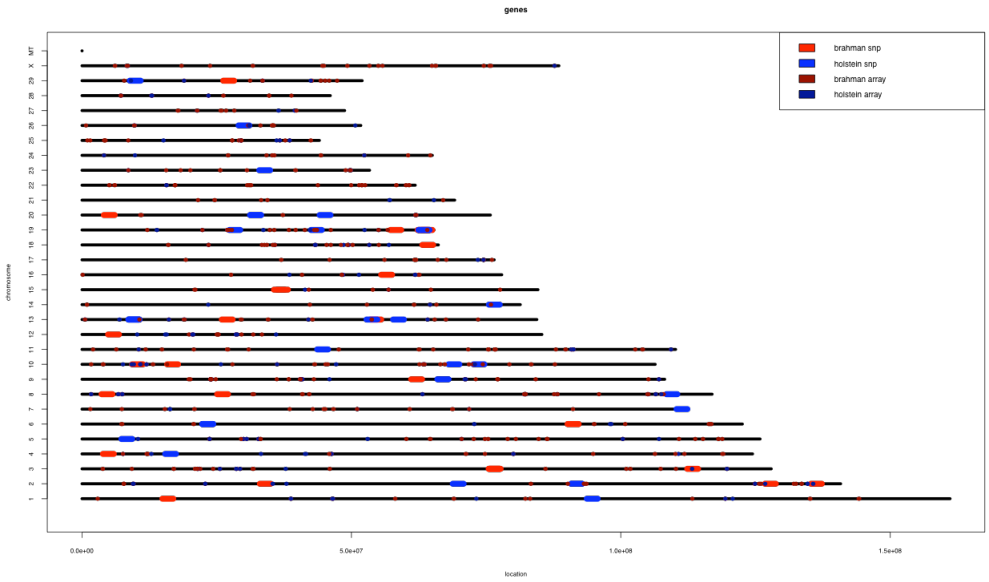

Figure 1. Chromosomal plot from chromosome 1 (bottom) to chromosome 29, X and mitochondrial (top). Significant SNP (GWAS) and gene probes (GES) are colour coded referring to respective experiment.

We perfomed a gene ontology analysis with all genes identified across projects (1,377 genes). There were several significant GO terms for biological process in our analysis. We excluded those that consisted of sequentially arranged genes, which are an artefact of the SNP window used, and considered only the GO terms supported by genes located on different chromosomes

Among the significant GO terms the most overrepresented ones are potentially related to tick resistance. GO0006953 (Acute-phase response) is supported by three genes in two chromosomes and GO0030595 (Leukocyte chemotaxis), also represented by three genes in two chromosomes, can be directly associated with parasite resistance while GO0016311 (Dephosphorilation) represented by ten genes located in seven different chromosomes is one of the most overrepresented.

There are at least two strategies to further explore the result of this meta-analysis: 1) explore candidate genes of a specific GO term cluster of interest or 2) investigate the top candidate genes from GWAS that overlap with GES according to the level of significance. Table 2 shows the top ten genes by significance level located in overlap regions.

Table 2. Top 10 candidate genes by significance level

Chr	*Gene Symbol*	*Gene Name*
2	LAPTM5	Lysosomal protein transmembrane 5
2	ALPL	Alkaline phosphatase, liver/bone/kidney
3	MFSD2A	Major facilitator superfamily domain containing 2A
10	DHRS7	Dehydrogenase/reductase (SDR family) member 7
10	GNPNAT1	Glucosamine-phosphate N-acetyltransferase 1
10	SERINC5	Serine incorporator 5
13	SIRPA	Signal-regulatory protein alpha
14	CA2	Carbonic anhydrase II
19	MPDU1	Mannose-P-dolichol utilization defect 1
26	ADD3	Adducin 3 (gamma)

Importantly some overlap regions not only were overrepresented in number of significant SNP in the GWAS but also in the number of significant probes in the GES.

Conclusion

These results show that by incorporating data from different sources we can 1) increase the reliability of GWAS studies and 2) remove noise from array data. Here we have identified new candidate genes and regions that appear to be involved in conferring tick resistance in cattle.

References

Barker S.C. & Murrell A. (2004) Systematics and evolution of ticks with a list of valid genus and species names. *Parasitology* 129, S15-S36.

Barwick S.A., Johnston D.J., Burrow H.M. *et al.* (2009) *Animal Production Science* 49, 367-82.

de Castro J.J. (1997) *Veterinary Parasitology* 71, 77-97.

Goodswen S.J., Kadarmideen H.N., Gondro C. *et al.* (2009) *Proceedings of the 18th Conference of the Australian Association for the Advancement of Animal Breeding and Genetics*. 18: 454-457.

Jongejan F. & Uilenberg G. (2004) *Parasitology* 129, S3-S14.

Piper E.K., Jackson L.A., Bagnall N.H. *et al.* (2008) *Veterinary Immunology and Immunopathology* 126, 110-9.

Prayaga K.C., Corbet N.J., Johnston D.J. *et al.* (2009) *Animal Production Science* 49, 413-25.

Turner L.B., Harrison B.E., Bunch R.J. *et al.* (2010) *Animal Production Science* Provisionally accepted.