A genomic assessment of the correlation between milk production traits and claw and udder health traits in Holstein dairy cattle

Helen Schneider,1* Dierck Segelke,2 Jens Tetens,3 Georg Thaller,4 and Jörn Bennewitz1

1Institute of Animal Science, University of Hohenheim, 70599 Stuttgart, Germany
2Vereinigte Informationssysteme Tierhaltung w.V. (VIT), 27283 Verden, Germany
3Department of Animal Sciences, University of Göttingen, 37077 Göttingen, Germany
4Institute of Animal Breeding and Husbandry, Christian-Albrechts University of Kiel, 24098 Kiel, Germany

ABSTRACT

Claw diseases and mastitis represent the most important disease traits in dairy cattle with increasing incidences and a frequently mentioned connection to milk yield. Yet, many studies aimed to detect the genetic background of both trait complexes via fine-mapping of quantitative trait loci. However, little is known about genomic regions that simultaneously affect milk production and disease traits. For this purpose, several tools to detect local genetic correlations have been developed. In this study, we attempted a detailed analysis of milk production and disease traits as well as their interrelationship using a sample of 34,497 50K genotyped German Holstein cows with milk production and claw and udder disease traits records. We performed a pedigree-based quantitative genetic analysis to estimate heritabilities and genetic correlations. Additionally, we generated GWAS summary statistics, paying special attention to genomic inflation, and used these data to identify shared genomic regions, which affect various trait combinations. We performed a pedigree-based quantitative genetic analysis to estimate heritabilities and genetic correlations. Additionally, we generated GWAS summary statistics, paying special attention to genomic inflation, and used these data to identify shared genomic regions, which affect various trait combinations. The heritability on the liability scale of the disease traits was low, between 0.02 for laminitis and 0.19 for interdigital hyperplasia. The heritabilities for milk production traits were higher (between 0.27 for milk energy yield and 0.48 for fat-protein ratio). Global genetic correlations indicate the shared genetic effect between milk production and disease traits on a whole genome level. Most of these estimates were not significantly different from zero, only mastitis showed a positive one to milk (0.18) and milk energy yield (0.13), as well as a negative one to fat-protein ratio (−0.07). The genomic analysis revealed significant SNPs for milk production traits that were enriched on Bos taurus autosome 5, 6, and 14. For digital dermatitis, we found significant hits, predominantly on Bos taurus autosome 5, 10, 22, and 23, whereas we did not find significantly trait-associated SNPs for the other disease traits. Our results confirm the known genetic background of disease and milk production traits. We further detected 13 regions that harbor strong concordant effects on a trait combination of milk production and disease traits. This detailed investigation of genetic correlations reveals additional knowledge about the localization of regions with shared genetic effects on these trait complexes, which in turn enables a better understanding of the underlying biological pathways and putatively the utilization for a more precise design of breeding schemes.

Key words: disease traits, milk traits, genetic correlations, GWAS

INTRODUCTION

During the last decades, the genetic gain for milk production traits in Holsteins increased steadily, and was nearly doubled by genomic selection (VIT, 2021). Recently, new functional and disease traits entered the total merit index in many countries. Mastitis and claw diseases are especially important in this context, as they appear to be the most frequent reasons for culling of cows with increasing incidences (Weber et al., 2013). Rising incidences are nowadays specifically precarious because of their ecological (Mostert et al., 2018) and economical importance, accompanied by a rising public awareness for animal welfare (Miglior et al., 2005). Disease cases are major cost factors in dairy production, mainly induced by milk loss, treatment costs, extended days open, disease recurrence, and increased risks of culling and death (Dolecheck et al., 2019; Heringstad et al., 2000), augmented by a reduced selection capacity (Pritchard et al., 2013). It has been frequently reported that unfavorable genetic correlations between disease and milk traits exist (van Dorp et al., 1998; Fleischer et al., 2001; Wu et al., 2007; Koeck et al., 2012). They are mainly caused by the negative energy balance (NEB), and subsequent physiological imbalances that occur.
Thus, counteracting potential negative side effects of decreasing mastitis resistance (Cai et al., 2020) and to increase the protein yield and mastitis resistance (Nilsen et al., 2009). The region is a region on chromosome 6 around 89 Mbp. It affects mastitis and milk production. Indeed, they detected a region around 89 Mbp on chromosome 6 with a significant effect, which might, for example, increase milk yield and mastitis resistance (Olsen et al., 2016; Pimentel et al., 2011; Bulik-Sullivan et al., 2015b; Li et al., 2017). Pleiotropy was identified to exist on a widespread level also in cattle (Xiang et al., 2021). Recent studies aimed to dissect the connection between the milk production traits on the one hand, the disease traits on the other hand, toward local genetic correlation estimates within specific genomic regions. Global estimates are limited in their interpretation as they might be biased in cases where local correlations appear in antagonistic directions. Moreover, they also give merely averaged estimates across the genome, hiding strong local correlations in the absence of any global one (Shi et al., 2017; van Rheezen et al., 2019). Therefore, common global estimates of genetic correlation cannot be directly transferred to the local scale (Werme et al., 2022). Several tools to break down global estimates toward a local scale have been invented, which are either focusing on single SNP shared genetic effect (e.g., Cai et al., 2020; Werme et al., 2022) or the quantification of (co)variance within specific genomic regions or whole chromosomes (e.g., Pimentel et al., 2011; Bulik-Sullivan et al., 2015b; Li et al., 2017). Pleiotropy was identified to exist on a widespread level also in cattle (Xiang et al., 2021). Recently, studies aimed to dissect the connection between mastitis and milk production. Indeed, they detected a region on chromosome 6 around 89 Mbp. It affects both traits in the direction to increase milk production and to decrease mastitis resistance (Olsen et al., 2016; Cai et al., 2020), and to increase the protein yield and mastitis resistance (Nilsen et al., 2009). The region is connected to the group-specific component (GC) gene, which encodes the vitamin D-binding protein and plays a role in milk production and the immune defense (Olsen et al., 2016). Hence, illuminating the shared genetic architecture in detail is beneficial not only for scientific interest because it enhances the understanding of biological pathways. It may also enable improved genomic selection decisions due to a different weighting of markers based on their potential pleiotropic and antagonistic effect, which might, for example, increase milk yield but decrease mastitis resistance (Cai et al., 2020). Thus, counteracting potential negative side effects of high production on disease prevalence induced by pleiotropic SNPs is facilitated. Recently, Guo et al. (2021) proposed a tool to detect shared genomic regions called LOGODetect (“LOcal Genetic eOrrelation Detector”) with an intuitive search for these regions because it is detached from knowledge based prior assumptions about their localization in the genome.

The LOGODetect tool is based on GWAS summary data. When generating these summary data in dairy cattle, one has to account for the large population structure with many half-sib families to prevent an inflation of type I errors. This is typically done by a full genomic relationship matrix or principal components (PC; van den Berg et al., 2019; Yin and König, 2019). Choosing the optimal stringency for genomic correction of inflation is a substantial part of doing a GWAS because it ensures the optimal balance between false-positive and false-negative findings (van den Berg et al., 2019). This holds true especially if the GWAS summary statistics are used not only for the inference of potential QTL positions or causal mutations, but also for post GWAS analyses such as the local genetic correlation estimation. A too stringent control would compromise power to detect local correlations because only few usable GWAS summary hits remain after a conservative correction.

