GENOMICS OF HEIFER PREGNANCY, DAYS OPEN, AND DAYS TO CONCEPTION IN RED ANGUS HEIFERS

A Thesis
Presented to
The Faculty of the Graduate School
At the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for Degree
Master of Science

By
Esdras Tuyishimire
Dr. Jared Decker, Thesis Supervisor
May 2022
The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

**GENOMICS OF HEIFER PREGNANCY, DAYS OPEN, AND DAYS TO CONCEPTION IN RED ANGUS HEIFERS**

presented by **Esdras Tuyishimire**, a candidate for the degree of **Master of Science**, and hereby certify that, in their opinion, it is worthy of acceptance.

________________________________________

**Dr. Jared Decker**

________________________________________

**Dr. Robert Schnabel**

________________________________________

**Dr. Elizabeth King**

________________________________________

**Dr. Ruthie Angelovici**
ACKNOWLEDGMENTS

I would like to thank my committee members for their unwavering support and comments during the analysis and writing process. Also, I would thank Dr. Decker and Dr. Schnabel’s lab members for providing a good working environment and their comments that helped me improve every process that I was taking.

Many thanks to my family; my mom, dad, brothers, and sisters always believed in me and supported my decisions. In addition, thanks to my roommate and friends who encouraged me to reach my academic potential.

Finally, I would like to extend my special thanks to the Red Angus Association of America for providing data.

This project was supported by Agriculture Food Research Initiative Competitive Grant no. 2020-67015-31132 from the USDA National Institute of Food and Agriculture.
## Table of Contents

ACKNOWLEDGMENTS .................................................................................................................. II

*List of Tables* ................................................................................................................................V

*List of Figures* ............................................................................................................................. VI

0. ABSTRACT .................................................................................................................................... VII

CHAPTER 1 LITERATURE REVIEW ................................................................................................. 1

1.1. THE CONTRAST BETWEEN PRODUCTION AND FERTILITY TRAITS ......................... 1

1.2. IMPORTANCE OF FERTILITY TRAITS ............................................................................. 4

1.3. PREVIOUSLY STUDIED REPRODUCTIVE TRAITS AT THE GENETIC LEVEL. ............ 5

CHAPTER 2 GENOMICS OF HEIFER PREGNANCY, DAYS OPEN, AND DAYS TO CONCEPTION IN RED ANGUS HEIFERS ............................................................................................................ 9

2.1. INTRODUCTION .................................................................................................................. 9

2.2. MATERIALS AND METHODS .......................................................................................... 11

2.3. RESULTS .......................................................................................................................... 19

2.4. DISCUSSION ...................................................................................................................... 33

2.5. CONCLUSION ..................................................................................................................... 38

2.6. REFERENCE: .................................................................................................................... 40

2.7. APPENDIX .......................................................................................................................... 46
List of Tables

Table 1 Economic fertility traits and their respective indicator traits .......................................................... 6

Table 2 Summary statistics for Heifer Pregnancy, Days to Conception, Days Open, Heifer's Age (in days), and Season Length (in days). n stands for sample size, and SD is a standard deviation. ........................................................................................................................................ 13

Table 3 Estimates of genetic variance $\sigma_g^2$, residual variance $\sigma_e^2$, and heritability $h^2$ from univariate GCTA and GEMMA .................................................................................................................. 20

Table 4. Heifer’s age and season length estimated effect from linear and Gibbs sampling models on Heifer Pregnancy, Days to Conception, and Days Open .......................................................................................... 22

Table 5 The number of samples, mean, standard deviation, high posterior density (HPD), autocorrelation, and heritability ($h^2$) estimates from the posterior distribution of the univariate threshold model ........................................................................................................... 23

Table 6 Reliabilities for heifer pregnancy (HP), days to conception, and days open on genotypes heifer (4097 individuals). SD is the standard deviation ............................................................................................................. 31

Table 7 Gene terms at Benjamin-Hochberg correction p-value < 0.05 in association with HP and DO .......................................................................................................................... 36
List of Figures

Figure 1 Genetic trends in growth (a), carcass (b), and reproduction (c) traits in Red Angus cattle [18].................................................................................................................................................. 3

Figure 2 Distribution of days open and days to conception after data filtering. Most heifers were pregnant by 100 days, while very few heifers went beyond that (DC histogram)....................... 12

Figure 3 Correlation between estimated days open (DO) and days to conception (DC) breeding values (EBVs)......................................................................................................................... 22

Figure 4 Trace and histogram plots for (a) Heifer pregnancy, (b) days to conception, (c) days open that show the convergence in threshold models using Gibb’s sampling. The red line and their corresponding value show the mean posterior variance and heritability for trace plots and histograms. ...................................................................................................................................................... 26

Figure 5 The relationship between full and partial breeding values in the validation set of (a) heifer pregnancy, (b) days to conception, and (c) days open.............................................................................. 27

Figure 6 Manhattan plot and QQ plot for (a) GCTA and (b) GEMMA univariate GWAS in HP. The red line in Manhattan plot shows a cutoff level at p-value of 1e-05. .............................................. 29

Figure 7 Manhattan plot and QQ plot for GEMMA bivariate GWAS in HP and DO. The red line shows the genome-wide significant level (p-value = 5e-08) while the blue line shows the suggestive significant level (p-value = 1e-05). .......................................................................................................................... 30

Figure 8 Manhattan QQ plots for (a) Heifer Pregnancy, (b) Days to Conception, and (c) Days Open based on de-regressed breeding values in the SNP1101 program. ........................................... 32
0. Abstract

Although it is challenging to collect information on reproductive traits, fertility traits are believed to play a significant role in the growth of the beef industry and bring revenue to the producers. It has been challenging to make genetic improvement of fertility traits because of scarce records and their low heritability. The joint analyses of economic and indicator traits provided much information for the analysis of other production traits, resulting in favorable improvement. The same method can be applied to reproductive traits to increase prediction accuracy. Furthermore, the ubiquity of high-throughput genotypes and the development of advanced computational methods give hope that there is a possibility of determining the genetic influence on fertility traits. The advancement in phenotypic collection technology eases the keeping of records on both economic and indicator fertility traits. Also, it allows the development of new traits to be analyzed while reducing the cost of data availability.

Heifer Pregnancy (HP) - the ability for a heifer to conceive by the end of the breeding season; Days Open (DO) - days of breeding season a heifer remained open; and Days to Conception (DC) – the number of days it took for a heifer to get pregnant, are easy to measure fertility traits with enough information to make beef reproductive genetic improvement in the industry. The objective of the current study is to identify DNA markers tagging genes influencing HP, DC, and DC among 18,039 Red Angus heifers, find the genetic relationship among those traits, and creating the genomic predictions of days to conception and days open.

The results show a small heritability of 0.119 to 0.131, 0.10 to 0.102, and 0.0749 to 0.112 for HP, DO, and DC, respectively. The DC model resulted in higher accuracy than other models. There was a high correlation between HP and DO ($r = -0.61$ and $0.85$ from the linear model and the
liability scale model, respectively). The de-regressed estimated breeding value (EBV) genome-wide association (GWAS) yielded 58 and 2 significant SNPs at suggestive significant level (p-value < 1.0e-05) for HP and DC, respectively. This study found GRID2 and ZMIZ1 genes to be associated with heifer pregnancy, and it has been speculated that the central nervous system related genes ontology and the hormones it controls might suggest the physiology behind some of the reproductive differences.

The present study confirms the low heritability status of fertility traits and shows the possibilities of genetic enhancement based on the obtained accuracies. The identified genes and gene terms will serve as starting points for future studies that might focus on different phenotypes while reducing the cost of phenotyping and improving the accuracy of fertility traits genomic predictions.
Chapter 1 Literature Review

1.1. The contrast between production and fertility traits

Since the introduction of *Bos taurus* in the US around the 16th century [1, 2], livestock producers were already applying artificial selection focusing on visual features. The involvement of academic researchers and government agencies in the 20th century revolutionized the industry as they put much effort into increasing productivity [3–5]. The early work focused on the growth traits like weaning weight and yearling weight, then on carcass traits like carcass marbling score [6–9]. The records on growth traits are abundant because they are quite easy to measure, and their quantitative nature facilitates the analysis, hence enabling genetic progress. The more the beef industry grew, the researchers were able to juxtapose the indicator traits in association with some economic traits. It has been possible to improve some of the hard-to-measure traits because additional information would come from indicator traits [10]. Examples of indicator traits include birth weight, which is associated with weaning weight, yearling weight, and calving ease; ultrasound measures give the status for carcass marbling score.

