Discovery of expression quantitative trait loci associated with Johne's disease using both RNA-seq and DNA variants

N. Bissonnette[1], J.-S. Brouard[1], O. Ariel[1], N. Gévry[2], E. Ibeagha-Awemu[1], and F. Miglior[3,4]

[1] *Sherbrooke Research and Development Centre, Agriculture and Agri-Food Canada, Sherbrooke, Qc, Canada J1M 0C8*
Nathalie.Bissonnette@canada.ca (Corresponding Author)
[2] *Département de Biologie, Université de Sherbrooke, Qc, Canada J1K 2R1*
[3] *Center of Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada N1G 2W1*
[4] *Canadian Dairy Network, Guelph, ON, Canada N1K 1E5*

Summary

Johne's disease (**JD**) is a debilitating chronic disease in ruminants caused by Mycobacterium avium ssp. paratuberculosis (**MAP**) which manipulates gut macrophages as survival strategies and for its dissemination. The endemic situation of JD can be in part explained by the lack of genetic resistance to MAP infection in cattle populations. A successful genetic improvement strategy results from a comprehensive understanding of the genetic variability associated with disease susceptibility/resistance, while providing information about the affected biological pathways. In this functional genomics study, accurate phenotypic data (e.g. diagnosis records) for JD were used to identify 22 MAP-infected (JD(+)) and 28 healthy/resistant (JD(–)) cows. The transcriptome of their blood circulating primary macrophages were analysed using the next-generation RNA sequencing (**RNA-Seq**) technology. DNA genotypes were also identified using a complementary strategy: the BovineSNP50 DNAchip imputed to the high density (HD) DNAchip. More than 60% of the 1,356,248 variants (call rate ≥ 0.2, minor allele frequency ≥ 0.05) were identified by RNA-seq, among which 12% (161,951) were novel variants. Genome-wide association study identified two major expression quantitative trait loci (**eQTL**) on BTA4 and 11 at $-\log_{10}(P) \geq 7$. Interestingly, 2,435 RNA-seq variants are predicted to produce high functional effect on known genes, while only 33 DNA genotypes (HD) were found in this category. RNA-Seq is an effective strategy to identify eQTL and thus increases the power to detect functional genetic variants. In the present study, we succeeded to identify eQTL and the regulatory pathways that discriminate MAP infected from healthy/resistant cows. Genetic variations in susceptible cows allow MAP to proliferate and escape the normal mycobacterial killing process of the macrophages. This strategy is thus highly relevant in genetic selection, as it may reduce disease susceptibility. An integration of the findings of this research (genomic information), into the conventional young sire selection and progeny testing program could yield a better, more accurate and rapid genetic improvement of resistance to bovine paratuberculosis in Canadian dairy herds.

Keywords: bovine paratuberculosis, primary macrophage, RNA-sequencing, GWAS, expression quantitative trait locus

Introduction

Johne's disease (**JD**) is a livestock disease caused by a zoonotic pathogen, leading to chronic diarrhoea and ill thrift in adult cattle. This is a slow progressive disease with unpredictable clinical signs, making it difficult to identify infected cows. Diagnosis of the disease is

challenging (Fock-Chow-Tho et al., 2017) and vaccines are ineffective at preventing infection (Bannantine et al., 2014, Ghosh et al., 2015). The prevalence of infected farms has been increasing worldwide, and JD (paratuberculosis) is now a global concern.

Enhancing the animals' natural disease resistance by improving the genetics is slow but the results are permanent. With the availability of the full bovine genomic sequence, numerous SNPs are annotated. A number of genome-wide association studies (**GWAS**) have been carried out to identify quantitative trait loci (**QTL**) associated with JD. These studies had found evidence for association on multiple and varying chromosomal locations. While numerous QTL were identified, we still have the daunting task of predicting 'genotype-to-phenotype' relationships mainly because markers for many of the traits of importance still explain a relatively small proportion of the JD resistance/susceptibility variance.

The JD causative pathogen, *Mycobacteria avium* spp. *paratuberculosis* (MAP), is an obligatory mycobacteria requisitioning macrophage for its multiplication and survival. The hypothesis of the present study is that genetic variations affecting macrophage activity in susceptible cows would allow MAP to proliferate and escape the normal mycobacterial killing process. Under the assumption that a better understanding of the transcriptome can lead to the identification of functional genetic variants, the primary goal of our study was to use JD case and control macrophages to identify biomarkers associated with JD.

The association between a genetic variant at a genomic locus and a trait is not directly informative but expressed QTL is informative with respect to the mechanism whereby the variant influence the phenotype. Nowadays, RNA-sequencing (**RNA-seq**) strategy is becoming increasingly affordable and increases the power to detect subtle pathway activity changes. The originality of this research is to perform a GWAS using RNA-Seq and DNA genotypes to provide a high-resolution genomic analysis. These findings provide a comprehensive strategy to unravel the relationship between genotype and phenotype in the context of host-pathogen interaction.

Materials and Methods

Animal selection, JD diagnosis, Differentiation of monocyte-derived macrophages, and RNA-seq data processing

Twenty-eight JD negative [JD(–)] and 22 JD positive [JD(+)] cows tested using serum ELISA and fecal PCR as described (Fock-Chow-Tho et al., 2017) were selected for isolation of blood circulating monocytes and *in vitro* differentiation into macrophages. All steps have been performed as previously described (Ariel et al., 2018).

The RNA from 12 cows (companion project) and from an additional 38 animals was extracted and cDNA libraries were constructed as described (Ariel et al., 2018). Variants were identified using the Genome Analysis Toolkit (GATK, (McKenna et al., 2010)) Joint Genotyping method using RNA-seq data (Brouard and Bissonnette, 2018) totalizing 9,820 billion paired-end reads.

