


## Article

# A Large-Scale Genome-Wide Association Study of Epistasis Effects of Production Traits and Daughter Pregnancy Rate in U.S. Holstein Cattle

Dzianis Prakapenka <sup>1</sup>, Zuoxiang Liang <sup>1</sup>, Jicai Jiang <sup>2</sup>, Li Ma <sup>3</sup>  and Yang Da <sup>1,\*</sup>

<sup>1</sup> Department of Animal Science, University of Minnesota, Saint Paul, MN 55108, USA; prak032@umn.edu (D.P.); zliang@umn.edu (Z.L.)

<sup>2</sup> Department of Animal Science, North Carolina State University, Raleigh, NC 27695, USA; jicai\_jiang@ncsu.edu

<sup>3</sup> Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA; lima@umd.edu

\* Correspondence: yda@umn.edu

**Abstract:** Epistasis is widely considered important, but epistasis studies lag those of SNP effects. Genome-wide association study (GWAS) using 76,109 SNPs and 294,079 first-lactation Holstein cows was conducted for testing pairwise epistasis effects of five production traits and three fertility traits: milk yield (MY), fat yield (FY), protein yield (PY), fat percentage (FPC), protein percentage (PPC), and daughter pregnancy rate (DPR). Among the top 50,000 pairwise epistasis effects of each trait, the five production traits had large chromosome regions with intra-chromosome epistasis. The percentage of inter-chromosome epistasis effects was 1.9% for FPC, 1.6% for PPC, 10.6% for MY, 29.9% for FY, 39.3% for PY, and 84.2% for DPR. Of the 50,000 epistasis effects, the number of significant effects with  $\log_{10}(1/p) \geq 12$  was 50,000 for FPC and PPC, and 10,508, 4763, 4637 and 1 for MY, FY, PY and DPR, respectively, and  $A \times A$  effects were the most frequent epistasis effects for all traits. Majority of the inter-chromosome epistasis effects of FPC across all chromosomes involved a Chr14 region containing *DGAT1*, indicating a potential regulatory role of this Chr14 region affecting all chromosomes for FPC. The epistasis results provided new understanding about the genetic mechanism underlying quantitative traits in Holstein cattle.

**Keywords:** epistasis; GWAS; milk production; fertility



**Citation:** Prakapenka, D.; Liang, Z.; Jiang, J.; Ma, L.; Da, Y. A Large-Scale Genome-Wide Association Study of Epistasis Effects of Production Traits and Daughter Pregnancy Rate in U.S. Holstein Cattle. *Genes* **2021**, *12*, 1089. <https://doi.org/10.3390/genes12071089>

Academic Editor: Ottmar Distl

Received: 11 June 2021

Accepted: 14 July 2021

Published: 18 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Epistasis effects are widely considered important [1–4], but the number of reported epistasis effects is far behind the number of single-point effects [2,4,5]. Genome-wide association study (GWAS) in several dairy cattle breeds have reported many single-SNP effects on dairy traits [6–20]. However, epistasis studies in dairy cattle were limited to candidate gene tests [21–23] or had a small sample size [24]. The U.S. Holstein cattle have uniquely large sample sizes [25,26] and provide an opportunity to identify epistasis effects associated with dairy traits using GWAS. The purpose of this study was to identify pairwise epistasis effects associated with five dairy production traits and one fertility trait using the same Holstein cattle population that we previously used for a large-scale GWAS to detect SNP additive and dominance effects [19]. For each SNP pair, we tested four types of epistasis effects, additive  $\times$  additive, additive  $\times$  dominance, dominance  $\times$  additive, and dominance  $\times$  dominance, and investigated the intra- and inter-chromosome epistasis effects.

## 2. Materials and Methods

### 2.1. Holstein Populations and Genotyping Data

The sample for GWAS analysis contained 294,079 first lactation Holstein cows with phenotypic observations for five milk production traits (milk, fat and protein yields, and

fat and protein percentages), and one fertility trait (daughter pregnancy rate). Daughter pregnancy rate is the percentage of cows that become pregnant during each 21-d period [26]. The number of phenotypic observations ranged from 294,079 for milk yield to 245,214 for daughter pregnancy rate. The 294,079 cows had two sets of SNP genotypes, the 80 K with original and imputed 76,109 SNPs with minor allele frequency of 0.05 and the 60 K with original and imputed 60,671 SNPs. Both the 80 K and 60 K SNPs were based on the same set of SNP chips previously described [19]. The 80 K is used in the current U.S. Holstein genomic evaluation, whereas the 60 K was used for the U.S. Holstein genomic evaluation when we reported SNP additive and dominance effects [19]. Results in this article were based on the 80 K except three inter-chromosome epistasis figures for fat percentage that used the 60 K results because the 60 K had more inter-chromosome epistasis effects for fat percentage than the 80 K. SNP positions were based on the ARS-UCD1.2 bovine genome assembly.

## 2.2. GWAS Analysis

The GWAS analysis used the method of an approximate generalized least squares (AGLS) analysis [19,27]. The AGLS method combines the least squares (LS) tests implemented by EPISNP1mpi [28,29] with the estimated breeding values from routine genetic evaluation using the entire U.S. Holstein population as a covariable. The statistical model was:

$$\mathbf{y} - \mathbf{Z}\tilde{\mathbf{a}} = \mu\mathbf{I} + \mathbf{X}_g\mathbf{g} + \mathbf{Z}\mathbf{a} + \mathbf{e} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (1)$$

where  $\mathbf{y}$  = column vector of phenotypic deviation after removing fixed nongenetic effects such as heard-year-season (termed as 'yield deviation' for any trait) using a standard procedure for the CDCB/USDA genetic and genomic evaluation;  $\tilde{\mathbf{a}}$  = column vector of 2(PTA), PTA = predicted transmission ability from routine genetic evaluation,  $\mu$  = common mean;  $\mathbf{I}$  = identity matrix;  $\mathbf{g}$  = column vector of pairwise genotypic values;  $\mathbf{X}_g$  = model matrix of  $\mathbf{g}$ ;  $\mathbf{b} = (\mu, \mathbf{g}')'$ ,  $\mathbf{X} = (\mathbf{I}, \mathbf{X}_g)$ ;  $\mathbf{a}$  = column vector of additive polygenic values;  $\mathbf{Z}$  = model matrix of  $\mathbf{a}$  = identity matrix if each individual has one observation;  $\mathbf{e}$  = random residuals. The first and second moments of Equation (1) are  $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$ , and  $\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \sigma_a^2\mathbf{Z}\mathbf{A}\mathbf{Z}' + \sigma_e^2\mathbf{I}$ , where  $\sigma_a^2$  = additive variance,  $\mathbf{A}$  = additive relationship matrix, and  $\sigma_e^2$  = residual variance. Equation (1) achieves the benefit of sample stratification correction from mixed models without the computing difficulty of inverting  $\mathbf{V}$  or  $\mathbf{A}$ . Four types of epistasis effects were tested: additive  $\times$  additive ( $A \times A$ ), additive  $\times$  dominance ( $A \times D$ ), dominance  $\times$  additive ( $D \times A$ ), and dominance  $\times$  dominance ( $D \times D$ ). The significance tests used the t-tests of the epistasis contrasts of the estimated SNP genotypic values [10,30], and the t-statistic was calculated as:

$$t_j = \frac{|L_j|}{\sqrt{\text{var}(L_j)}} = \frac{|\mathbf{s}_j\hat{\mathbf{g}}|}{v\sqrt{\mathbf{s}_j(\mathbf{X}'\mathbf{X})_{gg}^{-1}\mathbf{s}_j'}}, \quad j = A \times A, A \times D, D \times A, D \times D \quad (2)$$

where  $L_j$  = epistasis contrast,  $\sqrt{\text{var}(L_j)}$  = standard deviation of the additive or dominance contrast,  $v^2 = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})/(n - k)$  = estimated residual variance;  $\hat{\mathbf{g}}$  = column vector of the AGLS estimates of the pairwise SNP genotypic effects  $(\mathbf{X}'\mathbf{X})_{gg}^{-1}$  = submatrix of  $(\mathbf{X}'\mathbf{X})^{-1}$  corresponding to  $\hat{\mathbf{g}}$ . A limitation in EPISNP1mpi causes a p value less than  $10^{-308}$  to be printed as '0'. For such p values, empirical p values as a power function of the observed t-values were calculated using the empirical formula of  $p = 0.2416(t^{1.9713})$  [19]. For 76,109 SNPs and eight traits, the total number of pairwise tests was  $2,896,251,886 \times 4 \times 8 = 92,680,060,352$ . The threshold p value with the Bonferroni correction for 0.05 genome-wide false positives was  $p = 5.39(10^{-13})$ , and a pairwise epistasis effect was declared significant if  $\log_{10}(1/p) > 12.26 \approx 12$ . The Manhattan plots of observed  $\log_{10}(1/p)$  values for each trait were made using SNPEVG2 version 3.2 [31] for a global view of statistical significance, and the circular plots for global views of intra- and inter-chromosome epistasis

effects were created using CIRCOS version 0.69-9 [32,33]. The 29 autosomes and the X chromosome had 1225–4121 SNPs per chromosome. These SNPs had 104,203,650 intra-chromosome SNP pairs and 2,791,287,201 inter-chromosome SNP pairs, or, 3.6% of all SNP pairs were intra-chromosome and 96.4% were inter-chromosome. The ratio of  $(A \times A):(A \times D + D \times A):(D \times D)$  tests was 1:2:1. These percentages and ratios were used for comparison with the observed percentages and ratios.