The aim of the present study was to gain detailed insight into the genetic background of claw disease, mastitis, and milk production traits on a large cow sample of German Holstein cattle. The paper is split into 3 parts. First, we performed a pedigree-based quantitative genetic analysis of milk traits, as well as claw diseases and mastitis, to estimate heritabilities and global genetic correlations. Next, we generated GWAS summary statistics for these traits, paying special attention to genomic inflation. The last step was to use these summary data to break down the genetic correlations between the milk production traits on the one hand, and the disease traits on the other hand, toward local genetic correlations within specific genomic regions.

**METHODS**

No live animals were used in this study; therefore, ethical approval was deemed unnecessary.

**Phenotypic and Genotypic Data**

We analyzed 50K genotypes, raw phenotypes, and the pedigree data of 34,497 primiparous Holstein cows with lactation records for milk and disease traits collected between 2015 and 2020, whose pedigree consisted of 90,407 animals. This data set was a subset...
of a larger data set and was filtered with the aim to have a sufficient number of observations in each effect class because there might be a strong environmental impact, which needs to be considered in the analyses. A sufficient number of observations in each environmental class implies that their effects can be estimated precisely and they can be separated from genetic effects. The data were provided by the national computing center VIT (Vereinigte Informationssysteme Tierhaltung w.V.). Milk traits were first lactation records of milk yield in kilograms (MY), fat, and protein yield in kilograms. Diseases were binary coded in the way that 0 means “no disease” and 1 means “disease,” and multiple events within a disease were not considered. They were recorded on the farms mainly by farmers, veterinarians, and claw trimmers. Recorded were the following claw diseases: claw ulcers (CUL), digital dermatitis (CDD), interdigital hyperplasia (CIH), laminitis (CLM), digital phlegmon (CPH), and white line disease (CWL), as well as mastitis (MAS). For all claw diseases, the risk period ranged from calving until 305 DIM, and for MAS, the span was from 10 d before calving until 305 DIM. The data set also contained measurements of the overall DIM, ranging from 270 to 305 d with a mean of 302.43 d. Thus, the cows completed or nearly completed the first lactation. Age at first calving (AFC) was recorded in months and varied between 21 and 36 mo with a mean of 24.83 mo. The data editing followed the following criteria. We decided to exclude farms for which less than 10 cows had an event of diseases to avoid biased results due to incomplete reporting of diseases. This led to a total of 103 farms with an average size of 345 cows per farm. The farm, year, and season of calving for the cow were combined to herd-year-season (hys), where each season covered 3 mo according to the calendric partitioning of seasons. Each class of hys contained at least 20 individuals, as well as each class of AFC contained a minimum of 20 individuals. The milk energy yield in MJ was calculated with the formula of Nöstitz and Mielke (1995). In this formula, milk yield in kilograms was multiplied with 0.802, and added to 38.4 times fat yield plus 23.6 times protein yield. Fat-protein ratio was calculated as the ratio of fat to protein content.

**Statistical Analysis**

**Pedigree-Based Variance Component Estimation.** Variance components were estimated with Bayesian uni- and bivariate animal models that made usage of Markov chain Monte Carlo sampling techniques with a Gibbs sampler as implemented in the R-package MC-MCglimm (Hadfield, 2010). Because the disease traits were binary coded, we analyzed them with the following threshold model that is independent from incidences (Gianola and Foulley, 1983):

$$\text{liab} = \mathbf{X}\beta + \mathbf{Z}\alpha + \mathbf{W}h + e,$$

where liab is a vector of unobservable variables of the cows. Subsequently, these unobservable variables were denoted as liabilities, which were sampled as random variables from truncated normal distributions while conditioning on the other fixed and random parameters in the model. The vector $\beta$ contains the mean and AFC as a fixed effect for all claw diseases, but not for MAS because the effect was not significant; that is, the corresponding highest posterior density (HPD) interval (95%) did not overlap zero. The corresponding incidence matrix that links the trait records to the fixed effect was denoted with $X$. The terms $\alpha$ and $h$ are the random additive genetic and hys effects, respectively, with their corresponding matrices $Z$ and $W$. It was assumed that both follow a normal distribution with $\alpha \sim N(0, \sigma^2_\alpha)$ and $h \sim N(0, \sigma^2_h)$, where $A$ is the pedigree-based relationship matrix and $I$ is an identity matrix. The prior for both random effects was set to follow a $\chi^2$ distribution with 1 df. The term $e$ is the residuum with a fixed residual variance of $\sigma^2_e = 1$. The heritability on the liability scale was calculated as $h^2_{\text{liab}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_h + \sigma^2_e}$.

Milk yield, FPR, and MEY are so-called Gaussian traits because they follow a normal distribution. Thus, we analyzed them with the following linear model that was similar to the one for binary traits:

$$y = \mathbf{X}\beta + \mathbf{Z}\alpha + \mathbf{W}h + e,$$

where $y$ is the vector of phenotypic observations and the residual variance $e$ was not fixed, but had a variance $e \sim N(0, \sigma^2_e)$. Here, the vector $\beta$ contains the mean as well as the fixed effects AFC and DIM. The age at first calving was included for all 3 milk production traits, whereas DIM was excluded for FPR because it was not significant. Prior assumptions in the linear model were weak with an uninformative prior following $\sigma^2_\alpha \sim \text{inv-gamma}(0.01, 0.01)$ and $\sigma^2_h \sim \text{inv-gamma}(0.01, 0.01)$. The heritability on the observed scale, $h^2_{\text{obs}}$, was calculated in the common manner. Both univariate models had a Markov chain Monte Carlo (MCMC) length of 15,000 iterations with a burn-in of 1,000.

To estimate the genetic correlations between the 3 milk traits and disease traits, we performed a series
of bivariate analyses. Throughout these analyses, the binary nature of the disease traits was ignored and they were treated as Gaussian traits. We followed this procedure for computational reasons because several previous studies (e.g., Vinson et al., 1976; König et al., 2008) showed that the difference between linear and threshold models for this ease is marginal. The MCMC chain length was 25,000 iterations with a burn-in of 5,000 iterations. For both, univariate and bivariate applications, the number of iterations, and the burn-in period were chosen according to a visual inspection of the trace plots and to ensure an effective sample size for the variance components >100.