The application of statistical analysis and genetics allowed to estimate some of the parameters that breeders based on for selection decisions of production traits; as a result, there was great progress in beef growth and carcass traits (Figure 1a, b). When a high-density SNP array was introduced [11], the industry changed forever as some traits like weaning weight, yearling weight, carcass weight, and back fat could be studied at the genomic level [12, 13], hence the increased accuracy of prediction [14]. It did not take very long to realize that the reproductive traits were left behind while analyzing production and carcass characteristics. Primarily in the dairy industry, they identified that increased production impaired fertility [15, 16]. Wiltbank, 1970 called for research in beef cattle reproduction and identified the loss of calf crop production and the increased calving
season as the primary concerns; thus, he posed some questions that the scientific community would answer to solve the problem [17]. The Wiltbank and other scientists’ call allowed the transition to add the fertility traits among the economic traits that should be studied.

While many factors were identified to affect fertility traits, people in different domains tried to have input. They developed some techniques, including artificial insemination, embryo transfer, improved management practices, and selection on fertility; consequently, the currently observed improvement (Figure 1c), yet the problem remains not fully solved. There is still a need for further studies to understand the genes governing reproduction in beef heifers.
Figure 1 Genetic trends in growth (a), carcass (b), and reproduction (c) traits in Red Angus cattle [18].
From the 1980s to 2004, the growth and carcass traits increased while the fertility traits except stayability decreased. The improvement in fertility traits after 2004 might indicate researchers’ effort after they started to include those traits in selection decisions, although that selection is in Red Angus and a very few breeds, as it has been indicated by Rowan et al. 2021 [19].

1.2. Importance of fertility traits

Although reproductive traits are notorious for being lowlily heritable [17, 20], which restricts quick genetic progress, they are pivotal in sustaining the animal industry. It might be very hard to ignore the importance of fertility in the beef industry, especially for those in a cow-calf operation. In fact, the whole industry is based on calf production; therefore, it is crucial to ensure that a heifer is and remains pregnant until it calves a healthy calf. Some of the fertility traits that have been identified to be essential to the industry are heifer pregnancy, stayability, and calving ease [21–23]. On the other side, to have successful fertility, puberty onset should happen; thus, puberty-related traits are considered fertility indicators. Not only that a heifer that reaches puberty early in the breeding season weans a heavier calf, but in their study, Roberts et al. found that there was an 11 to 13 % less calving rate in the first 21 days among the heifers that failed to reach puberty during the breeding season than their contemporaries [24]. Therefore, interest is made for every pound increased on a weaned calf whose dam reached puberty early in the breeding season. In addition, while there is a cost related to raising and maintaining a heifer, culling comes with a loss to the producers. Fertility problems are among the top reasons for culling in the livestock industry [25]. In the dairy industry, the increased insemination rate, linked to long days open and pregnancy rate, reduced the profit by $205 per year per cow in 2004 [26]. Also, in 2008, it was shown that in Florida, $75 was lost per failed pregnancy during the breeding season [27], and that loss might
even be higher if adjusted for inflation. Despite the increase in technology, reproductive failure and loss of embryo issues are still around, and an estimated loss of $2.8 billion in the US beef industry per year might happen if nothing is done [28]. Given the identified failures and what could be gained if a heifer gets pregnant early in the breeding season, it is essential to study the fertility traits further to increase the number of calves born alive and reduce the cost related to reproductive failure.

1.3. Previously studied reproductive traits at the genetic level.

Burns et al. reviewed factors that affect reproductive performance in beef cattle, and they classified them into four main categories: management, environmental, physiological, and genetics [29]. Nowadays, technology and scientific research growth helped improve management, environmental, and physiological conditions. At the same time, genetic improvement remains a challenge, and little is known about the genetics underlying reproductive traits. Most fertility traits are hard to measure, and due to their binary nature, it is also challenging to make accurate predictions because of their low heritability.

After realizing a decrease in fertility, scientists put effort into improving these traits; hence, there are some noticeable improvements. Some of the studied traits in the beef industry include heifer pregnancy, stability, calving ease, age at puberty, first pregnancy rate, and to less extent, days open [20, 21, 23]. As suggested, the economic traits might be analyzed jointly with indicator traits (Table 1) to provide enough information, but a very few studies performed those joint analyses.
Table 1 Economic fertility traits and their respective indicator traits

<table>
<thead>
<tr>
<th>Economic Traits</th>
<th>Indicator Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calving ease</td>
<td>Birth weight</td>
</tr>
<tr>
<td></td>
<td>Gestation length</td>
</tr>
<tr>
<td>Stayability</td>
<td>Days to calving</td>
</tr>
<tr>
<td></td>
<td>Days open</td>
</tr>
<tr>
<td></td>
<td>Day to conception</td>
</tr>
<tr>
<td></td>
<td>Calving records</td>
</tr>
<tr>
<td>Heifer pregnancy</td>
<td>Age at puberty</td>
</tr>
<tr>
<td></td>
<td>Follicular diameter</td>
</tr>
<tr>
<td></td>
<td>Antral follicle count</td>
</tr>
<tr>
<td></td>
<td>Postpartum interval</td>
</tr>
<tr>
<td></td>
<td>Scrotal circumference</td>
</tr>
<tr>
<td></td>
<td>Reproductive tract score</td>
</tr>
</tbody>
</table>

Due to the complexity of fertility traits, it is challenging to perform across breed predictions [5, 14]; consequently, because of its various breeds, the beef industry runs behind dairy. On the other hand, there are some accomplished within breed research on HP, stayability, calving ease, days open, and scrotal circumference [30–37]. In those studies, the reported heritability was 0.085 to 0.27; 0.059 to 0.57; 0.19; 0.091 to 0.192; and 0.41 to 0.71 for HP, stayability, calving ease, days open, and scrotal circumference, respectively. Speidel et al. found some quantitative traits loci
(QTL) that are presumed to be associated with heifer pregnancy and stayability [38]. Some additional fertility traits that have been studied include antral follicular count (AFC) and reproductive tract score (RTS), with a sample size of 293 heifers [39]. Even though all those studies were able to find some loci and candidate genes associated with reproduction, the researchers concluded that validation studies were needed as they were not sure about the results. In addition to uncertain results, those studies used a small sample size which might not be reliable when studying complex heritable traits, and the genomic information was from low density markers, while high densities with much genomic information are currently available.

Even though their higher heritability can explain the success made in production and carcass than some reproductive traits, the quantitative nature, and the joint analyses included indicator traits added advantages to that success. For instance, the birth weight would be regarded as an indicator trait for weaning and yearling weight, and ultrasound measures for intramuscular fat, marbling score, and ribeye area; thus, they were usually analyzed together, which provides enough information for statistical analysis, hence improved accuracy [10, 13, 40, 41].

Despite the small improvement in HP and Stayability in Red Angus since 2012, there is still room for progress given the advancement in statistical and genomic tools that increase prediction accuracy [42]. The ubiquity of sequencing technology coupled with the development of advanced computational tools in the 21st century gives hope that there is a possibility of determining the genetic influence of fertility traits. The present studies should perform analyses that blend both pedigree and genomic information to increase fertility traits' accuracy while scrutinizing economic and indicator traits to maximize available information. Cushman et al. provide a list of some fertility traits and their corresponding indicators as some of them are presented in table 1 [43], but most of those traits are expensive to collect and require much work.
While the overall goal in the beef industry is to maximize profit while reducing production costs, the increase in phenotyping technology would help select the most relevant and introduce new traits for the analysis, which does not cost much and with less intensive work.
Chapter 2 Genomics of Heifer pregnancy, days open, and days to conception in Red Angus heifers

2.1. Introduction

Despite the previously reported data scarcity and low heritability status of reproductive traits that hinder the genetic progress in beef cattle [44–46], these traits are believed to have a pivotal role in the development of the beef industry. Mercadante et al. predicted a loss of $21 billion of revenue due to the failure to produce a live calf at weaning and the cost of breeding [28]. Morrey and Biased predicted that the beef industry would lose over $210 million due to late breeding heifers [47] as it is phenotypically related to cows continued reproductive success (i.e., stayability). Given that loss, much effort is required to improve reproductive traits. Previous studies have found reproductive problems associated with management, environment, physiology, and genetics [20, 29, 48, 49]. Some of those have been addressed; consequently, there has been improvement in reproductive traits among beef cattle. Yet, the genetic part of the reproductive traits is not fully understood. In the past, researchers made conclusions about fertility traits on heritability estimates only without any consideration of genetic effect and response to selection which led them to conclude that genetic improvement was almost impossible for reproductive traits [17, 50].