### 3. Results

#### 3.1. Overview of Epistasis Effects

The pairwise epistasis analysis focused on the top 50,000 pairs of epistasis effects for each trait. All 50,000 SNP pairs of fat and protein percentages had significant epistasis effects; milk yield had 10,485 (21.5%), fat yield had 4746, (10.4%), protein yield had 4629 (15.4%) significant epistasis effects, and daughter pregnancy rate only had one significant epistasis effect (Table 1). The six traits had sharp contrasts in percentages of intra- and inter-chromosome epistasis effects and in the distribution of  $A \times A$ ,  $A \times D$ ,  $D \times A$  and  $D \times D$  effects among the top 50,000 SNP pairs. Fat percentage (98.1%) and protein percentages (98.4%) had the highest frequencies of intra-chromosome epistasis effects, followed by milk yield (89.4%), fat yield (70.1%), protein yield (61.7%), and daughter pregnancy rate (15.8%) (Table 1). The  $A \times A$  effects were the most frequent effects (Table 2). A major difference between daughter pregnancy rate and the five production traits was that inter-chromosome  $A \times A$  effects were the most frequent  $A \times A$  effects, 52.9% of the 50,000 effects for this trait. In contrast, protein yield that had the highest frequency of inter-chromosome epistasis effects among the five production traits had 34.9% of inter-chromosome  $A \times A$  effects (Table 2). A global view of statistical significance of the top 50,000 epistasis effects for each trait is shown in Figure 1 for intra-chromosome epistasis effects and in Figure 2 for inter-chromosome epistasis effects. Significant epistasis effects for the five production traits are shown in Figure S1 for intra-chromosome epistasis, and in Figure S2 for inter-chromosome epistasis. Selected tests results including locations and  $\log_{10}(1/p)$  values of significant epistasis effects are given in Table S1 for intra-chromosome epistasis effects, and Table S2 for inter-chromosome epistasis effects.

**Table 1.** Frequencies and statistical significance of intra- and inter-chromosome pairwise epistasis effects among the top 50,000 epistasis effects.

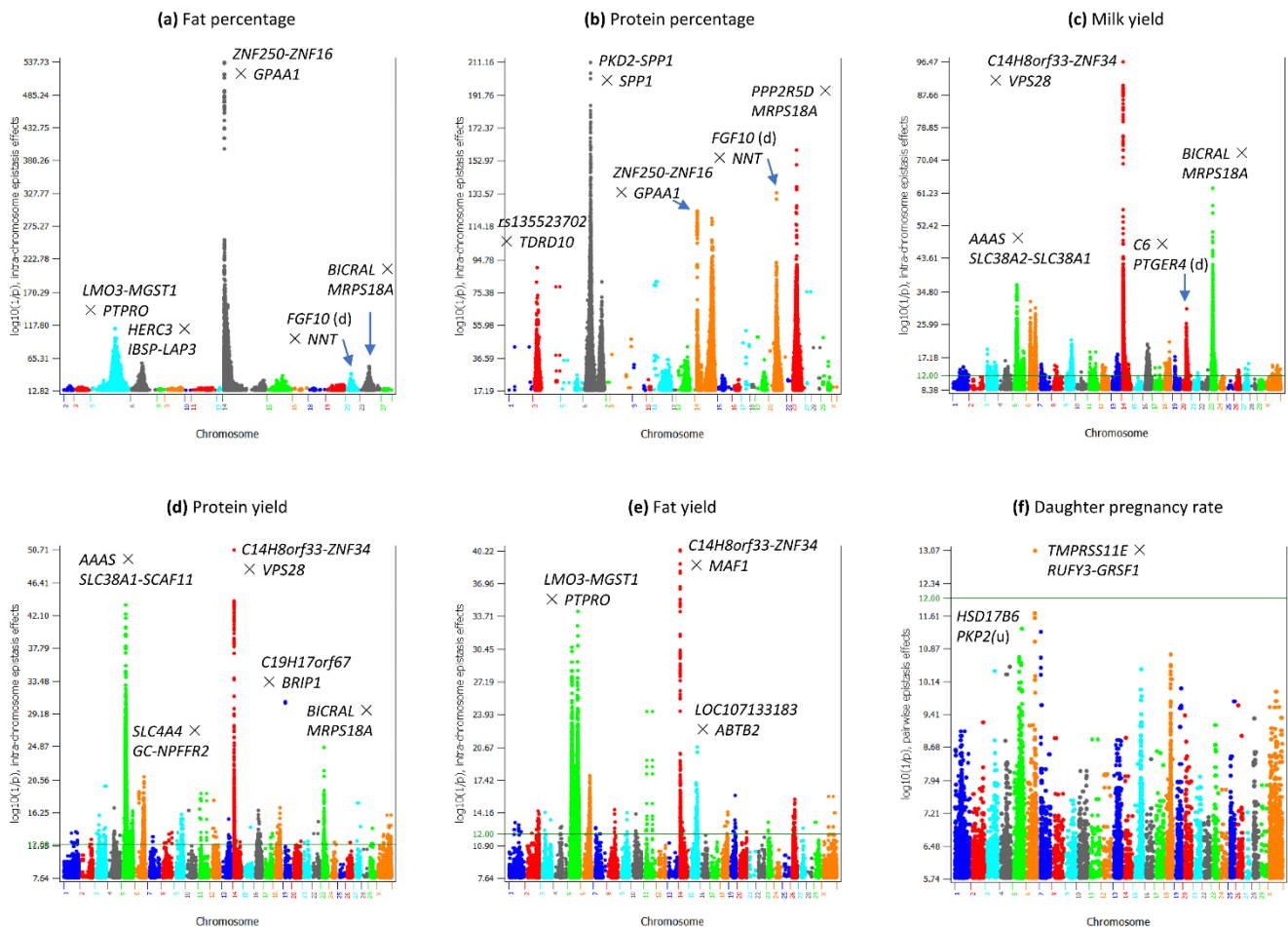
	Intra-Chromosome Epistasis (% of Intra-Chromosome SNP Pairs: 3.6)			Inter-Chromosome Epistasis (% of Inter-Chromosome SNP Pairs: 96.4)		
	Freq (%)	Log <sub>10</sub> (1/p)	Pairs with log <sub>10</sub> (1/p) ≥ 12	Freq (%)	Log <sub>10</sub> (1/p)	Pairs with log <sub>10</sub> (1/p) ≥ 12
FPC	98.1	13–537	49,046 (98.1%)	1.9	13–23	954 (1.9%)
PPC	98.4	17–211	49,206 (98.4%)	1.6	17–82	794 (1.6%)
MY	89.4	9–97	10,440 (20.9%)	10.6	13–19	68 (0.6%)
FY	70.1	7–40	4715 (9.4%)	29.9	7–14	48 (1.0%)
PY	61.7	7–51	4325 (8.7%)	39.3	7–21	312 (6.7%)
DPR	15.8	6–13	1 (0.002%)	84.2	6–11	0

Freq = frequency. FPC = fat percentage. PPC = protein percentage. MY = milk yield. FY = fat yield. PY = protein yield.

