J. Dairy Sci. 101:1–10
https://doi.org/10.3168/jds.2017-13848

Genome-wide association study and in silico functional analysis of the number of embryos produced by Holstein donors

C. Jaton,*†[1] F. S. Schenkel,* M. Sargolzaei,*† A. Cánova,* F. Malchiodi,* C. A. Price,‡ C. Baes,* and F. Miglior*§

*Centre for Genetic Improvement of Livestock (CGIL), Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada, N1G 2W1
†The Semex Alliance, Guelph, Ontario, Canada, N1G 3Z2
‡Faculté de Médecine Vétérinaire, Université de Montréal, St-Hyacinthe, Québec, Canada, J2S 2M2
§Canadian Dairy Network (CDN), Guelph, Ontario, Canada, N1K 1E5

ABSTRACT

Superovulation or ovum pick-up and in vitro fertilization are technologies used to produce an increased number of embryos from elite females. Embryo production traits have been shown to be heritable, but the genes that cause this variability have not yet been assessed. The main objectives of this study were to perform a genome-wide association study (GWAS) to find single nucleotide polymorphisms (SNP) associated with embryo production traits and to identify candidate genes affecting the number of embryos produced by Holstein donors in Canada that may provide insight into the regulation of embryo production. Breeding values were estimated and de-regressed for all donors and sires using a data set of 150,971 records of superovulation or ovum pick-up and in vitro fertilization. A total of 11,607 animals were genotyped, but of that number only 5,118 were genotyped with at least a 50K SNP panel and had a de-regressed estimated breeding value reliability of at least 10%. For the GWAS, 606,406 imputed SNP on 29 autosomal chromosomes were considered after applying quality control measures. A single-SNP univariate mixed linear animal model was used to perform the GWAS, and a 5% false discovery rate was applied to adjust for multiple testing. We found 36 and 14 significant SNP associated with the total number of embryos and the number of viable embryos, respectively, with most of them located on chromosome 11. Using these significant SNP, positional genes located within 10,000 bp upstream and downstream of the SNP were retrieved. Thirteen genes were harboring or near the significant SNP for the total number of embryos, 4 of them also being near the significant SNP for viable embryos. Some of these genes (*CRB2*, *DENND1A*, *MAD1L1*, *NDUFA8*, *PTGS1*) could be considered as potential positional candidate genes related to the number of embryos produced by a donor. This list will need to be validated in an independent population to confirm the role of the genes for embryo production.

Key words: GWAS, candidate gene, embryo production, Holstein

INTRODUCTION

Assisted reproductive technologies, such as superovulation and ovum pick-up and in vitro production (**OPU-IVP**) of embryos, are frequently used in the Canadian dairy industry to produce more offspring from elite donor cows. Although technical improvement of these procedures has been the subject of much research (Mapletoft and Bó, 2011), there is evidence that genetic improvements may also be possible. The number of embryos produced by a donor is moderately heritable in Holstein dairy cattle (Jaton et al., 2016a; Cornelissen et al., 2017; Parker Gaddis et al., 2017), so that it could be possible to genetically select donors that would produce more embryos. Moreover, finding genes that affect embryo production traits may also help with the selection of donors that respond well to superovulation or OPU-IVP. Genome-wide association studies (**GWAS**) have been performed for fertility traits in dairy (Kolbehdari et al., 2008; Huang et al., 2010; Berry et al., 2012; Cochran et al., 2013; Minozzi et al., 2013; Parker Gaddis et al., 2016) and beef cattle (McDaneld et al., 2014); however, very few studies have performed a GWAS for traits related to embryo production using a large data set. Some studies have used a candidate gene approach to investigate the association of specific genes and embryo production traits in Holstein donors. Some significant SNP on the follicle-stimulating hormone receptor (*FSHR*) gene were found to be associated with the total number of ova and the number of

Received September 18, 2017.
Accepted April 5, 2018.
[1] Corresponding author: cjaton@semex.com; cjaton@uoguelph.ca

transferable embryos produced by donors (Yang et al., 2010). Another study indicated that some SNP in the *FSHR* gene could help to identify donors that produce many or few embryos (Cory et al., 2013). Significant SNP on the gonadotropin-releasing hormone receptor (*GNRHR*) gene were also found to be significantly associated with the number of degenerated embryos, the number of transferable embryos, and the total number of ova produced by Holstein donors (Yang et al., 2011). One SNP located in the insulin-like growth factor 1 receptor (*IGFR1*) gene was significantly associated with the total number of ova (Yang et al., 2013). Similarly, 1 SNP on the glutamate receptor AMPA 1 (*GRIA1*) gene was reported to influence the number of ova and embryos collected in superovulated donors (Sugimoto et al., 2010). However, in all cases, few donors were used, which constitutes an important limitation.