With the advancement in computational and genomic technology, many studies were performed to predict some genetic parameters but very few incorporated genomic data with a small sample size. Cammack et al. reviewed most of the studies performed using only pedigree information [20], and most of those studies resulted in low heritability as expected for fertility traits [21, 23, 31, 37]. The reduced cost of genome sequencing and SNP array development [11] helped scientists collect genotypic information and analyze some of the reproductive traits at the genomic level. A couple of studies have used genomic information to estimate *Bos taurus* beef
reproductive traits heritability and find the associated single nucleotide polymorphism (SNP) and candidate genes [38, 51–54]. Nevertheless, the sample size and the number of SNPs in those studies were small. Reproductive traits are polygenic and require a significant sample size and high-density markers [52] to find the true SNPs in association and the genes that have an impact on the traits. Evans et al. 1999 found that the accuracy of heifer pregnancy would increase when analyzed with scrotal circumference [21], and Toghiani found that multi-trait genomic evaluation would elaborate on the accuracy of selection of fertility traits when both quantitative and discrete variables are analyzed jointly [52].

Not only a large dataset is needed to make accurate predictions on reproductive traits, but also high-quality phenotypes that are aligned with genotype are required [48]. Imputation allows inferring from low to high marker density with a large reference population of genotyped animals, which increases the ability to detect variants associated with the phenotype [55, 56]. This study aims to use high-density markers to find genes influencing fertility traits. Heifer Pregnancy (HP) – the ability for a heifer to conceive by the end of the breeding season and remain pregnant, Days Open (DO) – days of a breeding season a heifer remained open; and Days to Conception (DC) – numbers of days it took for a heifer to get pregnant, are promising, easy to measure fertility traits. We hypothesize these traits have sufficient information to improve the genetic prediction of reproduction in beef cattle, especially when analyzed jointly.

Eler et al. mentioned that it is irrelevant to use estimates from indicator traits while direct reproductive measures are available [57]. But in some circumstances, estimates from fertility traits are not available because they demand intensive labor with a high cost of data collection; as a result, the indicator traits might be used for indirect prediction of fertility traits. The first objective of the current study is to estimate the variance components of HP, DO, and DC using three software
suites by fitting a linear and threshold model for a binary trait (HP). The second objective is to estimate the accuracy of genomic prediction for each trait and identify DNA markers associated with the observed phenotypes and the genetic relationship among the traits. It has been presumed that genes in common influence both heifer pregnancy and days intervals (DO and DC).

2.2. Materials and Methods

2.2.1. Phenotypes and data quality control

The Red Angus Association of America provided phenotypes, genotypes, and pedigree records. Information was provided for 4,097 genotyped heifers with heifer pregnancy records, as well as 13,942 that were in the same contemporary groups as the genotyped heifers. The full phenotype dataset had 18,485 individuals born between 1988 to 2018. There were records on heifer pregnancy, start and end dates of the breeding season, contemporary group, calf birthdate, and sex. Animals in a contemporary group share information on the farm, calving season, and birth year. Heifer Pregnancy (HP) was recorded as 1 for pregnant and 0 for non-pregnant. In contrast, Days Open (DO) and Days to Conception (DC) were calculated as continuous variables. Both DO and DC were calculated based on the mean gestation length of 283 days, as has reported in [58]. Thus, DC was expressed as the difference in calf birth date, conception length, and the start of the breeding season; $DC = Calf\, birth\, date - conception\, length - start\, of\, the\, breeding\, season$. The same applies for DO, but heifers who did not become pregnant by the end of the breeding season were assigned the season length as their Days Open value. The season length was calculated as the difference between the end and the start of the breeding season. Appendix 2.6.1. shows distribution of quantitative variables before quality control.
While it is possible to have some noise during data collection, it was necessary to perform quality control on the phenotype data. The phenotype quality control was done by removing all individuals who were 300 days old or younger at the start of the breeding season. Heifers whose DO records were more than 365 or less than -30 days were removed from the dataset. Again, heifers with more than 250 or less than -21 days on DC and DO were considered missing (only DO and DC) based on the data distribution (Figure 2). Heifers are typically bred at 15 months of age[59–62]; thus, 300 days of age was set as cutoff (Figure 3 (c) in Appendix 2.6.2.). While some of the tools used during this study consider -9 as missing, DC and DO with -9 days were assigned -9.1 days to avoid that problem.

![Days Open Histogram](image1)

![Days to Conception Histogram](image2)

Figure 2 Distribution of days open and days to conception after data filtering. Most heifers were pregnant by 100 days, while very few heifers went beyond that (DC histogram).

Days open include heifers who reached the end of the breeding season without becoming pregnant, and most of those heifers are between 40 days and 150 days (DO histogram), indicating that most of the herds followed 90 days breeding season length.

After data quality control, 18,039 individuals remained for the analysis. Below is a summary table for heifers with variable’s number of records, minimum, maximum, mean, and standard deviation (Table 2).
Table 2 Summary statistics for Heifer Pregnancy, Days to Conception, Days Open, Heifer's Age (in days), and Season Length (in days). n stands for sample size, and SD is a standard deviation.

<table>
<thead>
<tr>
<th>Trait</th>
<th>n</th>
<th>minimum</th>
<th>maximum</th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heifer Pregnancy</td>
<td>18,039</td>
<td>0</td>
<td>1</td>
<td>0.8</td>
<td>0.39</td>
</tr>
<tr>
<td>Days to Conception</td>
<td>14,231</td>
<td>-21</td>
<td>249</td>
<td>18</td>
<td>33</td>
</tr>
<tr>
<td>Days Open</td>
<td>17,938</td>
<td>-21</td>
<td>249</td>
<td>31</td>
<td>42</td>
</tr>
<tr>
<td>Age</td>
<td>17,949</td>
<td>301</td>
<td>546</td>
<td>422</td>
<td>30</td>
</tr>
<tr>
<td>Season Length</td>
<td>17,763</td>
<td>0</td>
<td>244</td>
<td>86</td>
<td>37</td>
</tr>
</tbody>
</table>

2.2.2. Genotypes

The genotype information on the individuals used in this study came from the Red Angus Association of America. The ARS-UCD1.2 bovine reference genome was used to identify the SNP positions [63]. Heifers were genotyped using a variety of arrays used for genomic-enhanced EPDs. The genotypes passed through the University of Missouri multibreed imputation pipeline described by Rowan et al. [56]. Rowan et al. performed quality control on autosomal variants only using PLINK software [64] and removed individuals and variants with less than 90% call rate [56]. Among the Red Angus, 811,679 SNPs were included in the imputed data. Apart from the initial quality control done during imputation, SNPs with MAF < 0.01 were removed from this study.
2.2.3. Variance components estimation, and genome-wide association with Genome-wide Complex Trait Analysis (GCTA)

The GCTA software (v1.93.3beta2) was used to estimate the variance components of HP (binary), DO, and DC. Both univariate and bivariate linear mixed model analyses were performed for variance estimation. Only genotyped heifers (n = 4,097) were analyzed with GCTA [65]. Among those heifers, the number of analyzed individuals varied from trait to trait as the missing phenotypes were filtered out. Each model fitted the contemporary group and Genomic Relationship Matrix (GRM) to have a random effect. Contemporary group was fit as random due to the small number of animals within many contemporary groups [66]. In these models, heifer’s age and breeding season length were fit as fixed effects (Equation 1). The genomic relationship matrix (GRM) was calculated using autosomal SNPs, with the method of Yang et al [65]. The variance components were calculated by genome-based restricted maximum likelihood using the average information [65]. The mathematical representation of the univariate model to estimate variance and heritability is in equation 1. In addition, the estimated breeding values were calculated using the same equation. The heifer pregnancy heritability was adjusted on the liability scale considering the prevalence of 81.3% (The value obtained as the pregnancy rate among 18,039 heifers). Equation 2 represents the bivariate model for genetic correlation between HP and DO.

\[ Y = Hh + Ss + Cp + Zu + e \] (Equation 1)

Where \( Y \) is a vector of the phenotype of interest (can take the values of HP, DO, or DC)

\( h \) is the vector of heifer’s age fixed effect

\( s \) is the vector of season length fixed effect

\( p \) is the random effect of the contemporary group with \( p \sim N(0, \sigma_p^2) \)
\( u \) is the random additive genetic effect with \( u \sim N(0, A\sigma_u^2) \).

\( e \) is the random residual effect with \( e \sim N(0, I\sigma_e^2) \).

\( H, S, C, \) and \( Z \) are the incidence matrices relating traits observation in \( Y \) to the vectors \( h, s, p, \) and \( u \), respectively.

\( A \) and \( I \) being genetic relationship and identity matrices, respectively, among the individuals.