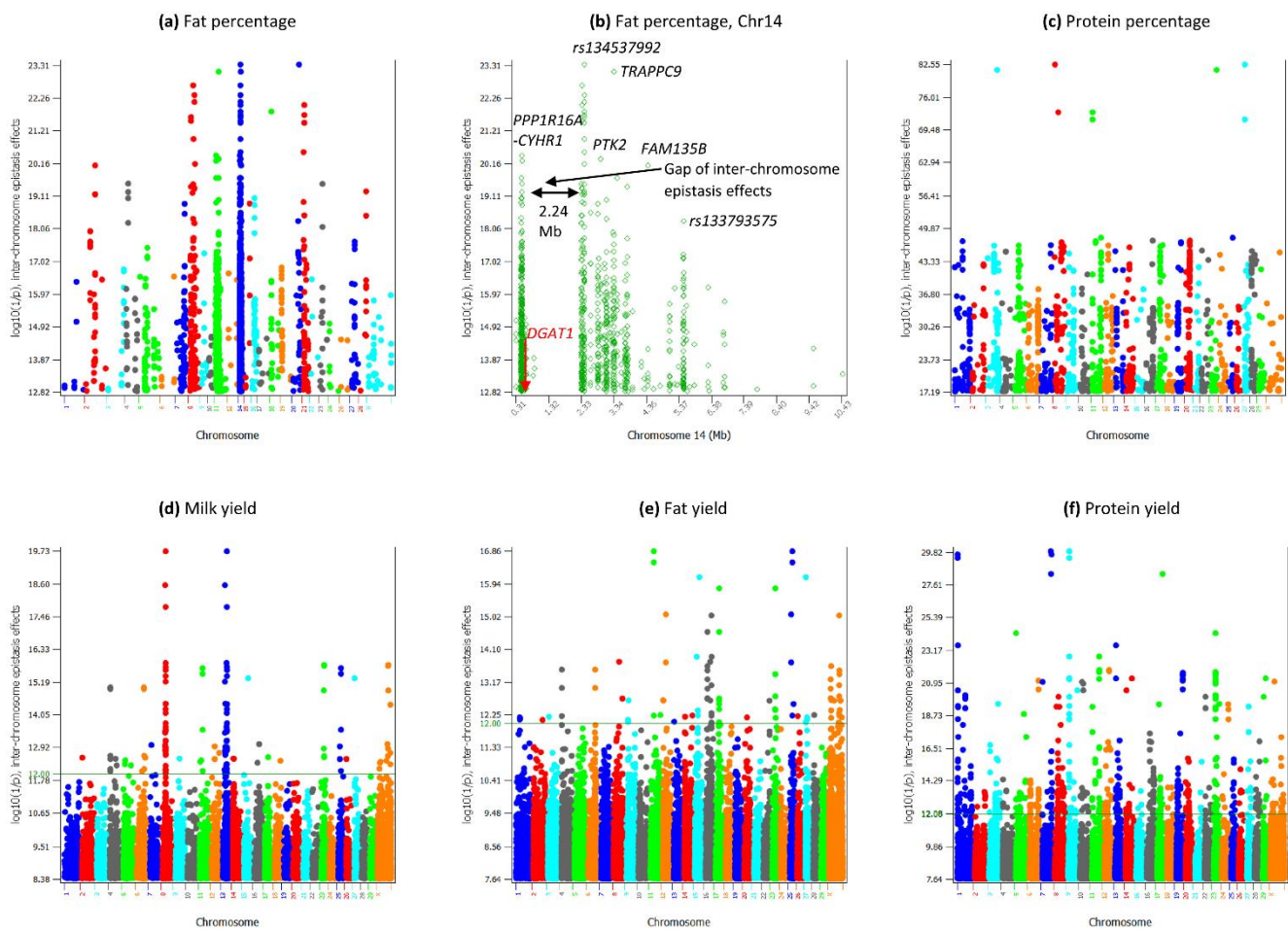
**Table 2.** Frequencies of A × A, A × D/D × A and D × D epistasis effects within intra- and inter-chromosome pairwise epistasis effects among the top 50,000 epistasis effects.

		Intra-Chromosome Epistasis			Inter-Chromosome Epistasis		
		A × A	A × D and D × A	D × D	A × A	A × D and D × A	D × D
FPC	Count	47,752	1023	271	952	0	1
	%	95.5	2.1	0.5	1.9	0.0	0.0
PPC	Count	45,379	2936	1116	10	3	780
	%	90.8	5.9	2.2	0.0	0.0	1.6
MY	Count	36,387	5978	2355	5197	35	48
	%	72.8	12.0	4.7	10.4	0.1	0.1
FY	Count	30,225	3767	1401	14,317	188	92
	%	60.5	7.5	2.8	28.6	0.4	0.2
PY	Count	20,582	7461	2840	17,468	200	1448
	%	41.2	14.9	5.7	34.9	0.4	2.9
DPR	Count	7122	602	196	26,430	10,531	5118
	%	14.2	1.2	0.4	52.9	21.1	10.2

FPC = fat percentage. PPC = protein percentage. MY = milk yield. FY = fat yield. PY = protein yield.



**Figure 1.** Manhattan plots of statistical significance of intra-chromosome epistasis effects for five milk production traits and daughter pregnancy rate. (a): Fat percentage. (b): Protein percentage. (c): Milk yield. (d): Protein yield. (e): Fat yield. (f): Daughter pregnancy rate.



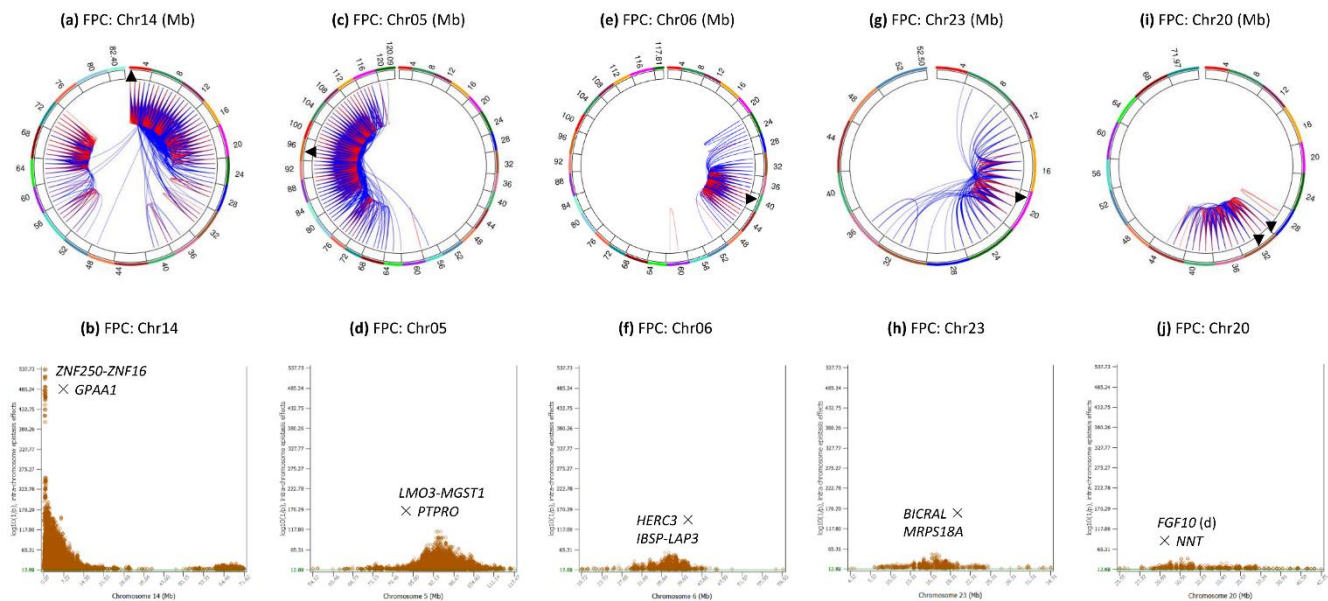
**Figure 2.** Manhattan plots of statistical significance of inter-chromosome epistasis effects for milk production traits. (a): Fat percentage. (b): Inter-chromosome epistasis effects of the 0.3–10.5 Mb region of Chr14 for fat percentage. (c): Protein percentage. (d): Milk yield. (e): Fat yield. (f): Protein yield.

### 3.2. Intra-Chromosome Epistasis Effects of Fat Percentage

Five chromosomes had the most significant intra-chromosome epistasis effects for fat percentage, chromosomes 14, 5, 6, 23 and 20, and 13 chromosomes did not have significant intra-chromosome epistasis effects for fat percentage, chromosomes 1, 4, 7, 12, 17, 21, 22, 24, 25, 26, 28, 29 and X (Figure 1a). The five chromosomes with the highly significant epistasis effects also had highly significant SNP effects [19]. This trait had the highest frequency of  $A \times A$  effects, 97.4% of the top 50,000 SNP pairs had  $A \times A$  effects, and the second highest frequency of intra-chromosome effects, 98.1% of the top 50,000 pairs had intra-chromosome epistasis effects (Table 2).

Chr14 had intra-chromosome epistasis effects in two regions: the 0.25–40 Mb region and the 52–74 Mb region (Figure 3a), but the latter was not among the most significant regions (Figures 1a and 3b). These two regions had the largest number of intra-chromosome epistasis effects, 27,806 of the 49,046 pairs, or 56.7% of all intra-chromosome epistasis effects, but the 0.3–40 Mb region had the most significant and the largest number of effects. Of the top 82 most significant intra-chromosome epistasis effects, 76 pairs each was between two SNPs flanking *DGAT1*. The genes with such SNPs upstream of *DGAT1* included *ZNF250*, *ZNF16*, *C14H8orf33*, *ZNF34*, *CYHR1*, *VPS28*; the genes downstream of *DGAT1* included *GPAA1*, *MAF1*, *MROH1* and *PARP10*. The 2.86–3.77 Mb region with *PTK2*, *AGO2*, *CHRAC1*, *TRAPPC9* and *KCNK9* had the highest concentration of intra-chromosome epistasis effects, 10,963 pairs out of the 27,806 Chr14 SNP pairs, or 39.4% of the Chr14 intra-chromosome epistasis effects. The most significant intra-chromosome epistasis effect was between

*rs110508680* (*ZNF250-ZNF16*) and *rs110929299* in *GPA1*. The *DGAT1* gene was not heavily involved in intra-chromosome epistasis effect for any of the five milk production traits, with #2245 ranking and 120 epistasis effects for fat percentage, although *DGAT1* had the most significant SNP effects for all five milk production traits in the same dataset [19]. As to be described, *DGAT1* also was not heavily involved in inter-chromosome epistasis effects.