**GWAS.** For the subsequent genetic analysis 44,126 SNPs remained after filtering out SNPs on sex chromosomes and those with a MAF below 0.01. We performed the GWAS with mixed linear models as implemented in the GCTA software version 1.92.3 beta3 (Yang et al., 2011a). Four different models were applied, which differed in the way of accounting for population structure to prevent an inflation of type I errors and to ensure a sufficient power to detect true associations. The first method included a polygenic term that implemented the full genomic relationship matrix \( G \) (Yang et al., 2014) across all chromosomes (known as MLMA in GCTA applications). The second method was the so-called “leave-one-chromosome-out” approach (known as LOCO in GCTA), where the chromosome on which the candidate SNP is located was excluded from the calculation of \( G \) (Yang et al., 2014). Two additional methods were an extension of the LOCO approach that aim to account for the underestimated relationship among individuals due to the missing chromosome in the LOCO analysis. Twenty PC, computed with GCTA, were included as covariates in the model. The choice of 20 PC was based on a preliminary analysis with either 5 or 10 PC, which did not result in a sufficient decline of genomic inflation factors. The 20 PC were either on a genome-wide (method PCgenome) or chromosomal (method PCchromosome) level, where the latter involved only the PC for the chromosome, on which the candidate SNP was located. We performed both of these methods because PC on a genome-wide level are likely to include overlapping information with the \( G \) and those on a chromosomal level do not (Yin and König, 2019). All methods applied the following mixed linear model that is implemented in GCTA:

\[
y = Xb + g + e.
\]

Here, \( y \) is the vector of raw phenotypes that were standardized to \( y \sim N(0,1) \) for the Gaussian traits to facilitate the interpretation of SNP effects, but remained binary coded for the disease traits. The vector \( b \) contains the mean and fixed effects including the SNPs effect, the lys, AFC, or DIM (the same effects were fitted to each trait as described in the previous section). In addition, \( b \) contains the PC for the methods PCgenome and PCchromosome. The incidence matrix \( X \) relates \( b \) to the SNPs number of allele copies and further links the trait records to the covariate effects. The variable \( g \) is the polygenic term, with variance \( g \sim N(0,G\sigma_g^2) \), implementing either the complete or the LOCO-adjusted \( G \), and \( e \) the residual term with variance \( e \sim N(0,\sigma_e^2) \). Because our aim was the detection of QTL positions rather than to perform a fine-mapping of causal mutations, we decided to follow van den Berg et al. (2019) and adjusted the \( P \)-values from GWAS for the false discovery rate following the Benjamini and Yekutieli (2001) approach. We chose a threshold of false discovery rate \( \leq 0.1 \) to declare a SNP as significantly trait-associated. To assess the level of inflation of type I errors, we calculated the genomic inflation factor \( \lambda_{GC} \) from the \( P \)-values of SNPs as the ratio of the median of the observed chi-squared test statistics with 1 df to the corresponding expected median (van den Berg et al., 2019).

**Mapping of Shared Genomic Regions for Milk Production and Disease Traits.** To identify genomic regions with significant shared effects on trait pairs, we applied LOGODetect, following the methods of Guo et al. (2021). Several 21 trait combinations were considered, always with 1 milk production trait (MEY, FPR, or MY) and 1 of the 7 disease traits. Summary statistics from the GWAS results were used as input data. Association z-scores of SNPs were calculated as the estimated SNP effect divided by its standard error. For FPR, MY, and CDD, we decided to proceed with the summary statistics that resulted in second lowest inflation (i.e., method PCgenome for FPR and MY, and LOCO for CDD). For all other traits, we used the summary statistic that yielded the lowest inflation. The reason for this was that the method, which resulted in less inflated summary statistics for MY, FPR, and CDD, hampered the detection of shared genomic regions, likely as a consequence of a too stringent correction. Thus, our decision related to the narrow ridge between stringent correction of inflation and thereby compromising power and loose correction with an increased number of false-positive results on the other side. Our motivation to accept a slightly increased inflation for the traits FPR, MY, and CDD seemed to be also justified by the fact, that LOGODetect contains powerful filtering steps for genomic inflation from...
GWAS summary statistics through the normalization of SNP effects via linkage disequilibrium (LD), as described in the following.

The key point of LOGO Detect is the scan statistic \( Q \) of a chromosomal segment \( R \) with \( z_i \) and \( z_2 \), as the association z-scores of SNP \( i \) for trait 1 and 2, respectively, which is defined as

\[
Q(R) = \frac{\sum_{i \in R} (z_i z_2_i)}{\sum_{i \in R} (z_i)^2}.
\]

The LD score of the \( i \)th SNP, \( l_i \), was computed within a window of 2.5 Mbp with the software “LD score regression” (Bulik-Sullivan et al., 2015b). We decided to apply this window size because a previous study (Qa-nbari et al., 2010) observed a substantial LD up to a range of 2.5 Mbp in the German Holstein population. The term \( \theta \), defined to take values between 0 and 1, controls for the effect of the LD score. It is a weighting factor that determines the length of the identified segment with putative shared genetic effects. A large (small) value implies a strong (weak) penalty and thus reduces (increases) the segment size. Following Guo et al. (2021), we used 9 values between 0.3 and 0.7 with step sizes of 0.05 for \( \theta \) and applied stratified LD score regression (Finucane et al., 2015) in the following manner to choose the optimal \( \theta \). Up to 9 sets of segments were identified with the scan statistic, because each of the 9 values for \( \theta \) is used in the scan statistic. All 44,126 SNPs were partitioned into 2 categories for each of these sets. The first category included all SNPs that lay within the identified segments, whereas the second category included the remaining SNPs. Next, we measured the genetic covariance and covariance enrichment within these 2 categories, respectively. Here, covariance enrichment is defined as the proportion of covariance divided by the proportion of SNPs within each category (Finucane et al., 2015). A larger genetic covariance points to a larger shared genetic effect on the 2 traits under consideration, whereas a larger covariance enrichment indicates a larger genetic covariance that is explained per SNP. Finally, we selected the \( \theta \) for which the first category showed the highest covariance enrichment.

A sliding window approach was applied to determine \( R \) along the genome, which was set to consist of not more than 10 consecutive SNPs. If a region \( R \) harbors strong (weak) concordant effects on traits 1 and 2, the numerator \( \sum_{i \in R} (z_i z_2_i) \) yields in large (small) absolute values and this indicates a genomic segment with (without) shared genetic effects. Because the inference from absolute values might be biased by strong LD within \( R \), the denominator \( \sum_{i \in R} (l_i)^2 \) is applied to account for LD among SNPs.

To assess if an estimate \( \hat{Q}(R) \) with the previously identified optimal \( \theta \) points to an enrichment of covariance and, thus, to significant shared genetic effects, we had to define a distribution of \( \hat{Q}(R) \) under the null hypothesis of no shared genetic effects. The LOGO Detect method provides a step where this is obtained by Monte Carlo simulations. Details are described in Guo et al. (2021). Briefly, 5,000 pseudo-samples of z-scores for the 2 traits of interest are sampled from a normal distribution with mean 0, a variance that assumes an equal distribution of the trait heritability across the genome, and a covariance that is fixed to 0. During the sampling process, the LD structure of the SNPs is considered. Subsequently, sampled pseudo z-scores are scanned as described above, which resulted in a null distribution of the scan statistics \( Q(R) \).

Finally, we performed an iterative procedure of scanning the genome for segments with significant genetic sharing. The first segment, \( R_1 \), is determined in the way that \( Q(R_1) \) maximizes the scan statistic from the real data. Then, \( Q(R_1) \) is compared with the upper 95% quantile of \( Q_{\text{max}} \) under the null distribution \( Q_{0.95} \). This threshold level was taken from Guo et al. (2021) and corresponds to a 5% genome-wide significance. If \( Q(R_1) \geq Q_{0.95} \), \( R_1 \) is declared to be significant, the SNPs within \( R_1 \) are excluded from the data set and the scanning procedure is repeated, until no further significant segment is identified. Because of the aforementioned strong LD in German Holstein, we merged the identified segments that were not more than 2.5 Mbp apart from each other to one region and averaged the scan statistic \( Q(R) \) and \( P \)-value of these segments to obtain a value for the merged region.