Only HP and DO were used for the bivariate model since there was no record of DC for non-pregnant heifers, leading to convergence problems. Also, DO and DC are highly correlated; thus, there was no need to fit them together. Equation 2 shows the mathematical representation of the bivariate model.

\[
\begin{bmatrix}
Y_1 \\
Y_2
\end{bmatrix} = 
\begin{bmatrix}
H_1 & 0 \\
0 & H_2
\end{bmatrix}
\begin{bmatrix}
h_1 \\
h_2
\end{bmatrix} + 
\begin{bmatrix}
S_1 & 0 \\
0 & S_2
\end{bmatrix}
\begin{bmatrix}
s_1 \\
s_2
\end{bmatrix} + 
\begin{bmatrix}
C_1 & 0 \\
0 & C_2
\end{bmatrix}
\begin{bmatrix}
p_1 \\
p_2
\end{bmatrix} + 
\begin{bmatrix}
Z_1 & 0 \\
0 & Z_2
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix} + 
\begin{bmatrix}
e_1 \\
e_2
\end{bmatrix}
\]

(Equation 2)

All letters are as described in the first equation. In contrast, numbers 1 and 2 in the second equation represent heifer pregnancy and days open, respectively.

The random genetic effects \( \begin{bmatrix}
u_1 \\
u_2
\end{bmatrix} \sim N(0, A \otimes V) \) in which \( V = \begin{bmatrix}
\sigma_1^2 & \sigma_1\sigma_2 \\
\sigma_1\sigma_2 & \sigma_2^2
\end{bmatrix} \).

\( V \) being the genetic variance-covariance matrix, \( \sigma_1^2 \) is the genetic variance for HP, \( \sigma_2^2 \) is the genetic variance for DO, and \( \sigma_1\sigma_2 \) is the genetic covariance of HP and DO.

Also, the random contemporary group effect \( \begin{bmatrix}
p_1 \\
p_2
\end{bmatrix} \sim N(0, I \otimes P) \) in which \( P = \begin{bmatrix}
p_1^2 & p_1p_2 \\
p_1p_2 & p_2^2
\end{bmatrix} \).
\( P \) is the contemporary group variance-covariance matrix; \( p_1^2 \) is the contemporary group variance for HP; \( p_2^2 \) is the contemporary group variance for DO, and \( p_1p_2 \) is the contemporary group covariance of HP and DO.

Finally, the random residuals effect \( [e_1, e_2] \sim N(0, I \otimes E) \) in which \( E = \begin{bmatrix} e_1^2 & e_1e_2 \\ e_1e_2 & e_2^2 \end{bmatrix} \),

\( E \) is the residual variance-covariance matrix, \( e_1^2 \) is the residual variance for HP, \( e_2^2 \) is the residual variance for DO, and \( e_1e_2 \) is the residual covariance of HP and DO.

The estimation of phenotypic variance explained by all SNPs, the so-called heritability, was calculated as \( h^2 = \frac{\sigma^2}{\sigma^2 + e^2} \).

\( h^2 \) is the heritability, \( \sigma^2 \) genetic variance for a given trait, and \( e^2 \) residual variance.

The genetic correlation \( (\sigma_g) \) was obtained as \( \sigma_g = \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2\sigma_2^2}} \).

The ‘-mlma’ option within GCTA was used to perform univariate GWAS for Heifer Pregnancy, Days Open, and Days to Conception, using the same univariate model described for genomic REML analysis, except each SNP is independently fit as a fixed effect.

2.2.4. **Analysis with Genome-wide Efficient Mixed-Model Association (GEMMA) using pre-corrected phenotypes**

GEMMA was also used for genome-wide association analysis, due to the ability of GEMMA to use linear mixed models to control the population structure and relatedness [67] coupled with increased statistical power from multi-trait models [68]. However, there had been previously difficulty getting GEMMA to run with large numbers of fixed effects. Consequently, the analyses in GEMMA were performed using pre-corrected phenotypes from GCTA. The
adjusted phenotypes were the sum of individual estimated breeding values and corresponding residual. Because the phenotypes were pre-corrected, the model fit in GEMMA contained the phenotypes and genomic relationship matrix, and its mathematical representation is as follows:

\[ Y = \mu + Zu + e \quad (equation.3) \]

\[ h^2 = \frac{\sigma^2}{\sigma^2 + e^2} \quad Equation.4 \]

And the multivariate equation is as,

\[
\begin{bmatrix}
Y_1 \\
Y_2
\end{bmatrix} =
\begin{bmatrix}
Z_1 & 0 \\
0 & Z_2
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix} +
\begin{bmatrix}
e_1 \\
e_2
\end{bmatrix} \quad (Equation.5)
\]

Where \( \mu \) is the mean, and other letters and distribution are as described in previous equations.

**2.2.5. Variance Components and Genomic predictions with Gibbs sampling**

Due to the binary nature of heifer pregnancy, the ideal way of estimating variance and heritability would be to consider the liability scale [69]. The THRGIBBS1F90 [70, 71] is one of the BLUP90 programs [72] that can handle both continuous and binary variables; thus, it has been used to generate Gibbs samples, and in turn, those samples were used by the POSTGIBBS1F90 program to estimate the posterior variances and breeding values. The threshold model (THRGIBBS1F90) was run for 1,000,000 MCMC samples and samples were saved at every 200 iterations for HP and 50 for DO and DC. The initial burn-in was 100,000 samples. Because the BLUPF90 suite of programs accepts both genotyped and non-genotyped individuals, all heifers with phenotype information were used for variance component estimation and genomic predictions. For every analysis, the sample and pedigree sizes (3 generations) changed as the individuals with missing phenotypes were not included in the study. The HP univariate and multivariate models had the same sample size of 18,039 heifers, and the pedigree contained 47,050
individuals. The univariate DO model had 17,938 heifers and 46,919 individuals within the pedigree. On the other hand, the DC model among others, had the smallest sample and pedigree size of 14,231 and 40,473 individuals.

The number of individuals within the contemporary group has increased due to the expanded sample size. Hence, the contemporary group was fit as a fixed effect (rather than random as in the GCTA models). Quality control and relationship matrices were performed using the PREGSF90 program[73]. Analyses used the inverse of the H matrix (H⁻¹), which was built as a blended matrix of genomic (G) and pedigree relationship matrices as described in Aguilar et al. [74–76]. In addition, the genomic relationship matrix was produced following the method described by VanRaden [77].

\[
H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \text{(Equation. 6)}
\]

The equations described in the GCTA analysis were used to estimate the variance component and breeding values, except that in the models solved using THRGIBBS1F90, the contemporary group had a fixed effect. Consequently, there was no distribution of contemporary group random effect. The model accuracy and individuals' reliability were calculated using Linear Regression methods [78, 79] to test model bias. Selection decisions are typically made regarding the youngest generation, and to avoid the data leakage problem, the individuals within the contemporary groups were kept together when forming validation subsets. As a result, the youngest 15% of individuals were grouped into three validation sets. The first 5% youngest had a missing phenotype for the first validation. The second youngest 5% group had a missing record for the second validation and the same for the third group. The model's accuracy was calculated by comparing the estimated
breeding values of each partial set with the estimated breeding values from the complete set of individuals, as described by Bermann et al.[79].

2.2.6. De-regressed breeding values for Genome-Wide Association (GWAS)

SNP1101 program suite [80] has been used to perform de-regressed genome-wide association studies (GWASs).

The estimated breeding values divided by their reliabilities from the Gibbs sampling models were used as new phenotypes to perform GWAS, weighted by $\frac{1}{\text{Reliability}} - 1$ [80]. The individual's reliability was calculated using $1 - \frac{PVE}{(1 + F_i)\sigma_a^2}$. $PVE$ is the prediction error variance, $F_i$ is the individual’s breeding value coefficient, and $\sigma_a^2$ is the additive variance. The genomic relationship matrix was built by VanRaden's method [77].

2.2.7. Nearest Gene within 10Kb

After performing GWAS, the BovineMine v1.6 search tool [81] was used to find genes within 10kb of genome-wide significant SNPs for each trait separately, to ensure a strong linkage disequilibrium between the casual gene and associated variant [82].

2.3. Results

2.3.1. Variance Component, heritability, and Breeding values estimation

2.3.1.1. Univariate GCTA and GEMMA models

The heifer pregnancy rate among the individuals analyzed only in GCTA and GEMMA was 87.1%, while among all individuals was 81.3%. GCTA and GEMMA used linear mixed models during variance estimation, and both analyses were different in that pre-corrected
phenotypes were used in GEMMA throughout this study. Table 3 reports the genetic variance and heritability from univariate GCTA and GEMMA.

As expected, the reported heritability from GEMMA is higher than that of GCTA. That is because the property of continuous variables has more information than binary variables. Even though there is a small increase in heritability from GCTA to GEMMA, the difference is not significant. The increase in heritability might be associated with fewer noises introduced in GEMMA than GCTA during pre-correcting phenotypes.