**Figure 3.** Chromosomes with most significant intra-chromosome epistasis effects for fat percentage. The top row shows the circular plots with red color indicating intra-4 Mb epistasis effect and blue color indicating inter-4 Mb epistasis effects; the bottom row shows the Manhattan plots of  $\log_{10}(1/p)$  values. The black arrow indicates the location of the peak epistasis effect of the chromosome. Gene name indicates the gene with a peak SNP epistasis effect, and ‘-’ indicates the SNP was between the two genes. FPC = fat percentage. (a): Intra-chromosome epistasis effects of Chr14 for fat percentage. (b): Statistical significance of intra-chromosome epistasis effects of Chr14 for fat percentage. (c): Intra-chromosome epistasis effects of Chr05 for fat percentage. (d): Statistical significance of intra-chromosome epistasis effects of Chr05 for fat percentage. (e): Intra-chromosome epistasis effects of Chr06 for fat percentage. (f): Statistical significance of intra-chromosome epistasis effects of Chr06 for fat percentage. (g): Intra-chromosome epistasis effects of Chr23 for fat percentage. (h): Statistical significance of intra-chromosome epistasis effects of Chr23 for fat percentage. (i): Intra-chromosome epistasis effects of Chr20 for fat percentage. (j): Statistical significance of intra-chromosome epistasis effects of Chr20 for fat percentage.

Chr05 had intra-chromosome epistasis effects in the 64–116 Mb region (Figure 3c), about half of the chromosome, with the second largest number of intra-chromosome epistasis effects, 17,057 epistasis effects, or 34.8% of all intra-chromosome epistasis effects. The most significant intra-chromosome epistasis effect of this chromosome was between *rs41602750* (*LMO3-MGST1*) and *rs136373712* in *PTPRO* (Figure 3d). We reported this region to have the most significant SNP effects of fat percentage among all chromosomes except Chr14 [19].

Chr06 had intra-chromosome epistasis effects mostly in the 32–44 Mb region (Figure 3e) with peak effect between *rs109552132* in *HERC3* and *rs109263339* (*IBSP-LAP3*) (Figure 3f). The 32–44 Mb region had significant SNP effects for all five production traits [19]. The top 100 pairs were located in the 30.15–36.70 region with *HERC3*, *IBSP*, *LAP3*, *LCORL*, *SPP1*, *MEPE*, *PKD2*, *CCSER1*, and *ABCG2*.

Chr23 had intra-chromosome epistasis effects in the 8–24 Mb region with peak with peak effects in the 16–17 Mb region (Figure 3g), and the most significant effect on this chromosome was between *rs109288796* in *BICRAL* and *rs109567499* in *MRPS18A* (Figure 3h). The Chr23 epistasis effects were mostly D × D, A × D and D × A effects, and most of the D × D effects involved genotype combinations with small frequencies of < 0.0001

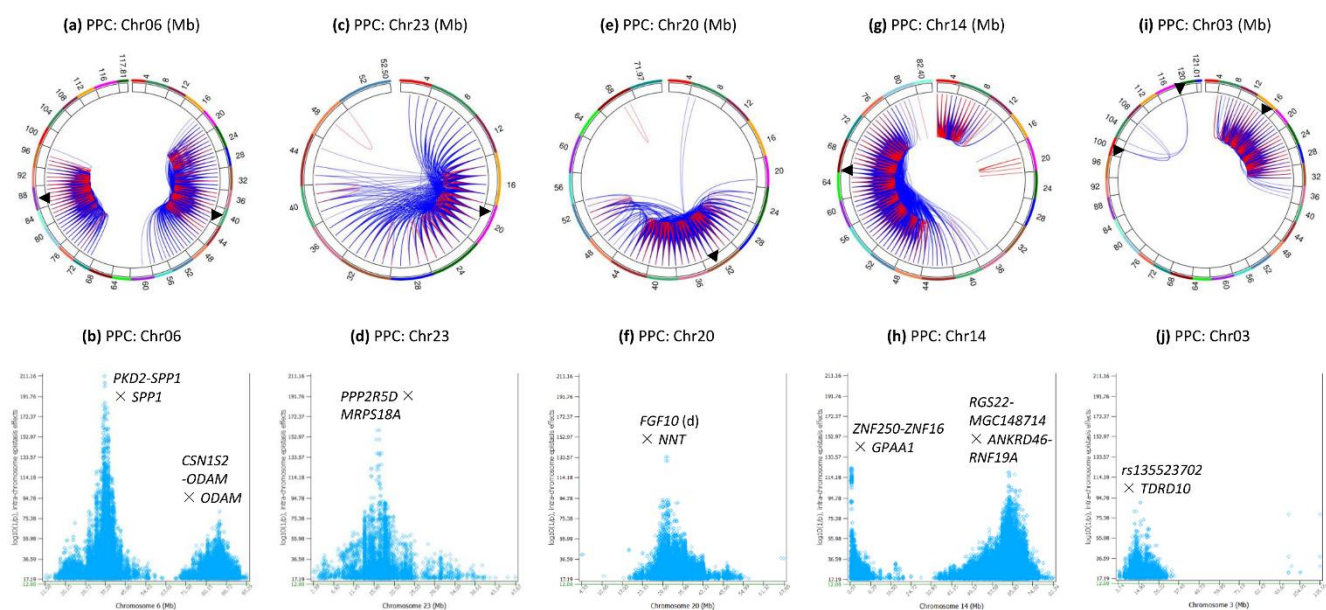
(Table S1). These effects were in the same region where significant SNP dominance effects were detected [19].

Chr20 had intra-chromosome epistasis effects in the 25–43 Mb region (Figure 3i) with peak effect between *rs134342292* about 10 Kb downstream of *FGF10* and *rs41580285* in *NNT* (Figure 3j). The 25–43 Mb region contained the *GHR-C6-PRLR* region that had highly significant SNP effects for all five production traits [19].

### 3.3. Intra-Chromosome Epistasis Effects of Protein Percentage

Five chromosomes had the most significant intra-chromosome epistasis effects for protein percentage, chromosomes 14, 6, 23, 20 and 3; six chromosomes did not have such effects for protein percentage, chromosomes 2, 4, 21, 24, 25 and 26 (Figure 1b). The five chromosomes with the highly significant epistasis effects also had highly significant SNP effects [19]. This trait had the second highest frequency of  $A \times A$  effects, 97.4% of the top 50,000 SNP pairs had  $A \times A$  effects, and the second highest frequency of intra-chromosome effects, 98.4% of the top 50,000 pairs had intra-chromosome epistasis effects (Table 2).

Chr06 had intra-chromosome epistasis effects in two regions: the 16–60 Mb region (approximately the same region with epistasis effects for fat percentage) and the 70–96 Mb region (Figure 4a). The most significant effects of all chromosomes were located in the 30.64–38.33 Mb region of Chr06 with *PKD2*, *SPP1*, *MEPE*, *IBSP*, *HERC3*, *ABCG2*; whereas the most significant effects within the 70–96 Mb region were located in 84.83–86.07 Mb with *CSN1S2*, *ODAM*, *CSN1S1*, *CSN2*, *LOC526979* (Figure 4b).



**Figure 4.** Chromosomes with most significant intra-chromosome epistasis effects for protein percentage. The top row shows the circular plots with red color indicating intra-4 Mb epistasis effect and blue color indicating inter-4 Mb epistasis effects; the bottom row shows the Manhattan plots of  $\log_{10}(1/p)$  values. The black arrow indicates the location of the peak epistasis effect of the chromosome. Gene name indicates the gene with a peak SNP epistasis effect, and ‘-’ indicates the SNP was between the two genes. PPC = protein percentage. (a): Intra-chromosome epistasis effects of Chr06 for protein percentage. (b): Statistical significance of intra-chromosome epistasis effects of Chr06 for protein percentage. (c): Intra-chromosome epistasis effects of Chr23 for protein percentage. (d): Statistical significance of intra-chromosome epistasis effects of Chr20 for protein percentage. (e): Intra-chromosome epistasis effects of Chr20 for protein percentage. (f): Statistical significance of intra-chromosome epistasis effects of Chr14 for protein percentage. (g): Intra-chromosome epistasis effects of Chr14 for protein percentage. (h): Statistical significance of intra-chromosome epistasis effects of Chr14 for protein percentage. (i): Intra-chromosome epistasis effects of Chr03 for protein percentage. (j): Statistical significance of intra-chromosome epistasis effects of Chr03 for protein percentage.

Chr23 had intra-chromosome epistasis effects in 12–24 Mb regions (Figure 4c), with peak effects in 13.85–18.61 Mb in or near *PPP2R5D*, *MRPS18A*, *BICRAL*, *FOXP4*, *SUPT3H*, *MOCS1*, *CCND3*, approximately the same region with significant epistasis and SNP dominance effects for fat percentage (Figure 4d). The Chr23 epistasis effects were mostly D × D, A × D and D × A effects, and most of the D × D effects involved genotype combinations with small frequencies < 0.0001.