**RESULTS**

**Genetic Parameter Estimation**

Some descriptive statistics can be taken from Table 1, and results from the pedigree-based genetic analysis are shown in Table 2. All disease traits showed low incidences, except of CDD (around 20%) and MAS (16%). The heritabilities on the liability scale were low to moderate with a range from 0.02 (CLM) to 0.19 (CIH) and the lower 95% quantile of the heritabilities for CWL and CLM was close to 0. The heritabilities for the milk production traits were moderate to high.

Table 3 shows the genetic correlations between milk production and disease traits. Most of the genetic correlations were not significantly different from 0 as their HPD interval overlapped 0. Mastitis showed a moder-
ate positive correlation to MY and to MEY, as well as a low negative correlation to FPR. In turn, CUL showed a positive correlation to MEY. A positive correlation between the diseases and MEY as well as MY indicates an unfavorable correlation because it implies a deteriorated udder and claw health with increasing milk production, which holds true in the opposite direction for negative correlations. Regarding the correlation to FPR it is important to mention, that this trait has an intermediate optimum. Thus, a positive correlation indicates increasing incidences with increasing FPR, whereas a negative one indicates increasing incidences as soon as the FPR decreases. As mentioned before, the latter is a hint for a subclinical acidotic state, whereas a high FPR is accompanied by several diseases such as lameness, mastitis, and deteriorated fertility. Some further correlations were almost significant, as one of the HPD interval’s boundaries was close to zero, for example between MY and CUL.

GWAS

The inflation factors from the 4 GWAS methods are shown in Table 4. Double fitting of SNPs in the MLA approach leads to deflated summary statistics. On the other side, as soon as the whole chromosome was excluded from the polygenic term (LOCO), inflation became a severe issue. This was most obvious for the milk production traits and made the application of the PCchromosome method for milk production traits unavoidable to reduce \( \lambda_{GC} \). Therefore, for the inference of the QTL position, we used the LOCO method for the disease traits CLM and CWL, the PCgenome method for the traits CUL, CDD, CIH, CPH, and MAS, and the PCchromosome method for MEY, MY, and FPR. For these methods, we found significant SNPs for all milk production traits and CDD. They were predominantly localized on BTA 5, 6, and 14 for MY, MEY, and FPR, for example, in the already known region of \( DGAT1 \), which is a well-known major gene on BTA 14 that affects milk production traits in Holsteins (Grisart et al., 2002; Winter et al., 2002). For FPR, also BTA 27 showed enriched findings of significant SNPs and for CDD, BTA 1, 5, 11, 13, 14, 22, and 23 harbor noticeable results. Even though the other disease traits did not appear with significant hits, some SNPs were close to reach significance for MAS (on BTA 6, 13, 16, and 24) and CIH (on BTA 8, 14, 16, and 25).

| Trait                | Incidence | Mean  | SD   | Minimum | Maximum |
|----------------------|-----------|-------|------|---------|---------|
| Milk yield (kg)      | 9,551.25  | 1,622.62 | 3,329.00 | 17,144.00 |
| Fat-protein ratio    | 1.14      | 0.11  | 0.66 | 1.68    |
| Milk energy yield (MJ)| 29,494.21 | 4,390.93 | 9,770.26 | 47,525.26 |
| Mastitis             | 16.02     |       |      |         |
| Claw ulcers          | 6.11      |       |      |         |
| Dermatitis digitalis| 19.85     |       |      |         |
| Interdigital hyperplasia | 2.54  |       |      |         |
| Digital phlegmon     | 4.27      |       |      |         |
| White line disease   | 3.87      |       |      |         |
| Laminitis            | 4.83      |       |      |         |

Table 1. Descriptive statistics, where the incidences (in %) of the disease traits and the mean, SD, minimum, and maximum of milk production traits are shown

Table 2. Genetic parameters from the univariate variance component estimation

| Trait                      | \( \sigma^2_a \) | \( \sigma^2_hys \) | \( \sigma^2_e \) | \( h^2 \) | HPD          |
|----------------------------|-----------------|-------------------|-----------------|----------|--------------|
| Mastitis                   | 0.207           | 0.565             | 1               | 0.074    | 0.050–0.102  |
| Claw ulcers                | 0.274           | 0.455             | 1               | 0.100    | 0.066–0.138  |
| Dermatitis digitalis       | 0.413           | 2.353             | 1               | 0.086    | 0.065–0.113  |
| Interdigital hyperplasia   | 0.661           | 0.766             | 1               | 0.191    | 0.129–0.254  |
| Digital phlegmon           | 0.250           | 0.821             | 1               | 0.081    | 0.041–0.118  |
| White line disease         | 0.066           | 0.519             | 1               | 0.025    | <0.001–0.061 |
| Laminitis                  | 0.083           | 2.351             | 1               | 0.018    | 0.001–0.038  |
| Milk yield (kg)            | 851,069.000     | 684,326.000       | 916,502.000     | 0.347    | 0.316–0.377  |
| Fat-protein ratio          | 0.006           | 0.002             | 0.004           | 0.482    | 0.450–0.511  |
| Milk energy yield (MJ)     | 4,792,681.000   | 5,684,912.000     | 7,328,647.000   | 0.269    | 0.243–0.295  |

1The heritabilities for the disease traits are on the underlying liability scale and for milk production traits on the observed scale; \( \sigma^2_a \) = additive genetic variance; \( \sigma^2_hys \) = herd-year-season variance; \( \sigma^2_e \) = residual variance, which was fixed at a value of 1 for the disease traits; HPD = 95% highest posterior density interval of \( h^2 \).
In total, 50 segments with significant local genetic correlations in 13 regions for all 21 trait combinations between production and disease traits were identified. They are located on BTA 1, 7, 10, 12, 13, 14, 19, 23, 25, and 29. The BTA 14 showed enriched findings with 31 identified segments, which was not surprising due to the substantial effect of DGAT1. All 31 segments fell within 1.49 and 3.30 Mbp of BTA 14, and were thus merged into a single region. The SNP density in our study limited a detailed inference from this region, thus we decided not to overrate this enriched finding. An overview over all regions can be taken from Table 5.

For a combination of a disease and production trait, 14 segments were identified for FPR (all on BTA 14), 15 for MEY, and 21 for MY. For FPR, $Q(R)$ varies between $-401.79$ (FPR-MAS) and $37.32$ (FPR-CDD), and appears with a negative sign for all trait combinations except of FPR-CDD. For MEY, the span is from $-5.54$ (MEY-CIH) to $151.94$ (MEY-CWL), and for MY from $-4.98$ (MY-CUL) to $14.31$ (MEY-CIH). Segments with negative values of scan statistics are located on BTA 13, 23, 25, and 29, and all chromosomes are 1-directional in their direction of $Q(R)$. A positive sign of $Q(R)$ indicates the same unfavorability as a positive global genetic correlation (i.e., rising incidences with increasing production). In contrast, negative signs can be seen as favorable because the incidences decline with decreasing production. Larger values imply a larger shared genetic effect. Mirrored Manhattan plots for MEY-MAS and MY-CIH are shown in Figures 1 and 2 as illustrative examples. These 2 trait combinations inherit highly significant ($P < 0.0005$) segments, whose localization is not limited to BTA 14. It can be seen that single SNPs within the segments do not necessarily have to reach genome-wide significance in the GWAS for each trait, respectively.