DNA genotyping using Bovine SNP50 BeadChip, variant filtering, imputation, annotation, and functional analysis

The fifty animals analyzed in RNA-seq were also genotyped from blood cells' DNA using the commercial Illumina BovineSNP50 BeadChip and imputed to HD using FImpute v2.2 as described (Sargolzaei et al., 2014). SVS (Golden Helix) and the BCFtools (Li et al., 2009)

were used to obtain high quality variants from the DNA and RNA genotypes, respectively. To prepare the final dataset used for GWAS, variants were concatenated. Imputation of the missing data in the Single Nucleotide Variant merged dataset was performed using Minimac (Pausch et al., 2017). Known variants detected in our study were annotated using NCBI SNP database (dbSNP version 150). The biological and molecular effects of the variants were predicted using SnpEff (Cingolani et al., 2012). Functional analysis (pathways and networks) was performed using BovineMINE (Elsik et al., 2016).

Results and Discussion

RNA-Seq and DNA-derived variants and their Enrichment in functional categories

For each cow, RNA library from uninfected macrophages and *ex vivo* MAP-infected macrophages collected at different post-infection (pi) time points (1 hpi, 4 hpi, 8 hpi, and 24 hpi) were sequenced. Overall, the whole primary macrophage transcriptome from 28 JD(–) and 22 JD(+) cows were analyzed. We identified 1,356,248 variants, including 542,080 unique to the DNA chip dataset (591,220 total SNP found in DNAchip) and 765,028 variants only detected by RNA-seq (814,168 total RNA-seq variants). Among them, 93,279 were indels. By comparing corresponding genotypes shared by two genotyping methods (e.g. RNA-seq and SNP50), we validated the accuracy of RNA-Seq variant calls. Among the variants identified, we identified 11.94% of novel variants which were all (161,951) from the RNA-seq data. To predict their effect (i.e. regulatory functions on known genes), the variants (SNP and indels) were annotated using SnpEff. The level of impact (high, low, or moderate effect) is described in Table 1.

Genome Wide Association Study (GWAS) and pathway analysis of the significant SNV

Genome-wide association analysis was performed using the 1,356,248 variants called from the imputed RNA-seq and SNP50 imputed to HD datasets from 28 JD(–) and 22 JD(+) cows, using a mixed linear model (MLM). Genome-wide association study identified two major eQTL loci on BTA4 and 11 at $-\log_{10}(P) \geq 7$. Network and pathway analysis using BovineMine from JD(+/–) macrophages revealed interesting cue regarding pathways that deserved further investigation (e.g. STAT transcription factor).

Table 1. Summary statistics of the identified variants using the respective methods.

Genotyping methods (counts)	RNA-seq	DNA chip
Variants processed		
SNP	720,889	591,220
Insertions	50,272	0
Deletions	43,007	0
Effects by impact[1]		
High	2,435	33
Low	28,359	5,109
Moderate	17,996	2,202
modifier	1,053,336	664,774

[1] Annotation and predicted effect of the genetic variants were performed using SnpEff

Conclusion

In the present study, we have identified functional biomarkers associated to JD that potentially explain the presence of putative dysfunctional allele(s) in these specialized phagocyte immune cells, named macrophages, which allow MAP to niche and proliferate. These functional SNP are part eQTL. Information on favorable or detrimental eQTL will eventually be used in genetic selection allowing beneficial allele(s) to disseminate in the progeny which would be able to efficiently prime host immune reaction. The originality of this research was to use RNA-Seq technology to provide a high-resolution genomic analysis of macrophages to identify novel mutations and transcripts associated with bovine paratuberculosis.

List of References

Ariel, O., N. Gévry, A. E. Ibeagha, G. Fecteau, and N. Bisseonnette. 2018. Mycobacterium avium ssp. paratuberculosis Infection Induces an Immune Tolerance Phenotype in Primary Bovine Macrophages from Johne's Disease Cows. In preparation.

Bannantine, J. P., M. E. Hines, 2nd, L. E. Bermudez, A. M. Talaat, S. Sreevatsan, J. R. Stabel, Y. F. Chang, P. M. Coussens, R. G. Barletta, W. C. Davis, D. M. Collins, Y. T. Grohn, and V. Kapur. 2014. A rational framework for evaluating the next generation of vaccines against Mycobacterium avium subspecies paratuberculosis. Frontiers in cellular and infection microbiology 4:126.

Brouard, J. S. and N. Bissonnette. 2018. The GATK Joint Genotyping for calling variants from RNA-Seq data. Anim. Genet. (submitted).

Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6(2):80-92.

Elsik, C. G., D. R. Unni, C. M. Diesh, A. Tayal, M. L. Emery, H. N. Nguyen, and D. E. Hagen. 2016. Bovine Genome Database: new tools for gleaning function from the Bos taurus genome. Nucleic Acids Res 44(D1):D834-839.

Fock-Chow-Tho, D., E. Topp, E. A. Ibeagha-Awemu, and N. Bissonnette. 2017. Comparison of DNA extraction kits and quantitative PCR systems for better sensitivity in detecting the causative agent of paratuberculosis in dairy cow fecal samples. J Dairy Sci 100(1):572-581.

Ghosh, P., D. C. Shippy, and A. M. Talaat. 2015. Superior protection elicited by live-attenuated vaccines in the murine model of paratuberculosis. Vaccine 33(51):7262-7270.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-2079.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing NGS data. Genome Res 20(9):1297-1303.

Pausch, H., I. M. MacLeod, R. Fries, R. Emmerling, P. J. Bowman, H. D. Daetwyler, and M. E. Goddard. 2017. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. Genet Sel Evol 49(1):24.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478.