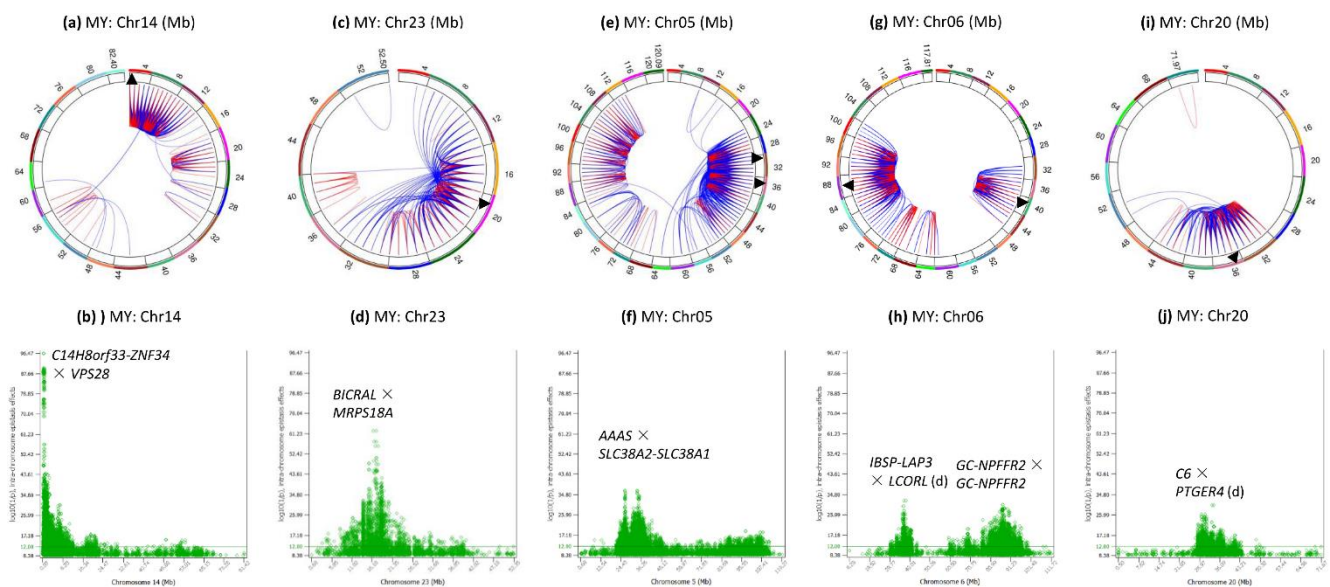
Chr20 had intra-chromosome epistasis effects in the 34–52 Mb region (Figure 4e) with peak effects in the 30.64–31.17 Mb of Chr20 involving a SNP 42 Kb downstream of *FGF10* and a SNP in *NNT* (Figure 4f).

Chr14 had intra-chromosome epistasis effects in two regions, the 0.25–20 Mb and the 36–82.4 Mb regions (Figure 4g). Peak intra-chromosome epistasis effects were in the 0.25–0.81 Mb region involving SNPs in or near *ZNF250*, *ZNF16*, *GPAA1*, *MAF1*, *CYHR1*, *MROH1*, *VPS28*; in the 64.36–67.10 Mb region in or near *RGS22*, *MGC148714*, *VPS13B*, *OSR2*, and *CPQ* (Figure 4h).

Chr03 had regions had intra-chromosome epistasis effects mostly in the 4–30 Mb region (Figure 4i), with peak effects in the 15.9–16.0 Mb region in or near *TDRD10*, *KCNN3*, *ADAR*; between *rs135012389* at 98.6 Mb and *rs136744726* at 116.1 Mb in *LOC101904767* (Figure 4j).

### 3.4. Intra-Chromosome Epistasis Effects of Milk Yield

Chromosomes 14, 5, 6, 23 and 20 had the most significant intra-chromosome epistasis effects (Figures 1c and 5), and all the chromosome regions with peak epistasis effects also had peak SNP effects [19].



**Figure 5.** Chromosomes with most significant intra-chromosome epistasis effects for milk yield. The top row shows the circular plots with red color indicating intra-4 Mb epistasis effect and blue color indicating inter-4 Mb epistasis effects; the bottom row shows the Manhattan plots of  $\log_{10}(1/p)$  values. The black arrow indicates the location of the peak epistasis effect of the chromosome. Gene name indicates the gene with a peak SNP epistasis effect, and ‘-’ indicates the SNP was between the two genes. MY = milk yield. (a): Intra-chromosome epistasis effects of Chr14 for milk yield. (b): Statistical significance of intra-chromosome epistasis effects of Chr14 for milk yield. (c): Intra-chromosome epistasis effects of Chr23 for milk yield. (d): Statistical significance of intra-chromosome epistasis effects of Chr23 for milk yield. (e): Intra-chromosome epistasis effects of Chr05 for milk yield. (f): Statistical significance of intra-chromosome epistasis effects of Chr05 for milk yield. (g): Intra-chromosome epistasis effects of Chr06 for milk yield. (h): Statistical significance of intra-chromosome epistasis effects of Chr06 for milk yield. (i): Intra-chromosome epistasis effects of Chr20 for milk yield. (j): Statistical significance of intra-chromosome epistasis effects of Chr20 for milk yield.



Chr14 had 2976 or 28.51% of the 10,440 intra-chromosome epistasis effects for milk yield mostly in the 0.25–12 Mb region (Figure 5a), with peak effect between *rs109752439* and *VPS28* (Figure 5b) that was also the peak effect for protein yield (Figure 1e). However, *DGAT1* that had the most significant SNP effects for all five production traits [19] only had five intra-chromosome epistasis effects with the best ranking of #6140 for milk yield and had no significant intra-chromosome epistasis effects for protein and fat yields.

Chr23 had intra-chromosome epistasis effects mostly in the 10–30 Mb region (Figure 5c), with peak D × D effects between two SNPs in *BICRAL* and *MRPS18A* in the 16.4–17.2 Mb region (Figure 5d). The 13.85–19.79 Mb region had a large number of D × D, D × A or A × D effects of milk and protein yields (Table S1) and was approximately the same region where significant SNP dominance effects were detected [19].

Chr05 had intra-chromosome epistasis effects mostly in the 24–50 and 84–112 Mb regions for milk yield (Figure 5e) with peak D × A effect between the dominance effect of a SNP in *AAAS* and the additive effect of a SNP between *SLC38A2* and *SLC38A1* (Figure 5f). In the 24–50 Mb region, majority of the epistasis effects were D × A, A × D and D × D. The chromosome region around the peak epistasis effect had highly significant dominance effects in or near *ATF7*, *AAAS*, *EIF4B* and *PLXNC1* [19].

Chr06 had two regions with intra-chromosome epistasis effects for the milk yield: the 30–46 Mb and the 62–102 Mb (Figure 5g), with peak effects between *rs109263339* (*IBSP-LAP3*) and *rs110012183* downstream of *LCORL* in the first region, and between a SNP (*rs109452259*) flanked by *GC* and *NPFRR2* and multiple SNPs in *SLC4A4* in the second region (Figure 5h). The *SLC4A4-GC-NPFRR2* region had the most significant SNP effects on milk and protein yields outside Chr14 [19].

Chr20 had intra-chromosome epistasis effects in the 30–52 Mb region for milk yield (Figure 5i) with peak effect between a SNP in *C6* and a SNP in *PTGER4* (Figure 5j). The 25–43 Mb region contained the *GHR-C6-PRLR* region that had highly significant SNP effects for milk and protein yields [19].

### 3.5. Intra-Chromosome Epistasis Effects of Protein Yield

Chromosomes 14, 5, 6, 23 and 20 had the most significant intra-chromosome epistasis effects (Figures 1d and 6), and all the chromosome regions with peak epistasis effects also had peak SNP effects [19]. Chromosome 28 was the only chromosome without intra-chromosome epistasis effects with  $\log_{10}(1/p) \geq 12$  (Figure 1d).

Chr14 had intra-chromosome epistasis effects for protein yield mostly in the 0.25–7 Mb region (Figure 6a), with peak effect between *rs109752439* and *VPS28* (Figure 6b) that was also the peak effect for milk yield (Figure 5b).

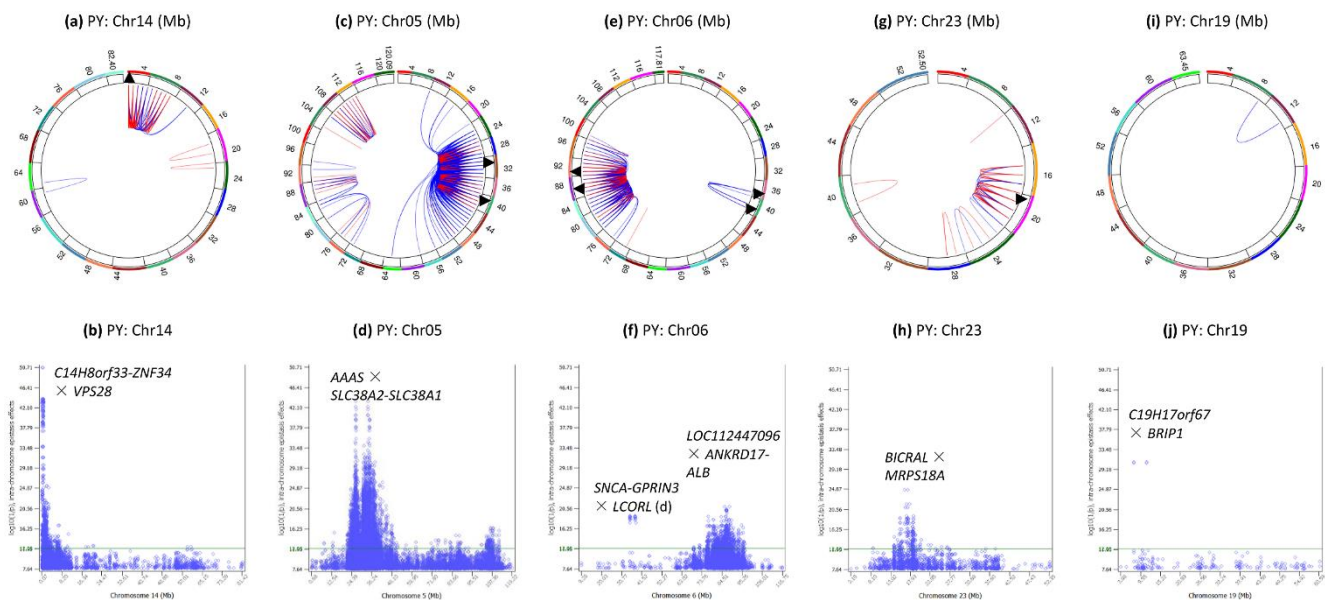
Chr05 had intra-chromosome epistasis effects mostly in the 24–40 Mb for protein yield (Figure 6c) with peak D × A effect between the dominance effect of a SNP in *AAAS* and the additive effect of a SNP between *SLC38A2* and *SLC38A1* (Figure 6d). The same SNP pair also had peak D × A for milk yield. As for milk yield, majority of the epistasis effects in the 24–40 Mb region were D × A, A × D and D × D. Genes with highly significant intra-chromosome epistasis effects near the peak effects included *ATF7*, *AAAS*, *SLC38A1*, *SCAF11* and *ANO6*.