**DISCUSSION**

To date, many studies aimed to gain accurate estimates of heritabilities and correlations between disease and production traits in Holstein cattle. Further, genomic analysis has also been performed to detect and fine-map QTL. Combining the estimation of genetic correlations and genomic information while aiming to detect regions that harbor local genetic correlations has not been applied to disease and production traits, to our best knowledge. In this study, we aimed to fill this gap by estimating genetic parameters, conducting

---

**Table 3.** Genetic correlations with their 95% highest posterior density interval of $h^2 (\text{HPD})$ interval in parentheses, results from the bivariate variance component estimation.

| Item               | Milk energy yield | Milk yield | Fat-protein ratio |
|--------------------|-------------------|------------|-------------------|
| Mastitis           | 0.13 (0.04, 0.22)* | 0.18 (0.10, 0.27)* | −0.07 (−0.14, −0.01)* |
| Claw ulcers        | 0.09 (0.01, 0.17)* | 0.08 (<0.01, 0.14) | 0.03 (−0.03, 0.09) |
| Digital dermatitis | <−0.01 (<−0.10, 0.08) | 0.03 (−0.05, 0.12) | −0.04 (−0.10, 0.02) |
| Interdigital hyperplasia | 0.04 (<−0.02, 0.10) | 0.07 (<−0.01, 0.14) | −0.05 (<−0.12, <0.01) |
| Digital phlegmon   | 0.07 (<−0.01, 0.14) | 0.05 (<−0.01, 0.12) | <0.01 (<−0.05, 0.06) |
| White line disease | 0.02 (<−0.04, 0.09) | 0.02 (<−0.07, 0.09) | 0.03 (−0.03, 0.08) |
| Laminitis          | 0.07 (<−0.01, 0.14) | 0.06 (<−0.02, 0.13) | −0.01 (−0.07, 0.04) |

*The superscript letter indicates significant correlations, whose HPD interval does not overlap.

**Table 4.** Genomic inflation factors ($\lambda_{GC}$) of the 4 GWAS methods

| Trait               | $\lambda_{GC}$ |
|---------------------|----------------|
|                     | MLMA | LOCO | PCgenome | PCchromosome |
| Milk yield          | 0.84 | 4.19 | 3.02 | 1.41 |
| Milk energy yield   | 0.96 | 3.80 | 2.83 | 1.25 |
| Fat-protein ratio   | 0.94 | 4.61 | 3.08 | 1.50 |
| Mastitis            | 0.96 | 1.26 | 1.16 | 0.63 |
| Digital phlegmon    | 0.98 | 1.11 | 1.03 | 0.59 |
| White line disease  | 0.96 | 1.06 | 0.98 | 0.59 |
| Interdigital hyperplasia | 0.96 | 1.25 | 1.17 | 0.62 |
| Claw ulcers         | 0.96 | 1.21 | 1.13 | 0.63 |
| Digital dermatitis  | 0.97 | 1.61 | 1.39 | 0.71 |
| Laminitis           | 0.97 | 1.02 | 0.97 | 0.60 |

$^{1}$MLMA = mixed linear model of GCTA; LOCO = “leave-one-chromosome-out” approach, extension of MLMA; PCgenome/PCchromosome = extensions of LOCO, including 20 principal components, computed for the whole genome (PCgenome) or for the chromosome, on which the candidate SNP is located (PCchromosome).
a GWAS, and subsequently detecting genomic regions with significant local genetic correlations and shared genetic effects contrasting 3 milk production with 7 disease traits. We found mostly significant heritabilities, some significant global genetic correlations, significant GWAS hits, and significant local genetic correlations in various chromosomal regions.

Genetic Parameters

The incidences in our study (Table 1) were similar to previously reported results. This includes incidences below 10% for most claw diseases (Heringstad et al., 2018), which are almost doubled for CDD, agreeing with Swalve et al. (2008). The same was found for MAS, where incidences range from 14% (Kelton et al., 1998) to 23% (Olde Riekerink et al., 2008). The average FPR of 1.14 in this study (Table 1) is approximately equal to what is found elsewhere (Bastiaansen et al., 2010; Buttchereit et al., 2010; Negussie et al., 2013), even though we could not dissect differences in early, mid and late lactation. Fat-protein ratio is said to have a peak during the postpartum NEB, declining with ongoing lactation (Negussie et al., 2013).

The heritabilities and genetic correlations were estimated with models using the pedigree-based relationship matrix \(A\). We preferred this over the use of the \(G\) matrix in the models because we were not sure how well the used 50K chip explained the genetic variance of the traits, especially of the disease traits. For the functional

### Table 5. Genomic regions with significant shared effects on a trait combination of a milk production and a disease trait

| Region | BTA | Start (Mbp) | Stop (Mbp) | Trait combination | \(Q(R)\) | \(\theta\) | \(P\)-value |
|--------|-----|-------------|------------|-------------------|--------|--------|---------------|
| 1      | 1   | 120.4794    | 120.4969   | MEY-CWL           | 6.77   | 0.50   | 0.0046        |
| 2      | 7   | 19.7152     | 19.7617    | MEY-MAS, MY-CPH   | 6.38   | 0.60   | 0.0002        |
| 3      | 10  | 3.7498      | 3.9911     | MEY-CHI           | 2.40   | 0.65   | 0.0011        |
| 4      | 10  | 10.1643     | 10.1959    | MEY-CPH, MY-CHI   | 8.22   | 0.40   | 0.0014        |
| 5      | 12  | 20.1803     | 20.4341    | MY-CWL, MY-CIH    | 3.61   | 0.70   | 0.0092        |
| 6      | 13  | 10.0837     | 10.1115    | MY-CPH, MY-CHI    | 2.90   | 0.65   | 0.0036        |
| 7      | 13  | 53.1797     | 53.8098    | MY-CPH, MEY-CLM   | 4.22   | 0.70   | 0.0262        |
| 8      | 13  | 57.5482     | 57.5701    | MY-CPH, MEY-MAS   | 4.54   | 0.70   | 0.0116        |
| 9      | 14  | 1.8695      | 1.8011     | MY-CPH, MEY-CDD   | 6.49   | 0.65   | 0.0002        |
| 10     | 19  | 61.4410     | 61.4784    | MY-CPH, MEY-MAS   | 2.90   | 0.70   | 0.0036        |
| 11     | 23  | 0.0602      | 0.7648     | MY-CUL, MY-CPH    | 1.88   | 0.70   | 0.0002        |
| 12     | 25  | 31.2032     | 31.2783    | MY-CPH, MEY-CUL   | 1.48   | 0.70   | 0.0002        |
| 13     | 29  | 45.1026     | 45.1291    | MY-CUL, MY-CPH    | 5.97   | 0.50   | 0.0044        |

1The regions were identified with LOGODetect; for each region the chromosome, start and stop position in megabase pairs (Mbp), combination of traits, scan statistic \([Q(R)]\), applied \(\theta\), and \(P\)-values are given.

2MEY = milk energy yield; CWL = white line disease; MAS = mastitis; MY = milk yield; CPH = digital phlegmone; CHI = interdigital hyperplasia; CLM = laminitis; CDD = digital dermatitis; FPR = fat-protein ratio; CUL = claw ulcers.
Heritabilities of disease traits were mostly low, which agrees with previous studies (Martin et al., 2018). Only CIH appears with a moderate heritability on the liability scale of 0.19, which coincides with König et al. (2008) and Gernand et al. (2012). The hys variance was large for the disease traits (Table 2), which underlined traits explained, variance might be reduced, as shown by Haile-Mariam et al. (2013).