The main objectives of our study were to perform a GWAS to find SNP associated with embryo production traits using a large data set of 5,118 genotyped animals and to identify potential key regulator genes affecting the number of embryos produced and to retrieve the biological pathways and mechanisms linked to the number of embryos produced. This study should provide insight into the regulation of embryo production.

MATERIALS AND METHODS

Data

A data set containing all superovulation or OPU-IVP procedures performed in Canada over the last 35 years was provided by Holstein Canada (Brantford, ON; www.holstein.ca). This data set was used in our previous studies on genetic analysis of superovulatory response of Holstein cows (Jaton et al., 2016a,b) and the same edits were performed. After editing, 150,971 records from 59,586 donors were considered for the analysis, with 1 record corresponding to 1 superovulation or OPU-IVP procedure performed on 1 donor.

Traits. The 2 embryo production traits that were analyzed from the data set provided were the total number of embryos (**NE**) and the number of viable embryos (**VE**) per procedure. The difference between these traits is that VE does not include degenerated or dead embryos recovered.

Pedigree. An animal pedigree file containing 32,587 animals was generated by tracing the pedigrees of the genotyped donors back to 1950.

EBV. Genetic parameters and EBV for both embryo traits were estimated from univariate analyses, as described by Jaton et al. (2016a). Breeding values from natural logarithmic transformation were considered for the GWAS. Overall, 63,482 donors and their sires had EBV available for both traits. The EBV were de-regressed using a simplified method for de-regression (VanRaden and Sullivan, 2010). Animals with a reliability of de-regressed EBV lower than 10% for both traits were not considered for further analysis.

Genotypes and Imputation. Genotypes were available for 11,607 donors and sires that had de-regressed EBV. Of that number, 6,734 animals were genotyped with at least a 50K SNP panel, and these genotypes were imputed to high-density genotypes using the FImpute software (Sargolzaei et al., 2014) and a reference population of 2,653 high-density genotypes. After accounting for the reliability threshold, 5,118 individuals (905 males and 4,213 females) were considered for further analyses of the NE and the VE, respectively.

GWAS

Quality Control. Quality control analysis was performed including the exclusion of SNP having a minor allele frequency lower than 1%, a call rate lower than 90%, an excess of heterozygosity higher than 15%, and Mendelian error frequency larger than 5%. The SNP that were out of Hardy-Weinberg equilibrium with very low probability (1×10^{-8}) and individuals with a call rate lower than 90% were also excluded. Parentage verification was performed on genotyped parent-progeny pairs, and we observed no parentage conflicts. Overall, 734,077 SNP on 29 autosomal chromosomes were considered for the association analysis before applying quality control measures.

Method. A single SNP univariate mixed linear animal model was used to perform the GWAS using the snp1101 software version 1.0 (Sargolzaei, 2014). Considering the large variation in the reliabilities of the de-regressed EBV, the mixed model equations were weighted by reliabilities. To account for population structure, a random animal effect was fit in the model with (co)variance structure based on the genomic relationship matrix (**G**). The model was

$$\mathbf{y} = \mu + \mathbf{Wb} + \mathbf{Zg} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of de-regressed EBV for either NE or VE; μ is the overall mean; $\mathbf{b}$ is a vector of fixed effects including the SNP effect; $\mathbf{g}$ is a vector of random animal effects; $\mathbf{e}$ is a vector of random residuals; and $\mathbf{W}$ and $\mathbf{Z}$ are the corresponding incidence matrices relating $\mathbf{y}$ to $\mathbf{b}$ and $\mathbf{g}$, respectively.

Random effects were assumed to be normally distributed, with means equal to zero and covariance structure equal to