Chr06 had two regions with intra-chromosome epistasis effects for protein yield at 35.1–38.3 and 76–96 Mb (Figure 6e), with peak effects between *rs1095336909* (*SNCA-GPRIN3*) and *rs110718778* downstream of *LCORL* in the first region, and between *rs109452259* in *LOC112447096* and *rs135020479* flanked by *ANKRD17* and *ALB* (Figure 6f). A SNP (*rs109452259*) between *GC* and *NPFRR2* and five SNPs in *SLC4A4* had the #2–7 peak effects of the chromosome, noting that this region also had highly significant SNP effects for protein yield [19].

Chr23 had intra-chromosome epistasis effects mostly in the 14–20 Mb region (Figure 6g), with peak D × D effects between two SNPs in *BICRAL* and *MRPS18A* in the 16.4–17.2 Mb region (Figure 6h), the same SNP pair with peak D × D for milk yield (Figure 5d). The 13.85–19.79 Mb region had a large number of D × D, D × A or A × D effects of milk and

protein yields (Table S1) and was approximately the same region where significant SNP dominance effects were detected [19].

Chr19 only had two pairs of highly significant intra-chromosome epistasis effect between *rs110881085* in *C19H17orf67* and two SNPs in *BRIP1* and *INTS2* (Figures 1d and 6i,j).



**Figure 6.** Chromosomes with most significant intra-chromosome epistasis effects for protein yield. The top row shows the circular plots with red color indicating intra-4 Mb epistasis effect and blue color indicating inter-4 Mb epistasis effects; the bottom row shows the Manhattan plots of  $\log_{10}(1/p)$  values. The black arrow indicates the location of the peak epistasis effect of the chromosome. Gene name indicates the gene with a peak SNP epistasis effect, and ‘-’ indicates the SNP was between the two genes. PY = protein yield. (a): Intra-chromosome epistasis effects of Chr14 for protein yield. (b): Statistical significance of intra-chromosome epistasis effects of Chr14 for protein yield. (c): Intra-chromosome epistasis effects of Chr05 for protein yield. (d): Statistical significance of intra-chromosome epistasis effects of Chr05 for protein yield. (e): Intra-chromosome epistasis effects of Chr06 for protein yield. (f): Statistical significance of intra-chromosome epistasis effects of Chr06 for protein yield. (g): Intra-chromosome epistasis effects of Chr23 for protein yield. (h): Statistical significance of intra-chromosome epistasis effects of Chr23 for protein yield. (i): Intra-chromosome epistasis effects of Chr19 for protein yield. (j): Statistical significance of intra-chromosome epistasis effects of Chr19 for protein yield.

### 3.6. Intra-Chromosome Epistasis Effects of Fat Yield

Chromosomes 14, 5, 15, 6 and 26 had the most significant intra-chromosome epistasis effects (Figures 1e and 7). Nine chromosomes did not have intra-chromosome epistasis effects with  $\log_{10}(1/p) \geq 12$ , chromosomes 12, 13, 16, 17, 21, 22, 24, 25, and 28.

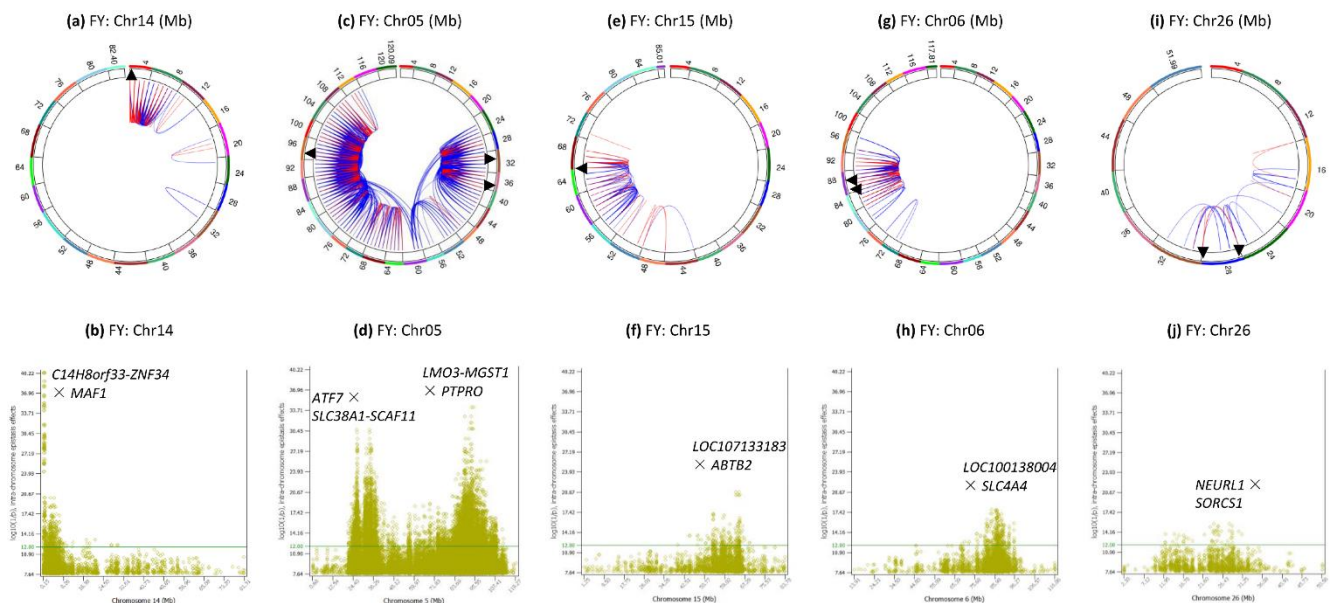
Chr14 had 259 intra-chromosome epistasis effects for fat yield mostly in the 0.24–8.11 Mb region (Figure 7a), with the most significant effects between *rs109752439* (*C14H8orf33-ZNF34*) and *MAF1* (Figure 7b) among all chromosomes. This 0.24–8.11 Mb region had 36 of the top 66 intra-chromosome epistasis effects and each of the 36 effects involved two SNPs on both sides of *DGAT1*, which did not have any significant intra-chromosome epistasis effect for fat yield.

Chr05 had 4111 or 86.31% of the 4763 intra-chromosome epistasis effects in the 22–112 Mb region that covered 93% of the chromosome (Figure 7c). The most significant effect was between *rs136373712* in *PTPRO* and *rs41602750* (*LMO3-MGST1*) (Figure 7d), and the 89.3–94.38 Mb region with *PTPRO*, *LMO3*, *MGST1*, *PDE3A*, *SLC15A5*, *DERA*, *PIK3C2G*, *RERGL* and *PLEKHA5* had the most significant  $A \times A$  effects on this chromosome. The epistasis effects in the 26.5–44.9 region mostly involved dominance,  $A \times D$ ,  $D \times A$  and  $D \times D$ , with the most significant such effect being  $D \times A$  between the dominance effect of *rs109675908* in *ATF7* and the additive effect of *rs42851612* (*SLC38A1-SCAF11*) (Figure 7d).

Chr15 had intra-chromosome epistasis effects mostly in the 52–66 Mb region (Figure 7e) with peak effect between *rs132656140* in *LOC107133183* and *rs29022897* in *ABTB2* in the 63.51–64.79 Mb region (Figure 7f).

Chr06 had intra-chromosome epistasis effects mostly in the 80–92 Mb region (Figure 7g), with peak effects between *rs136961414* in *LOC100138004* and *rs134391498* in *SLC4A4* (mboxfigfig:genes-1277371-f007h). Similar to milk and protein yields, *rs109452259* between *GC* and *NPF2R2* and five SNPs in *SLC4A4* had the #3 and #8–11 effects of the chromosome.