Figure 1. Mirrored Manhattan plot from GWAS of interdigital hyperplasia (CIH) and milk yield (MY). The solid line corresponds to a false discovery rate (FDR) of 0.1, computed with the method from Benjamini and Yekutieli (2001). There is no threshold for CIH because no SNP was significant at FDR <0.1. The plot for MY was truncated at $-\log_{10}(P) = 10$. Chromosomes are shown from 1 to 29. Pink dots highlight the SNPs that are located in the identified regions from LOGODetect.
the need for the careful data editing to ensure a sufficient number of observations within each hys class.

In the bivariate analysis, MAS showed the highest genetic correlation to production traits, whereas most other genetic correlations between disease and production traits were not significant. This supports the inconsistent results throughout the literature (Her-ingstad et al., 2018). Considering the possibility that

Figure 2. Mirrored Manhattan plot from GWAS of mastitis (MAS) and milk energy yield (MEY). The solid line corresponds to a false discovery rate (FDR) of 0.1, computed with the method from Benjamini and Yekutieli (2001). There is no threshold for MAS because no SNP was significant at FDR <0.1. The plot for MEY was truncated at \(-\log_{10}(P) = 10\). Chromosomes are shown from 1 to 29. Pink dots highlight the SNPs that are located in the identified regions from LOGODetect.
Single Trait-Associated Genomic Regions

Results from our GWAS agreed with previously reported results. Tetens et al. (2013) and Bastiaansen et al. (2010) reported significant SNPs for FPR on BTA 14, 17, and 27. This is complemented by findings for MY on BTA 6 and 14 (van den Berg et al., 2020), claw diseases on all but 9 chromosomes with enriched findings on BTA 8, 9, and 20 (van der Spek et al., 2015), and a significant hit for MAS on BTA 6 (Cai et al., 2018; Freebern et al., 2020).

Increasing the sample size is naturally accompanied by increased power in the detection of significantly associated signals (Schmid and Bennewitz, 2017), but also by inflated SNP effects that might lead to false-positive findings of SNP associations (Lango Allen et al., 2010; Yang et al., 2011b; Liu et al., 2020). Our sample size is sufficient to also detect significant hits for some of the disease traits, but also to generate strong inflation, if no appropriate correction was used. van den Berg et al. (2019) defined genomic inflation factors above 2 as high, but acceptable if their value drops below 1.5. The GWAS results in our study were substantially deflated using the MLMA method, confirming the lack of power due to double fitting of SNPs (Yang et al., 2014). Under the application of the LOCO approach, inflation became substantial unless PC were included as covariates. Their application is considered to be necessary in livestock populations because population stratification appears as a strong confounding bias (van den Berg et al., 2019; Yin and König, 2019). Despite the mentioned advantages of PC in GWAS, Hayes (2013) reminded to pay attention to the uncertainty in the source of variation that they explain. Bulik-Sullivan et al. (2015b) attributed most of inflation in GWAS to polygenicity. Additionally, LD structure, the number of causal variants, the trait heritability (Yang et al., 2011b), population stratification, systemic bias, or a truly strong association with the trait (van den Berg et al., 2019) have been described as potential sources for genomic inflation. Systemic bias in this case refers to a potential overrepresentation and unbalanced sampling of 1 phenotype that evokes a SNPs trait association rather than a true biological association. It is most likely that a combination of polygenic background and systemic bias has caused the inflation found in this study. Most production traits in livestock breeding are influenced by many genes with small effects and inflation factors of these traits are larger. Thus, as soon as inflation is reduced by including PC, one might have corrected for inflation that came from polygenicity and not from external bias, which increases the number of false negative findings.

What limits the inference regarding explicit QTL positions from the present genomic analysis is the low number of markers. This is because most common variants explain only a small proportion of heritability (Visscher and Yang, 2016) and it is unknown how well the applied SNP chip covers the genetic variance of all the traits considered. In addition, rare variants are most likely not in LD with the SNPs on the common SNP chip (Das et al., 2018), but might be important...
especially for disease traits. We assume that genotypic data with higher density, such as 700K chip data, would improve the precision of QTL mapping and of course of mapping shared genomic regions.

Local Correlations and Shared Genomic Regions for Milk Production and Disease Traits

We aimed to detect shared genomic regions and identify their genetic annotation and function to enhance a better understanding of biological pathways that result in the interrelationship between various traits. Global correlation estimates might fail to detect the real interrelationship in cases of strong local, but, in summary, no global correlation. This can be seen at the example of FPR and various disease traits, where all segments are solely located on BTA 14. Next to that, also the antagonistic sign of local estimates causes biased global estimates, shown in bidirectional scan statistics for MEY-CIH, MY-CIH, and MEY-CPH (Table 5). Although the pedigree-based estimation did not reveal significant correlations for all trait combinations (Table 3), at least 1 shared region for every combination was found.

Although a detailed in-depth analysis of QTL positions was not possible, we were successful in the detection of shared genomic regions. By applying LOGODetect to different trait combinations, we were able to identify such regions with significantly correlated effects on disease traits and milk production. Utilizing a sliding window approach is one of the advantages of LOGODetect. Regions with significant genetic sharing might vary substantially in size, depending on the combination of traits and localization in the genome. Thus, applying a fixed window or restricting the analysis to a single genome-wide significant SNP violates this variability (Guo et al., 2021). Moreover, unlike in other tools that are single SNP based (Werme et al., 2022), there is no limitation to distinguish between the true SNPs signal and the signal of any neighboring SNPs. By applying LD as a weighting factor for the marker effects on both traits, LOGODetect avoids unjustified preference of a sequence whose effect is merely LD-based. Originally, LOGODetect had been developed in human genetics, exhibiting substantially lower LD than livestock. Hence, the final choice of $\theta$ had to be considered carefully. We followed the suggestions by Guo et al. (2021) and adapted it to our sample by expanding possible values for $\theta$ toward a value of 1 and applied the grid search via stratified LD score regression. Nevertheless, this method is a trade-off between accuracy and computational burden because the analysis of every possible value between 0 and 1 is not practicable. In the original methods of LOGODetect, the authors (Guo et al., 2021) supposed a splitting of the genome into independent LD blocks. The limited number of markers hindered us to do so as each block would either have consisted of only few markers or would have covered a large part of the genome. Neither of these opportunities fits the idea behind this splitting to achieve more precision and power in the localization of significant regions (Guo et al., 2021; Werme et al., 2022).