Table 3 Estimates of genetic variance $\sigma_g^2$, residual variance $\sigma_e^2$, and heritability $h^2$ from univariate GCTA and GEMMA.

<table>
<thead>
<tr>
<th>Trait</th>
<th>GCTA</th>
<th>GEMMA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma_g^2$</td>
<td>$\sigma_e^2$</td>
</tr>
<tr>
<td>HP</td>
<td>0.0129 $\pm$</td>
<td>0.092 $\pm$</td>
</tr>
<tr>
<td></td>
<td>0.0024</td>
<td>0.0031</td>
</tr>
<tr>
<td>DC</td>
<td>100.641 $\pm$</td>
<td>836.79 $\pm$</td>
</tr>
<tr>
<td></td>
<td>23.138</td>
<td>29.740</td>
</tr>
<tr>
<td>DO</td>
<td>135.586 $\pm$</td>
<td>1382.27 $\pm$</td>
</tr>
<tr>
<td></td>
<td>29.615</td>
<td>39.700</td>
</tr>
</tbody>
</table>

Among all three traits, Heifer Pregnancy (HP) is showing a high heritability compared to Days Open (DO) and Days to Conception (DC).

The HP heritability obtained from the linear model might not be reliable because of the binary nature of HP, which violates the assumptions of normal distribution for the linear models. For that reason, the obtained heritability was adjusted on the liability scale. As a result, the adjusted
HP heritability in GCTA was 0.33. GEMMA's liability scale was not needed because the used variables were continuous.

2.3.1.2. Bivariate GCTA and GEMMA

The bivariate model fit heifer pregnancy and days open. HP and DO model resulted in a correlation of \(-0.61 \pm 0.095\) and \(-0.63 \pm 0.020\) for GCTA and GEMMA, respectively. The variance estimates and heritabilities were not different from the univariate model. The bivariate model of DC and HP could not converge, probably because only pregnant heifers had values for days to conception. Not only was there a convergence problem, but it would not even be easy for results interpretation of the model that includes DC due to the missing records on non-pregnant heifers.

2.3.2. Gibbs Sampling Models variance components and genomic predictions

2.3.2.1. Univariate and bivariate outcomes

The threshold univariate model used 18,039 heifers with records on pregnancy; some of those did not have DC or DO records; thus, the sample size for the later traits was smaller than HP. The fixed effects from the Gibbs sampling model were smaller than the effect from linear models, as can be seen in table 4. However, there was not much difference between heritability estimated from previous linear models (GCTA & GEMMA) and the estimates from the thresholds model. Days to conception only showed a drastic change (0.107 to 0.0749). Table 5 shows the univariate outcomes from posterior distributions while figure 4 shows the convergence of all univariate models and the corresponding genetic variance and heritability. The reported values are the means from posteriors. The HP and DO bivariate model yielded a correlation of 0.86, considering that the HP solutions were on the liability scale. Even though a direct relationship between HP and DC was not estimated, Figure 3 shows a positive correlation between estimated breeding values of DO
and DC, thus, an indirect relation between HP and DC. Furthermore, the heritabilities from bivariate were the same as univariate models.

Table 4. Heifer’s age and season length estimated effect from linear and Gibbs sampling models on Heifer Pregnancy, Days to Conception, and Days Open.

<table>
<thead>
<tr>
<th>Traits</th>
<th>Heifer Pregnancy</th>
<th>Days to Conception</th>
<th>Days Open</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>Gibbs</td>
<td>Linear</td>
</tr>
<tr>
<td>Heifer’s Age</td>
<td>0.0003 ± 0.0002</td>
<td>-0.004 ± 0.00024</td>
<td>-0.178 ± 0.0189</td>
</tr>
<tr>
<td>Season Length</td>
<td>0.0076 ± 0.0018</td>
<td>-0.00166 ± 0.000565</td>
<td>0.0643 ± 0.0168</td>
</tr>
</tbody>
</table>

Figure 3. Correlation between estimated days open (DO) and days to conception (DC) breeding values (EBVs)
Table 5 The number of samples, mean, standard deviation, high posterior density (HPD), autocorrelation, and heritability (h2) estimates from the posterior distribution of the univariate threshold model.

<table>
<thead>
<tr>
<th>Trait</th>
<th>samples</th>
<th>Posterior Mean</th>
<th>HPD</th>
<th>Autocorrelation</th>
<th>h2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Genetic</td>
<td>Residual</td>
<td>Genetic</td>
<td>Residual</td>
</tr>
<tr>
<td>HP</td>
<td>4500</td>
<td>0.117 ± 0.048</td>
<td>0.861 ± 0.0160</td>
<td>0.032; 0.21</td>
<td>0.833; 0.893</td>
</tr>
<tr>
<td>DC</td>
<td>18000</td>
<td>76.7 ± 18.84</td>
<td>947.0 ± 22.46</td>
<td>41.7; 113.8</td>
<td>903.6; 991.6</td>
</tr>
<tr>
<td>DO</td>
<td>18000</td>
<td>170.61 ± 31.57</td>
<td>1495.0 ± 33.83</td>
<td>112.40; 233.6</td>
<td>1432; 1563.0</td>
</tr>
</tbody>
</table>
(a) HP
(b) DC
Figure 4 Trace and histogram plots for (a) Heifer pregnancy, (b) days to conception, (c) days open that show the convergence in threshold models using Gibb’s sampling. The red line and their corresponding value show the mean posterior variance and heritability for trace plots and histograms.

2.3.2.2. Accuracy of the model

Three sets of the youngest heifers were selected as the validation sets to calculate the model accuracy. The accuracy formula was derived by Legarra et al. [78]. \( Acc = \sqrt{\frac{cov(EBV_f, EBV_p)}{(1-F)\sigma_a^2}} \) where Acc is the accuracy, \( cov \) is the covariance between the estimated full \( (EBV_f) \) and partial \( (EBV_p) \)
model breeding values, $F$ is the mean inbreeding value of the individuals in the validation set, and $\sigma^2_a$ is the additive genetic variance from the full model. Among all the three traits, the model that fit days to conception resulted in higher accuracy of $0.38 \pm 0.02$, while days open and heifer pregnancy accuracies were $0.33 \pm 0.03$ and $0.25 \pm 0.05$, respectively. Figure 5 shows the correlation among EBVs in full and partial models for HP, DC, and DO in all three validation sets used to calculate the accuracy. Overall, there was an underestimation of the estimated breeding values for partial set given that the slopes are less than 1.

Figure 5 The relationship between full and partial breeding values in the validation set of (a) heifer pregnancy, (b) days to conception, and (c) days open.
The rel_part is the reliability in the partial validation set with the corresponding number (1, 2, and 3). The R is the Pearson correlation coefficient, S is the slope which is expected to be 1 in case there is no dispersion, and the blue line indicates regression.

2.3.3. **Genome-wide association (GWAS) for SNP effect**

2.3.3.1. **Univariate GCTA and GEMMA**

There was a lack of power in the GCTA univariate GWAS for all three traits as none of the SNPs reached the standard genome-wide significance level (5.0e-08) for common variants. On the other side, 62 SNPs found on chromosomes 2, 3, 4, 17, 18, 23, 25, and 26 reached the suggestive genome significance threshold level (p-value = 1.0e-05) in both GCTA and GEMMA univariate models. Among those, 47 SNPs were significant at the False Discovery Rate (FDR) less than 0.05, distributed on chromosomes 2, 3, and 23. Also, 12 SNPs on chromosomes 1, 2, 6, 13, and 24 for DC and 10 SNPs on chromosomes 8, 10, 12, and 22 for DO, reached suggestive significance levels (Manhattan and QQ plots in Appendix 2.6.3.). There was a weaker SNP association with DC and DO than HP. Neither DC nor DO had a significant SNP at the FDR < 0.05. Overall, there was an increase in power from GCTA to GEMMA (Figure 6). Apart from the SNPs that reached the suggestive significance level in GCTA, GEMMA showed additional 5 SNPs to reach the suggestive level, and 2 SNPs on chromosome 23 were significant at the genome-wide cutoff level (p-value < 5e-08), and 60 SNPs were significant at the 0.05 FDR.
Figure 6 Manhattan plot and QQ plot for (a) GCTA and (b) GEMMA univariate GWAS in HP. The red line in Manhattan plot shows a cutoff level at p-value of 1e-05.

There was an increase in power from GCTA to GEMMA, and the deviation from red line in QQ plot show an association between SNPs and the traits of interest (HP).