Chr26 had scattered intra-chromosome epistasis effects in the 12–32 Mb region (Figure 7i) with peak effect between *rs42092371* in *NEURL1* and *rs41604819* in *SORCS1* (Figure 7j).

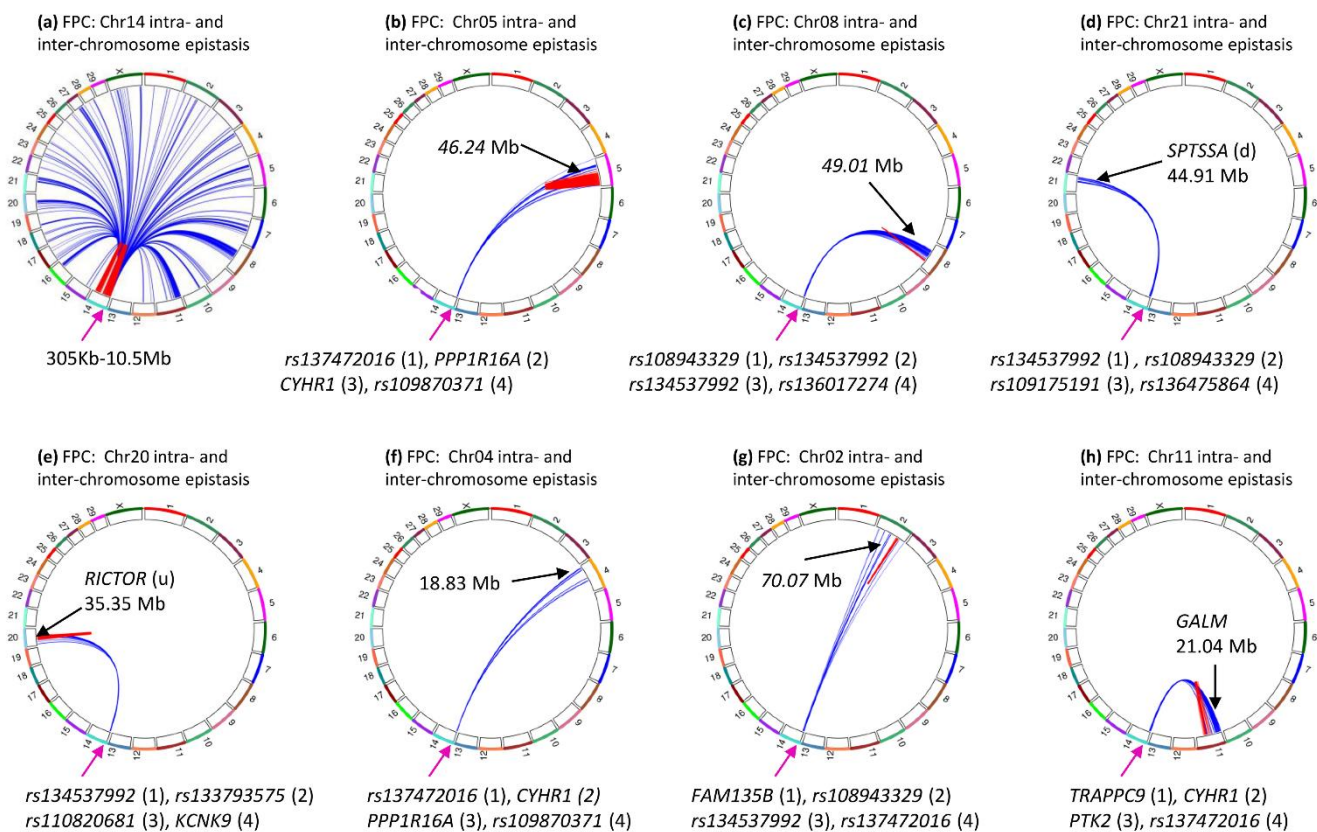


**Figure 7.** Chromosomes with most significant intra-chromosome epistasis effects for fat yield. The top row shows the circular plots with red color indicating intra-4 Mb epistasis effect and blue color indicating inter-4 Mb epistasis effects; the bottom row shows the Manhattan plots of  $\log_{10}(1/p)$  values. The black arrow indicates the location of the peak epistasis effect of the chromosome. Gene name indicates the gene with a peak SNP epistasis effect, and ‘-’ indicates the SNP was between the two genes. FY = fat yield. (a): Intra-chromosome epistasis effects of Chr14 for fat yield. (b): Statistical significance of intra-chromosome epistasis effects of Chr14 for fat yield. (c): Intra-chromosome epistasis effects of Chr05 for fat yield. (d): Statistical significance of intra-chromosome epistasis effects of Chr05 for fat yield. (e): Intra-chromosome epistasis effects of Chr15 for fat yield. (f): Statistical significance of intra-chromosome epistasis effects of Chr15 for fat yield. (g): Intra-chromosome epistasis effects of Chr06 for fat yield. (h): Statistical significance of intra-chromosome epistasis effects of Chr06 for fat yield. (i): Intra-chromosome epistasis effects of Chr26 for fat yield. (j): Statistical significance of intra-chromosome epistasis effects of Chr26 for fat yield.

### 3.7. Inter-Chromosome Epistasis Effects of Fat Percentages

The most surprising inter-chromosome epistasis effects were those for fat percentage: all 29 chromosomes interacted with a Chr14 region with *DGAT1*, which had the most significant SNP effect for fat percentage [19]. Figure 2a shows 27 of the 30 chromosomes had significant inter-chromosome epistasis effects for fat percentage ( $\log_{10}(1/p) \geq 12.82$ ) among the top 50,000 pairs. However, 26 of the 27 chromosomes each only interacted with the same 0.3–10.5 Mb region of Chr14 (Figure 8a) with A  $\times$  A effects. Only one inter-chromosome effect did not involve the Chr14 region: the D  $\times$  D effect between Chr22 and Chr23 (Figure S2, Table S2). These results were from the 80 K SNP set that had more intra-chromosome epistasis effects (98.1%, Table 1) and fewer inter-chromosome epistasis effects (1.9%, Table 1) than those from the 60 K SNP set that had 94% intra- and 6% inter-chromosome epistasis effects. From the 60 K, all 29 chromosomes interacted with the same Chr14 region (Figure S2, Table S2). Figure 8b–h shows some 80 K examples showing each chromosome specifically interacted with the 0.3–10.5 Mb region of chr14.

Such a high degree of specificity of the 0.3–10.5 Mb region of chr14 for interactions with other chromosomes provided evidence implicating a potential global regulatory role of this region for fat percentage and indicated that fat percentage likely had genetic contributions from the entire genome.



**Figure 8.** Examples of inter-chromosome epistasis effects between the same Chr14 region and different chromosomes for fat percentage ( $\log_{10}(1/p) \geq 12.82$ ). Each number on the outer circle is the chromosome number, red color represents intra-chromosome epistasis effects, and blue color represents inter-chromosome epistasis effects. Number in () is the effect ranking. The arrow indicates the Chr14 region that interacted with every chromosome for fat percentage. FPC = fat percentage. (a): Inter-chromosome epistasis effects between 26 chromosomes and the 0.3–10.5 Mb region of Chr14 for fat percentage. (b): Inter-chromosome epistasis effects between Chr05 and the 0.3–10.5 Mb region of Chr14 for fat percentage. (c): Inter-chromosome epistasis effects between Chr08 and the 0.3–10.5 Mb region of Chr14 for fat percentage. (d): Inter-chromosome epistasis effects between Chr21 and the 0.3–10.5 Mb region of Chr14 for fat percentage. (e): Inter-chromosome epistasis effects between Chr20 and the 0.3–10.5 Mb region of Chr14 for fat percentage. (f): Inter-chromosome epistasis effects between Chr04 and the 0.3–10.5 Mb region of Chr14 for fat percentage. (g): Inter-chromosome epistasis effects between Chr02 and the 0.3–10.5 Mb region of Chr14 for fat percentage. (h): Inter-chromosome epistasis effects between Chr11 and the 0.3–10.5 Mb region of Chr14 for fat percentage.