Biological Interpretation of Shared Genomic Regions

We found several regions to coincide with previous literature reports. One of them is a hit on BTA 7, harboring shared effects on MEY, MAS, and CPH, where the gene $FBN3$ is located in proximity to $C3$. It affects inflammatory reactions and the abnormal function of the immune system (Cai et al., 2018, 2020). The region for MY-CIH on BTA 25 is nearby a region that Cai et al. (2018) claimed to be located within the genes $ORAI2$, $CORO7$, and $CUX1$. These genes are important for T-cell activation ($ORAI2$), abnormal T-cell differentiation, tumor necrosis ($CORO7$), and enhanced wound healing ($CUX1$; Cai et al., 2018). Interdigital hyperplasia is a disease with enhanced growth in the interdigital area (Heringstad et al., 2018); thus, an abnormal function of skin proliferation might be a reason for this. $Bos taurus$ autosome 13 harbors the genes $HCK$ and $CCM2L$, contributing to innate immunity ($HCK$) and wound healing ($CCM2L$; Cai et al., 2018), which are located close to regions for MY and MAS. The regions on BTA 10 and BTA 19 are located 0.275 Mbp (BTA 10) and 3.86 Mbp (BTA 19) away from 2 variants with a significant trait association to energy balance (BTA 10) and ECM (BTA 19; Krattemacher et al., 2019).

$DGAT1$ showed up with very significant association on almost all trait combinations. These findings are supported by other studies that have, for example, shown pleiotropic effects not only on milk production but also on fertility (Oikonomou et al., 2009) and mastitis (Manga and Riha, 2011; Cai et al., 2020). Unlike the assumption that an increased FPR is accompanied by an aggravated NEB, which makes individuals prone to mastitis, lameness, and metabolic diseases (Heuer et al., 1999), we found effects in the opposite direction. Although the A allele had been detected to increase milk and protein yield and decrease fat yield (Thaller et al., 2003; Bennewitz et al., 2004), it was also found to increase the SCC (Manga and Riha, 2011). Thus, a lower FPR is accompanied by a deterioration in udder health, which ameliorates as soon as FPR is increasing. Presumably, the underlying genes for FPR are different from metabolic pathways (Tetens et al., 2013). This is strengthened by Klein et al. (2019), who did not find an
Effect of DGAT1 on ketosis as a metabolic disease and concluded that it affects FPR via milk composition and not via lipolytic processes during the NEB.

Next to DGAT1, most of the identified regions are located within or nearby genes, whose function in the immune system has been introduced by other studies (Cai et al., 2018, 2020). This supports the aforementioned assumption of NEB being causative for the interrelationship between disease and production traits via the energetic burden cows do face with increasing milk production.

CONCLUSIONS

Taking advantage of the large sample size in this study, it was possible to dissect the genetic background of disease and production traits in more detail. Heritabilities of disease traits were mostly low and, except of mastitis, almost none of them appeared with a significant global correlation to milk production. In the genomic analysis, on the one hand, correcting for relatedness turned out to be crucial to avoid inflation, which was highest for milk production traits, and on the other hand, not to compromise the power by a too stringent correction. Because global genetic correlations suffer (e.g., from antagonistic directions throughout the genome), we aimed to gain a detailed insight in their local correlations. Segments with local genetic correlations were predominantly, but not exclusively found on BTA 14, which was due to DGAT1. The study demonstrates the potential to detect shared genetic regions affecting quantitative traits. This sheds light on the detailed genetic interrelationship between milk production and disease traits, which is an important subject in dairy cattle breeding because it might allow for improved genomic selection decisions based on the informativity and the effect of specific regions on balanced breeding of cows with respect to milk production and disease resistance.

ACKNOWLEDGMENTS

This study is part of the project QTCC: From Quantitative Trait Correlation to Causation in Dairy Cattle and was supported by the Deutsche Forschungsgemeinschaft (DFG, Bonn, Germany; BE3703/15-1). The authors have not stated any conflicts of interest.

REFERENCES

Bastiaansen, J. W. M., H. Bovenhuis, S. Wijga, S. Mc Parland, E. Wall, E. Strandberg, and R. F. Veerkamp. 2010. GenomeWide association study for milk production and fat to protein ratio in dairy cattle. In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production; Leipzig. Accessed Feb. 22, 2022. https://library.wur.nl/WebQuery/wurpubs/fulltext/162631.

Becker, V. A. E., E. Stamer, H. Spikers, and G. Thaller. 2021. Residual energy intake, energy balance, and liability to diseases: Genetic parameters and relationships in German Holstein dairy cows. J. Dairy Sci. 104:10970–10987. https://doi.org/10.3168/jds.2021-20382.

Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29:1165–1188. https://doi.org/10.1214/aos/1013699998.

Benkehlund, J., N. Reinsch, S. Paul, C. Looff, B. Kappe, C. Weimann, G. Erhardt, G. Thaller, C. Kuhl, M. Schwierz, H. Thomsen, F. Reinhardt, R. Reents, and E. Kalm. 2004. The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. J. Dairy Sci. 87:431–442. https://doi.org/10.3168/jds.S0022-0302(04)73182-3.

Bergsten, C. 2003. Causes, risk factors, and prevention of laminitis and related claw lesions. Acta Vet. Scand. 44(Suppl_1):S157. https://doi.org/10.1186/1751-0147-44-S1-S157.

Bulik-Sullivan, B. H., K. Finncane, V. Anttila, A. Gusev, F. R. Day, P. R. Loh, ReproGen Consortium, Psychiatric Genomics Consortium, European consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, L. Duncan, J. R. B. Perry, N. J. Patterson, E. B. Robinson, M. J. Daly, A. L. Price, and B. M. Neale. 2015a. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47:1236–1241. https://doi.org/10.1038/ng.3406.

Bulik-Sullivan, B. K., P. R. Loh, H. K. Finncane, S. Ripke, J. Yang, Psychiatric Genomics Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, and B. M. Neale. 2015b. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47:291–295. https://doi.org/10.1038/ng.3211.

Butcher, E. N., E. Stamer, W. Junge, and G. Thaller. 2010. Evaluation of five lactation curve models fitted for fat:protein ratio of milk and daily energy balance. J. Dairy Sci. 93:1702–1712. https://doi.org/10.3168/jds.2009-2198.

Cai, Z., M. Dusza, B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2020. Distinguishing pleiotropy from linked QTL between milk production traits and mastitis resistance in Nordic Holstein cattle. Genet. Sel. Evol. 52:19. https://doi.org/10.1186/s12711-020-00538-4.

Cai, Z., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. BMC Genomics 19:656. https://doi.org/10.1186/s12864-018-5505-x.

Das, S., G. R. Abecasis, and B. L. Browning. 2018. Genotype imputation from large reference panels. Annu. Rev. Genomics Hum. Genet. 19:73–96. https://doi.org/10.1146/annurev-genom-083117.021602.

Dolecheck, K. A., M. W. Overton, T. B. Mark, and J. M. Bewley. 2019. Use of a stochastic simulation model to estimate the cost per case of digital dermatitis, sole ulcer, and white line disease by parity group and incidence timing. J. Dairy Sci. 102:715–730. https://doi.org/10.3168/jds.2018-14901.

Esposito, G., P. C. Irons, E. C. Webb, and A. Chapwanya. 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. Anim. Reprod. Sci. 144:60–71. https://doi.org/10.1016/j.anireprosci.2013.11.007.

Finncane, H. K., B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P. R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, The RACI Consortium, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, and A. L. Price. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47:1228–1235. https://doi.org/10.1038/ng.3404.

Fleischer, P., M. Metzner, M. Beyerbach, M. Hoedemaker, and W. Klee. 2001. The relationship between milk yield and the incidence of mastitis, almost none of them appeared with a significant global correlation to milk production.
R. J. F. Loos, M. Booneke, M. I. McCarthy, E. Ingelsson, C. M. Lindgren, G. R. Abecasis, K. Stefansson, T. M. Frayling, and J. N. Hirschhorn. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838. https://doi.org/10.1038/nature09410.