2.3.3.2. Bivariate GEMMA

As described by [83], assessing the SNP association using a bivariate model is advantageous when two or multiple traits are believed to be correlated or affected by the same genes. During the bivariate GWAS analysis in GEMMA, both HP and DO were analyzed, given that those traits are genetically correlated. Among 662,044 SNPs, 59 reached the significance level (p-value = 1.0e-05) and 38 SNPs were significant at the FDR less than 0.05. In addition to the chromosomes from
the univariate models, there was a significant SNP on one more chromosome (BTA1) from the bivariate (Figure 7).

![Manhattan plot and QQ plot for GEMMA bivariate GWAS in HP and DO. The red line shows the genome-wide significant level (p-value = 5e-08) while the blue line shows the suggestive significant level (p-value = 1e-05).](image)

2.3.3.3. De-regressed Estimated Breeding Value (EBV) GWAS

Before performing the de-regressed EBV GWAS, the reliability was calculated from the Gibbs sampling models (Table 6). The same problem of having no statistical power for significant SNPs was observed for de-regressed EBVs. Only two SNPs reached the genome-wide significance level (p-value = 5e-08) for HP, while there were no significant SNPs in association with DC and DO (Figure 8). At the suggestive significant level, 58 SNPs and 2 SNPs were found to be associated with HP and DC, respectively. The Manhattan plots below show the significant SNPs for each trait and the corresponding chromosome loci.
Table 6 Reliabilities for heifer pregnancy (HP), days to conception, and days open on genotypes heifer (4097 individuals). SD is the standard deviation.

<table>
<thead>
<tr>
<th>Trait</th>
<th>minimum</th>
<th>mean</th>
<th>maximum</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP</td>
<td>0</td>
<td>0.1931</td>
<td>0.422</td>
<td>0.072</td>
</tr>
<tr>
<td>DC</td>
<td>0</td>
<td>0.223</td>
<td>0.493</td>
<td>0.072</td>
</tr>
<tr>
<td>DO</td>
<td>0.0792</td>
<td>0.419</td>
<td>0.667</td>
<td>0.075</td>
</tr>
</tbody>
</table>

(a) HP

(b) DC
Figure 8 Manhattan QQ plots for (a) Heifer Pregnancy, (b) Days to Conception, and (c) Days Open based on de-regressed breeding values in the SNP1101 program.

The red line in the Manhattan plots shows the Genome-wide significant threshold level adjusted (p-value = 5e-08), and the blue line shows a suggestive significant level (p-value = 1e-05). The HP quantile-quantile (QQ) plot shows that there is an association (deviation from the red line) between a few SNPs and the traits while the plots for DC and DO reveal the lack of power.

2.3.4. Nearest genes within 10kb and Gene set enrichment analysis.

Using Bovinemine v.16 with ARS-UCD1.2 genome assembly, there were six unique genes for HP and DO linear bivariate and 4 and 2 unique genes for HP and DC, respectively, from de-regressed EBV GWAS. There were no genes for DO as none of the SNPs reached at least the suggestive genome-wide significant level. g:Profiler [84] assisted in gene set enrichment analysis. At the Benjamin-Hochberg cutoff level of 0.05, the bivariate GEMMA resulted in 8 gene set terms. On the other hand, HP and DC de-regressed EBV GWAS analysis yielded 75 and 5 gene terms at the same cutoff level, respectively. GRID2 and ZMIZ1 were the most predominant genes in many terms, and they had been identified in other studies to affect fertility in different species (discussed later).
2.4. Discussion

In this study, the heifer pregnancy rate was 87.1%, which falls in the range of 75-95%, as reported in previous studies [23, 85]. Thus, we proceeded with the analysis of heifer pregnancy, knowing the pregnancy rate was in the normal range. Maintaining a high pregnancy rate is critical because problems associated with this trait are the leading cause of culling in the beef industry [86]. Although fertility traits are pivotal in the beef and dairy industry [20, 87–89], they are notorious for their low heritability, making it hard to make genetic progress [86, 90].

The fixed effects were small, and the negative sign shows that the increase in heifer’s age will result in reduced days open and days to conception; however, the increase in season length will increase both days open and days to conception. The effect of both heifer’s age and season length was relatively smaller than that of the previous two traits (DO & DC), and they were close to zero. The reduced effect from the linear model to Gibbs sampling suggests that the effect of both heifer’s age and season length on the analyzed traits might be negligible when considering a very large sample size.

The current study confirms the low heritability nature of fertility traits. Even though the scaled heritability from the linear model was high (0.33), the threshold model (designed to handle binary data) did not confirm that heritability. Hence, the current study assumes algorithm problems during the adjustment and does not consider 0.33 heritability. The linear and threshold models resulted in HP heritability between 0.119 to 0.131 (Tables 3 and 5). This heritability is comparable to the 0.12 ± 0.01 HP heritability reported by Boldt et al. [23] in Red Angus. Various HP heritability estimates have been obtained across the breeds for the past years. For instance, Toelle and Robison reported the heifer pregnancy rate heritability of 0.06 among the Hereford breed [91]. Within the same breed, Evans et al. found the heritability of 0.138 ± 0.08 [21]. Also, in Angus
heifers, the pregnancy heritability was estimated to be between 0.13 and 0.27 [85, 92], and in Brangus, 0.07 [51]. Overall, many breeds show a low estimate of heritability. In their study, Martínez-Velázquez et al. combined nine breeds in one analysis, resulting in the heritability of 0.14 ± 0.03 [93]. Although the *Bos taurus* breeds show a low heifer pregnancy heritability, the inverse seems true for *Bos indicus* breeds. For example, the Nellore breeds showed a high HP heritability estimated to be between 0.37 to 0.57 [94–96]. It is not surprising that the heritability among *Bos taurus* and *Bos indicus* is different because these subspecies have distinct physiological differences that might explain such disparity.

Apart from heifer pregnancy, this study also considered two more fertility traits, days to conception and days open, which are novel in the beef industry. The linear models' genomic heritability estimates (without pedigree information) were 0.112 and 0.102 for DC and DO, respectively. In contrast, the combination of pedigree and genomic information models resulted in 0.0749 and 0.102 for DC and DO, respectively. The cause of the decline in DC heritability is not understood yet. However, it might be because of the increase in the ratio of pregnant heifers over non-pregnant, implying the bias in the previous model with a small sample size. Not many works have been done about DO and DC in the beef industry. In the study done in Spain, the reported heritability of DO in beef cattle ranged from 0.09 to 0.197 [36]. That range is not far from the obtained heritabilities in the current study, though the definition of days open (days from one calving to the next conception) does not quietly match the present study. In the dairy industry, the exact definition of DO - days from calving to next conception - has been used, and a couple of studies reported the heritability between 0.02 to 0.06 [44, 97–102]. The interval from first to successful AI in the study of black Japanese cows by Setiaji and Oikawa is close to DC in this study, and the obtained heritability was 0.028 [99]. The reported heritabilities in the current study
seem to be higher than those reported about the dairy industry. That might be predicated on the worsening of reproductive traits of the dairy cows that happened while focusing on improving the production traits (milk yield) [16].

The genetic correlation between DO and HP from the linear model was -0.61. The threshold model resulted in a 0.85 correlation, and the positive direction was because the solutions for HP were on the liability scale. The linear bivariate model for HP and DC failed to converge. While the threshold model could ignore missing values of DC and fit the bivariate model with only pregnant heifers, the results from that model would be biased due to the lack of not analyzing data on non-pregnant heifers. Consequently, the scaled model was not used to predict the correlation between HP and DC to avoid that bias. To the author's knowledge, there has not been any known study that evaluated the genetic correlation between DO and HP among beef cattle. However, some studies assessed the correlation with other fertility traits. The correlation between pregnancy rate and days open in Japanese black cows was -0.92 [99]. The correlation differences with the current study may be caused by how traits are defined and the unlikeliness of the breeds. A negative correlation between DO and HP suggests that some common genes affect both traits differently. Consequently, it is possible to improve one trait while meticulously selecting the other. That is, while selecting on lower days open, there is a high probability of increasing the heifer pregnancy.

In this study, the model accuracies were 0.38, 0.33, and 0.25 for DC, DO, and HP, respectively. Despite its high heritability compared to the other two traits, the HP model resulted in lower accuracy than DC and DO models. That was expected as the quantitative variables contain more information than binary variables [103]. Therefore, selecting days to conception or days open might lead to more rapid genetic improvement than selection on heifer pregnancy.
Most previous studies on beef fertility used traditional genetic approaches (prediction based on pedigree) for variance and heritability estimation, and almost none considered the model validation. It was easy to overlook the fertility traits during selection decisions based on the results that showed they are lowly heritable. The current study agrees with the low heritability status of HP, DO, and DC, but the achieved accuracies are the best indicators that genomic selection can elaborate on these traits. There is a hope that with sophisticated genomic tools and statistical methods, animal producers should start to invest in improving fertility traits, expecting that there will be some improvement while applying genomic selection.