Li, X., M. S. Lund, L. James, C. Wang, X. Ding, Q. Zhang, and G. Su. 2017. The patterns of genomic variances and covariances across genome for milk production traits between Chinese and Nordic Holstein populations. BMC Genet. 18:26. https://doi.org/10.1186/s12863-017-0491-9.

Liu, L., J. Zhou, C. J. Chen, J. Zhang, W. Wen, J. Tian, Z. Zhang, and Y. Gu. 2020. GWAS-based identification of new loci for milk yield, fat, and protein in Holstein cattle. Animals (Basel) 10:2048. https://doi.org/10.3390/ani10020128.

Manga, I., and H. Rïha. 2011. The DGAT1 gene K232A mutation is associated with milk fat content, milk yield and milk somatic cell count in cattle (Short Communication). Arch. Tierzucht 54:257–263. https://doi.org/10.5194/aab-54-257-2011.

Martin, P., H. W. Barkema, L. F. Brito, S. G. Narayana, and F. Miglior. 2018. Symposium review: Novel strategies to genetically improve mastitis resistance in dairy cattle. J. Dairy Sci. 101:2724–2736. https://doi.org/10.3168/jds.2017-13554.

Miglior, F., B. L. Muir, and B. J. van Doormaal. 2005. Selection insights of Holstein cattle of various countries. J. Dairy Sci. 88:2161–2170. https://doi.org/10.3168/jds.2005-2728.

Tetens, J., T. H. Meuwissen, and S. Lien. 2009. Characterization of a QT region affecting clinical mastitis and protein yield on BT6. Anim. Genet. 40:701–712. https://doi.org/10.1111/j.1365-2052.2009.01908.x.

Nötzli, B., and H. Mielke. 1995. Vergleich verschiedener Methoden in der Bestimmung des Milchengehaltes beim Schwarzburten Milchrind. J. Anim. Physiol. Anim. Nutr. (Berl.) 73:9–18. https://doi.org/10.1111/j.1365-2052.2008.01789.x.

Olken, K. M., T. Larsen, and K. L. Ingvarson. 2013. Generation of an index for physiological imbalance and its use as a predictor of primary disease in dairy cows during early lactation. J. Dairy Sci. 96:2161–2170. https://doi.org/10.3168/jds.2012-5646.

Negussie, E., I. Strandén, and E. A. Mäntysaari. 2013. Genetic associations of test-day fat:protein ratio with milk yield, fertility, and udder health traits in Nordic Red cattle. J. Dairy Sci. 96:1275–1280. https://doi.org/10.3168/jds.2013-6048.

Tetens, J., T. Tribout, A. B. Gjuvsland, D. Boichard, Ø. Nordbø, M. P. Sanchez, and M. E. Goddard. 2020. Meta-analysis for milk fat and protein percentage using imputed sequence variant genotypes in 94,321 cattle from eight cattle breeds. Genet. Sel. Evol. 52:37. https://doi.org/10.1186/s12711-020-00556-4.

Negus, L. M., T. James, C. C. Roberts, L. E. Shkolnik, J. A. Serre, R. D. Teskey, and M. D. P. Kenney. 2014. Genome-wide association study for claw disorders and trimming status in dairy cattle. J. Dairy Sci. 98:1286–1295. https://doi.org/10.3168/jds.2013-7395.

Schneider et al.: GENOMIC ASSESSMENT: MILK PRODUCTION AND HEALTH TRAITS

van der Spek, D., J. A. M. van Arendonk, and H. Bovenhuis. 2015. Genome-wide association study for claw disorders and trimming status in dairy cattle. J. Dairy Sci. 98:1286–1295. https://doi.org/10.3168/jds.2013-7395.

van Dorp, T. E., J. Dekkers, S. W. Martin, and J. Noordhuizen. 1998. Genetic parameters of health disorders, and relationships with 305-day milk yield and conformation traits of registered Holstein cows. J. Dairy Sci. 18:2264–2270. https://doi.org/10.3168/jds.S0022-0302(98)75806-0.

van Haren, W. J. P. M., and R. F. Veerkamp. 2019. Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. J. Anim. Breed. Genet. 136:418–429. https://doi.org/10.1111/jjbg.12419.

Visscher, P. M., and J. Yang. 2016. A plethora of pleiotropy across complex traits. Nat. Genet. 48:707–708. https://doi.org/10.1038/ng.3389.

Weber, A., E. Stamer, W. Junge, and G. Thaller. 2013. Genetic parameters for lameness and claw and leg diseases in dairy cows. J. Dairy Sci. 96:3310–3318. https://doi.org/10.3168/jds.2012-6261.

Werne, J., S. van der Sluis, D. Posthumus, and C. A. de Leeuw. 2022. An integrated framework for local genetic correlation analysis. Nat. Genet. 54:274–282. https://doi.org/10.1038/s41588-022-01017-y.

Winter, A., W. Kramer, F. A. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitink, J. E. Womack, G. Thaller, and R. Fries. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene
encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. Proc. Natl. Acad. Sci. USA 99:9300–9305. https://doi.org/10.1073/pnas.142293799.

Wu, X.-L., B. Heringstad, Y.-M. Chang, G. de los Campos, and D. Gianola. 2007. Inferring relationships between somatic cell score and milk yield using simultaneous and recursive models. J. Dairy Sci. 90:3508–3521. https://doi.org/10.3168/jds.2006-762.

Xiang, R., I. M. MacLeod, H. D. Daetwyler, G. de Jong, E. O’Connor, C. Schrooten, A. J. Chamberlain, and M. E. Goddard. 2021. Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. Nat. Commun. 12:860. https://doi.org/10.1038/s41467-021-21001-0.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011a. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Yang, J., M. N. Weedon, S. Purcell, G. Lettre, K. Estrada, C. J. Willer, A. V. Smith, E. Ingelsson, J. R. O’Connell, M. Mangino, R. Mägi, P. A. Madden, A. C. Heath, D. R. Nyholt, N. G. Martin, G. W. Montgomery, T. M. Frayling, J. N. Hirschhorn, M. I. McCarthy, M. E. Goddard, and P. M. Visscher. 2011b. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19:807–812. https://doi.org/10.1038/ejhg.2011.39.

Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price. 2014. Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. 46:100–106. https://doi.org/10.1038/ng.2876.

Yin, T., and S. König. 2019. Genome-wide associations and detection of potential candidate genes for direct genetic and maternal genetic effects influencing dairy cattle body weight at different ages. Genet. Sel. Evol. 51:4. https://doi.org/10.1186/s12711-018-0444-4.

Zschiesche, M., A. Mensching, A. R. Sharifi, and J. Hummel. 2020. The Milk fat-to-protein ratio as indicator for ruminal pH parameters in dairy cows: A meta-analysis. Dairy 1:259–268. https://doi.org/10.3390/dairy1030017.

ORCIDS
Helen Schneider ⓒ https://orcid.org/0000-0001-6272-5937
Dierck Segelke ⓒ https://orcid.org/0000-0002-7351-530X
Jens Tetens ⓒ https://orcid.org/0000-0001-5352-461X
Jörn Bennewitz ⓒ https://orcid.org/0000-0001-6450-1160