The bivariate linear mixed model of HP and DO gave 6 unique genes classified into 8 gene ontology terms (Table 7). Those genes are DLGAP1, FBXO21, LOC112445875, SCN3A, SLC37A1, and SLC38A1. According to the GWAS catalog [104], most of those genes (DLGAP1, FBXO21, SCN3A, and SLC38A11) are related to body mass index traits in humans; but it is not clear about the role they play in HP and DO or any other fertility traits in general.

Table 7 Gene terms at Benjamin-Hochberg correction p-value < 0.05 in association with HP and DO.

<table>
<thead>
<tr>
<th>Source</th>
<th>Term name</th>
<th>Intersections</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:BP</td>
<td>Hexose phosphate transport</td>
<td>SLC37A1</td>
<td>0.0224</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Glucose-6-phosphate transport</td>
<td>SLC37A1</td>
<td>0.0224</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Phosphate ion transmembrane transport</td>
<td>SLC37A1</td>
<td>0.0358</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Phosphate ion transport</td>
<td>SLC37A1</td>
<td>0.0391</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Inorganic ion transmembrane transport</td>
<td>SCN3A, SLC37A1</td>
<td>0.0391</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of postsynaptic neurotransmitter receptor activity</td>
<td>DLGAP1</td>
<td>0.0391</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------------------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>KEGG</td>
<td>Glutamatergic synapse</td>
<td>DLGAP1</td>
<td>0.0394</td>
</tr>
<tr>
<td>KEGG</td>
<td>Taste transduction</td>
<td>SCN3A</td>
<td>0.0394</td>
</tr>
</tbody>
</table>

From de-regressed breeding values GWAS, there were 4 and 2 unique genes with 75 and 5 gene set terms (Appendix 2.6.4. table 4 and 5) for HP and DC, respectively, at the Benjamin-Hochberg p-value $\leq 0.05$. Both genes identified from DC are indicated in the human GWAS catalog to be associated with body mass index [104]. If that is the case in this study, the body condition score effect on the fertility traits should be assessed in future research. Overall, most of the gene set terms for HP are related to the central nervous system, and it has been speculated that they may play a role in the secretion of some reproductive hormones. Moreover, GRID2 and ZMIZ1 genes have been identified in the previous studies to have an influence on reproductive traits among different species. For instance, Vohra et al. found GRID2 among the genes associated with the postpartum interval in Indian buffalo [105]. In addition, in human clinical studies, the intragenic deletion in GRID2 has been associated with a neurological disorder with developmental delay as one of the major symptoms [106]. Although it can be argued that the deletion in GRID2 resulted in a bigger effect that might not be seen when considering a small allele change (the case of SNPs), the fact that the function of this gene is not fully understood can be the reason to study in depth of its effect on fertility as it might be affecting the pubertal behavior.

Another group of researchers from China identified ZMIZ1 as one of the genes that were strongly associated with the interval from calving to first insemination among Chinese Holstein [107]. Not only was ZMIZ1 found to enrich the transcription activity of androgen receptors [108],
but May-Panloup et al. found that its less expression level was linked with a diminished ovarian reserve which is an indicator of subfertility [109]. Therefore, it would be important to conduct a validation study for this gene while assessing its effect on pregnancy.

Contrary to the expectation, a drastic reduction in power for the SNP effect was observed; thus, very few genes were associated with the analyzed traits while using de-regressed EBV for GWAS. One possible reason for that might be the small reliability of the estimated EBVs, and it implies that an enormous sample size with phenotypic and genotypic information is needed to conduct analyses on fertility that will result in high reliability, hence, the gain of power.

2.5. Conclusion

The availability of sophisticated computational methods and the improvement in DNA sequencing increase prediction accuracy in hard-to-measure reproductive traits. The current study estimated the genetic parameters for heifer pregnancy, days to conception, and days open which are the fertility traits that are easy to measure and contain enough information for genomic studies. Both linear and threshold models result in almost the same heritability for binary traits, indicating the robustness of the developed linear model tools that successfully handle binary data. Like the previous studies, the heritability of fertility was relatively low. Even though the multivariate between HP and DC did not converge, the high correlation between DO and HP shows that indirect selection for one of these traits can be used to improve the overall reproductive performance in Red Angus.

Furthermore, despite the variation in DC heritability, this variable yielded higher accuracy than the rest. Given its quantitative nature, future studies should assess the feasibility of using DC in multi-trait models that fit binary and continuous variables. Some genes were associated with HP, DC, and DO, but no known biological processes relate those genes to fertility. Further studies
are needed to confirm the influence of the obtained genes and their corresponding gene terms on heifer reproduction. While the current study considered only one breed, future studies should try multibreed analysis with a large sample size to confirm the effect of obtained genes (GRID2 & ZMIZ1) on beef heifer fertility and the feasibility of DC and DO, which are considered novel traits in the industry.
2.6. Reference:

49. Utter SD, Houghto PL, Corah LR, Simms DD, Spir MF, Butine MD. FACTORS INFLUENCING FIRST-SERVICE CONCEPTION AND OVERALL PREGNANCY RATES IN COMMERCIAL BEEF HEIFERS I.
60. Dow JS, Moore JD, Bailey CM, Foote WD. ONSET OF PUBERTY IN HEIFERS OF DIVERSE BEEF BREEDS AND CROSSES 1.
64. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 2007;81:559–75.
69. S Falconer BD. The inheritance of liability to certain diseases, estimated from the incidence among relatives. 1966;29:61.
80. Sargolzaei M. SNP1101 User Guide. 2014;0 March.
2.7. Appendix

2.7.1. Figures before the quality control

Figure 1. The histogram of (a) days to conception, (b) days open, (c) age, and season length before quality control.

For days interval (DC & DO), individuals less -50 and above 200 days seem to be outliers.

For age, many individuals are within 300 and 500 days old, while season length with many individuals ranges 0 to 200 days.
Figure 2. Scatter plots relating (a) days open and (b) days to conception with heifer’s age after quality control. Note that colors indicate birthdate.

2.7.2. Figures after data quality control

The quality control was done based on the literature, and the distribution of raw data.

- First, changed the -9 days interval (DO&DC) to -9.1
- At the same time, removed heifers with more than 365 and less than -30 days and less than 200 days of age (Based on the distribution and assuming errors in data collection)
- Afterward, more than 250 and less than -21 DO and DC to missing
- Finally, less than 300 and more than 550 days of age, as well as more than 360 days of season length as missing.
Figure 3. Distribution of (a) days to conception, (b) days open, (c) heifer’s age, and (d) season length after quality control
Figure 4. Scatter plots relating (a) days open and (b) days to conception with heifer’s age after quality control.

2.7.3. Days to conception and days open for GCTA and GEMMA Manhattan and QQ plots

(a)

(b)

(c)
There was a weak association between a few SNPs and the both traits (DC - Days to conception and DO - Days Open). Also, there was an increase in power from GCTA to GEMMA, probably because of less noise introduced in GEMMA and most of the variation was due to genetics as estimated breeding values were used to generate pseudo-phenotypes.

2.7.4. Figures that show traits relationship among estimated breeding values, as well as the breeding values from linear models and Gibbs sampling.

Figure 5. Manhattan and QQ plots for days to conception from GCTA (a) and GEMMA, and Manhattan and QQ plots for Days Open (DO) from GCTA (c) and GEMMA(d).

Figure 6. scatter plots that show the relationship between traits based on the estimated breeding values (a) heifer pregnancy and days to conception, (b) heifer pregnancy and days open, and (c) days open and days to conception.

The red lines show the regression line while R is the correlation between breeding values.
Figure 7. EBVs relationship from Gibbs sampling (THR) and linear model (GCTA) for days open (a) and days to conception (b). The red lines show the regression line while R is the correlation between breeding values.

2.7.5. Gene-Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) term enrichment from genes associated from linear and Gibbs sampling models

Table 1. Heifer pregnancy GO and KEGG terms from linear models (GCTA & GEMMA)

<table>
<thead>
<tr>
<th>Source</th>
<th>Term name</th>
<th>Intersection</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:BP</td>
<td>regulation of metanephros size</td>
<td>PAX2</td>
<td>0.03</td>
</tr>
<tr>
<td>KEGG</td>
<td>Mucin type O-glycan biosynthesis</td>
<td>GALNT13</td>
<td>0.026</td>
</tr>
<tr>
<td>KEGG</td>
<td>Other types of O-glycan biosynthesis</td>
<td>GALNT13</td>
<td>0.026</td>
</tr>
<tr>
<td>KEGG</td>
<td>Taste transduction</td>
<td>SCN3A</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Table 2. Days to conception GO and KEGG terms from linear models (GCTA & GEMMA)

<table>
<thead>
<tr>
<th>source</th>
<th>Term name</th>
<th>Intersection</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:BP</td>
<td>Negative regulation of peptidase activity</td>
<td>SERPINB12, SERPINB13</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of endopeptidase activity</td>
<td>SERPINB12, SERPINB13</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Hexose phosphate transport</td>
<td>SLC37A1</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Glucose-6-phosphate transport</td>
<td>SLC37A1</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Term name</td>
<td>Intersection</td>
<td>p-value</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>GO:BP</td>
<td>negative regulation of keratinocyte apoptotic process SERPINB13</td>
<td>---------------</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of keratinocyte apoptotic process SERPINB13</td>
<td>---------------</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Keratinocyte apoptotic process SERPINB13</td>
<td>---------------</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>negative regulation of proteolysis SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of hydrolase activity SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0183</td>
</tr>
<tr>
<td>GO:BP</td>
<td>regulation of endopeptidase activity SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0216</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Phosphate ion transmembrane transport SLC37A1</td>
<td>---------------</td>
<td>0.0224</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of peptidase activity SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0224</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of postsynaptic neurotransmitter receptor activity DLGAP1</td>
<td>---------------</td>
<td>0.0311</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Phosphate ion transport SLC37A1</td>
<td>---------------</td>
<td>0.0344</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of catalytic activity SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0344</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of proteolysis SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0386</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of molecular function DLGAP1, SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0407</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of epithelial cell apoptotic process SERPINB13</td>
<td>---------------</td>
<td>0.0407</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of protein metabolic process SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0476</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of molecular function SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0476</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of neurotransmitter receptor activity DLGAP1</td>
<td>---------------</td>
<td>0.0476</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of cellular protein metabolic process SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0476</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of epithelial cell apoptotic process SERPINB13</td>
<td>---------------</td>
<td>0.0476</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of hydrolase activity SERPINB12, SERPINB13</td>
<td>---------------</td>
<td>0.0476</td>
</tr>
<tr>
<td>KEGG</td>
<td>Glutamatergic synapse DLGAP1</td>
<td>---------------</td>
<td>0.0264</td>
</tr>
</tbody>
</table>

Table 3. Days to conception KEGG terms from linear models (GCTA & GEMMA)

<table>
<thead>
<tr>
<th>Source</th>
<th>Term name</th>
<th>Intersection</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG</td>
<td>Signaling pathways regulating pluripotency of stem cells ESRRB</td>
<td>---------------</td>
<td>0.0341</td>
</tr>
</tbody>
</table>

52
Table 4. Heifer pregnancy GO and KEGG terms using de-regressed EBV (SNP1101)

<table>
<thead>
<tr>
<th>source</th>
<th>Term name</th>
<th>Intersections</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:BP</td>
<td>Pyramidal neuron development</td>
<td>ZMIZ1</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Pyramidal neuron differentiation</td>
<td>ZMIZ1</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Pyramidal neuron migration to cerebral cortex</td>
<td>ZMIZ1</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Central nervous system neuron differentiation</td>
<td>GRID2, ZMIZ1</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellar granule cell differentiation</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of postsynaptic specialization assembly</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of postsynaptic density assembly</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellar granular layer formation</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellar granular layer morphogenesis</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Radial glia-guided pyramidal neuron migration</td>
<td>ZMIZ1</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellar granular layer development</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Positive regulation of long-term synaptic depression</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Prepulse inhibition</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Vitellogenesis</td>
<td>ZMIZ1</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of excitatory synapse assembly</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of postsynaptic density organization</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of long-term synaptic depression</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Postsynaptic density assembly</td>
<td>GRID2</td>
<td>0.0285</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Postsynaptic specialization assembly</td>
<td>GRID2</td>
<td>0.0299</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Brain development</td>
<td>GRID2, ZMIZ1</td>
<td>0.0299</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cytoplasm organization</td>
<td>ZMIZ1</td>
<td>0.0299</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Head development</td>
<td>GRID2, ZMIZ1</td>
<td>0.032</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Long-term synaptic depression</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Telencephalon glial cell migration</td>
<td>ZMIZ1</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Postsynaptic density organization</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of protein sumoylation</td>
<td>ZMIZ1</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Ionotropic glutamate receptor signaling pathway</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Postsynaptic specialization organization</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cell-cell adhesion</td>
<td>GRID2, ZMIZ1</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebral cortex radial glia-guided migration</td>
<td>ZMIZ1</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Postsynapse assembly</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellar cortex formation</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cell differentiation in hindbrain</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Excitatory synapse assembly</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Description</td>
<td>Gene(s)</td>
<td>Score</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Central nervous system development</td>
<td>GRID2, ZMIZ1</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Ligand-gated ion channel signaling pathway</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Startle response</td>
<td>GRID2</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Forebrain neuron development</td>
<td>ZMIZ1</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:BP</td>
<td>neuron development</td>
<td>GRID2, ZMIZ1</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Negative regulation of synaptic transmission</td>
<td>GRID2</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Positive regulation of fibroblast proliferation</td>
<td>ZMIZ1</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Androgen receptor signaling pathway</td>
<td>ZMIZ1</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Hindbrain morphogenesis</td>
<td>GRID2</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Positive regulation of Notch signaling pathway</td>
<td>ZMIZ1</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellar cortex morphogenesis</td>
<td>GRID2</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebral cortex radially oriented cell migration</td>
<td>ZMIZ1</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Forebrain neuron differentiation</td>
<td>ZMIZ1</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Glial cell migration</td>
<td>ZMIZ1</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellum morphogenesis</td>
<td>GRID2</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules</td>
<td>GRID2</td>
<td>0.0422</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebral cortex cell migration</td>
<td>ZMIZ1</td>
<td>0.0428</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Forebrain generation of neurons</td>
<td>ZMIZ1</td>
<td>0.0428</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Description</td>
<td>Gene(s)</td>
<td>p-value</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Glutamate receptor signaling pathway</td>
<td>GRID2</td>
<td>0.0439</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Positive regulation of developmental process</td>
<td>GRID2, ZMIZ1</td>
<td>0.0439</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Protein sumoylation</td>
<td>ZMIZ1</td>
<td>0.0439</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cerebellar cortex development</td>
<td>GRID2</td>
<td>0.0439</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Positive regulation of synapse assembly</td>
<td>GRID2</td>
<td>0.0439</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of multicellular organismal development</td>
<td>GRID2, ZMIZ1</td>
<td>0.0445</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Positive regulation of multicellular organismal process</td>
<td>GRID2, ZMIZ1</td>
<td>0.046</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Artery morphogenesis</td>
<td>ZMIZ1</td>
<td>0.046</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of postsynapse organization</td>
<td>GRID2</td>
<td>0.046</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Neuron differentiation</td>
<td>GRID2, ZMIZ1</td>
<td>0.046</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Telencephalon cell migration</td>
<td>ZMIZ1</td>
<td>0.046</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Forebrain cell migration</td>
<td>ZMIZ1</td>
<td>0.0472</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Cell adhesion</td>
<td>GRID2, ZMIZ1</td>
<td>0.0472</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Fibroblast proliferation</td>
<td>ZMIZ1</td>
<td>0.0472</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Central nervous system neuron development</td>
<td>ZMIZ1</td>
<td>0.0472</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Biological adhesion</td>
<td>GRID2, ZMIZ1</td>
<td>0.0472</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Synaptic transmission, glutamatergic</td>
<td>GRID2</td>
<td>0.0472</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Regulation of fibroblast proliferation</td>
<td>ZMIZ1</td>
<td>0.0472</td>
</tr>
<tr>
<td>Source</td>
<td>Term name</td>
<td>Intersections</td>
<td>p-value</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Activation of cysteine-type endopeptidase activity involved in apoptotic process by cytochrome c</td>
<td>DIABLO</td>
<td>0.0278</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Poly-N-acetyllactosamine metabolic process</td>
<td>B3GNT4</td>
<td>0.0278</td>
</tr>
<tr>
<td>GO:BP</td>
<td>Poly-N-acetyllactosamine biosynthetic process</td>
<td>B3GNT4</td>
<td>0.0278</td>
</tr>
<tr>
<td>KEGG</td>
<td>Glycosphingolipid biosynthesis - lacto and neolacto series</td>
<td>B3GNT4</td>
<td>0.0209</td>
</tr>
<tr>
<td>KEGG</td>
<td>Apoptosis - multiple species</td>
<td>DIABLO</td>
<td>0.0209</td>
</tr>
</tbody>
</table>

Table 5. Days to conception GO and KEGG terms using de-regressed EBV (SNP1101)