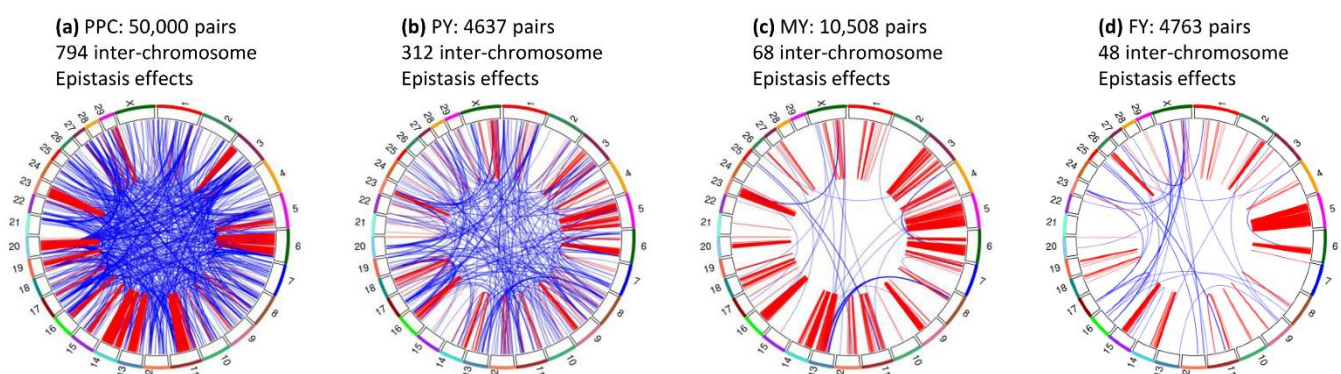
The analysis of the 0.3–10.5 Mb region of Chr14 showed that two regions flanking *DGAT1* had the most significant and frequent inter-chromosome epistasis effects, the *PPP1R16A-CYHR1* region at 0.46–0.50 Mb about 120 Kb upstream of *DGAT1* and the 2.33–3.34 Mb intergenic region about 1.71 Mb downstream of *DGAT1* including *rs134537992* at 242.1119 Kb with the most significant inter-chromosome epistasis effects (Figure 2b). Although *DGAT1* was only 0.12 Mb from the *PPP1R16A-CYHR1* region that had over 200 inter-chromosome epistasis effects, *DGAT1* was not involved in 98.75% of the 954 pairs of inter-chromosome epistasis effects, with a total of 12 (1.25%) effects and the highest effect ranking of #332 for inter-chromosome epistasis effects. In contrast, *TRAPPC9* had 73 (7.65%) of the 954 inter-chromosome epistasis effects and the #2 effect. The approximately 100 Kb intergenic region of 2.33–2.42 Mb had six SNPs and five of these six SNPs had 159

significant inter-chromosome epistasis effects with 19 chromosomes (chromosomes 1, 2, 7–9, 11, 12, 15, 17–24, 26, 28, X), including the most significant effect of *rs134537992*. The 0.3–10.5 Mb region had a gap of inter-chromosome epistasis effects at 87.65 Kb–2.33 Mb (Figure 2b), a 2.24 Mb gap without inter-chromosome epistasis effects but with many intra-chromosome epistasis effects (Figure 3a, Table S1), many SNP effects [19], and at least 44 coding genes.

Chr05 had 63 SNPs in the 0.13–119.07 Mb region interacting with the 0.46–5.68 Mb region of Chr14, showing that virtually the entire Chr05 of 120.09 Mb interacted with the same Chr14 region only, noting that nearly the entire Chr05 also had intra-chromosome epistasis effects for milk, fat and protein yields. Chr11 had the second largest number of SNPs interacting with the Chr14 regions, with 259 SNPs in the 0.65–62.52 Mb region interacting with the 0.46–6.25 Mb region of Chr14. Chr08 had the second largest number of SNPs interacting with the Chr14 regions, with 137 SNPs in the 32.72–81.71 Mb region interacting with the 0.46–6.76 Mb region of Chr14. It is interesting to note that both SNPs in *DGAT1* had significant inter-chromosome epistasis effects with *rs109477201* of Chr05, although *DGAT1* only had 1.25% of all inter-chromosome epistasis effects.

### 3.8. Inter-Chromosome Epistasis Effects of Protein Percentage and Yield Traits

The inter-chromosome epistasis effects of protein percentage and the three yield traits involved many chromosome pairs (Figure 9) but did not have the specificity observed for inter-chromosome epistasis effects of Chr14 for fat percentage (Figure 8).



**Figure 9.** Inter-chromosome epistasis effects with  $\log_{10}(1/p) \geq 12$  for protein percentage and three yield traits. Each number on the outer circle is the chromosome number, red color represents intra-chromosome epistasis effects, and blue color represents inter-chromosome epistasis effects. PPC = protein percentage. PY = protein yield. MY = milk yield. FY = fat yield. (a): Intra- and inter-chromosome epistasis effects of protein percentage. (b): Intra- and inter-chromosome epistasis effects of protein yield. (c): Intra- and inter-chromosome epistasis effects of milk yield. (d): Intra- and inter-chromosome epistasis effects of fat yield.

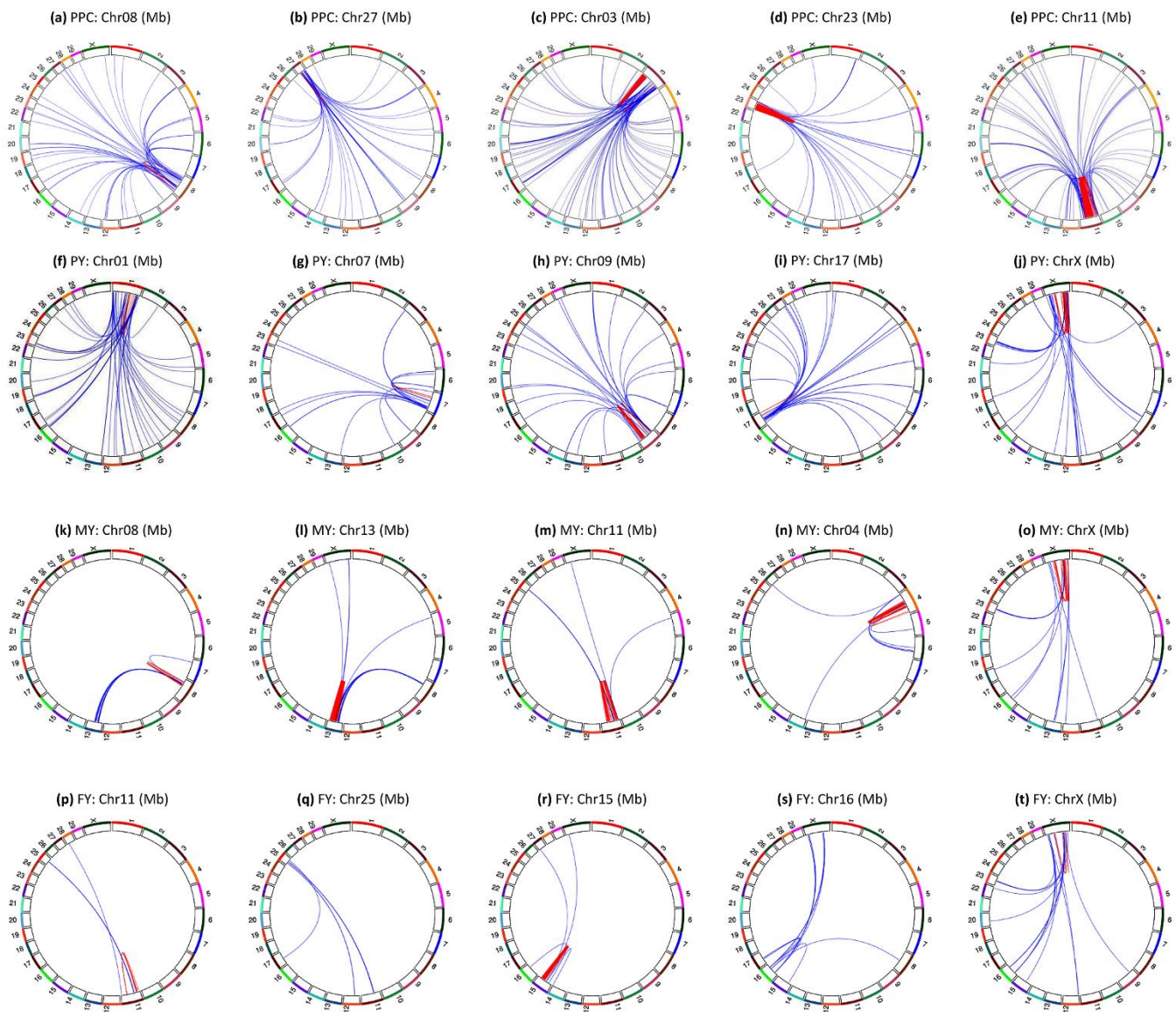
Protein percentage had 794 inter-chromosome epistasis effects, fewer than the 954 inter-chromosome epistasis effects for fat percentage (Table 1), and every chromosome interacted with most of the other chromosomes (Figure 9a and Figure S2), unlike fat percentage where every chromosome virtually interacted with the same Chr14 region only (Figure 8a and Figure S2). Figure 10a–e show examples where every chromosome that interacted with most of the other chromosomes for protein percentage.

Protein yield (Figure 9b) had 312 significant inter-chromosome epistasis effects with similar patterns as protein percentage, i.e., every chromosome had inter-chromosome epistasis effects with some other chromosomes. Figure 10f–j show examples of inter-chromosome epistasis effects of protein yield.

Milk yield had 68 (0.6%) inter-chromosome epistasis effects involving chromosome pairs between 20 chromosomes (Figure 9c). Half of these 68 effects were between Chr08 and Chr13, with 29 effects between 5 SNPs in *LOC101904827* on Chr08 and 14 SNPs in

the 20.63–46.24 Mb region of Chr13. Figure 10k–o show examples of inter-chromosome epistasis effects of milk yield.

Fat yield had 48 (1.0%) inter-chromosome epistasis effects involving chromosome pairs between 22 chromosomes (Figure 9d). A surprise was that Chr05 with intra-chromosome epistasis effects virtually covering the entire chromosome (Figure 7c) did not have significant inter-chromosome epistasis effects. Figure 10p–t show examples of inter-chromosome epistasis effects of fat yield.

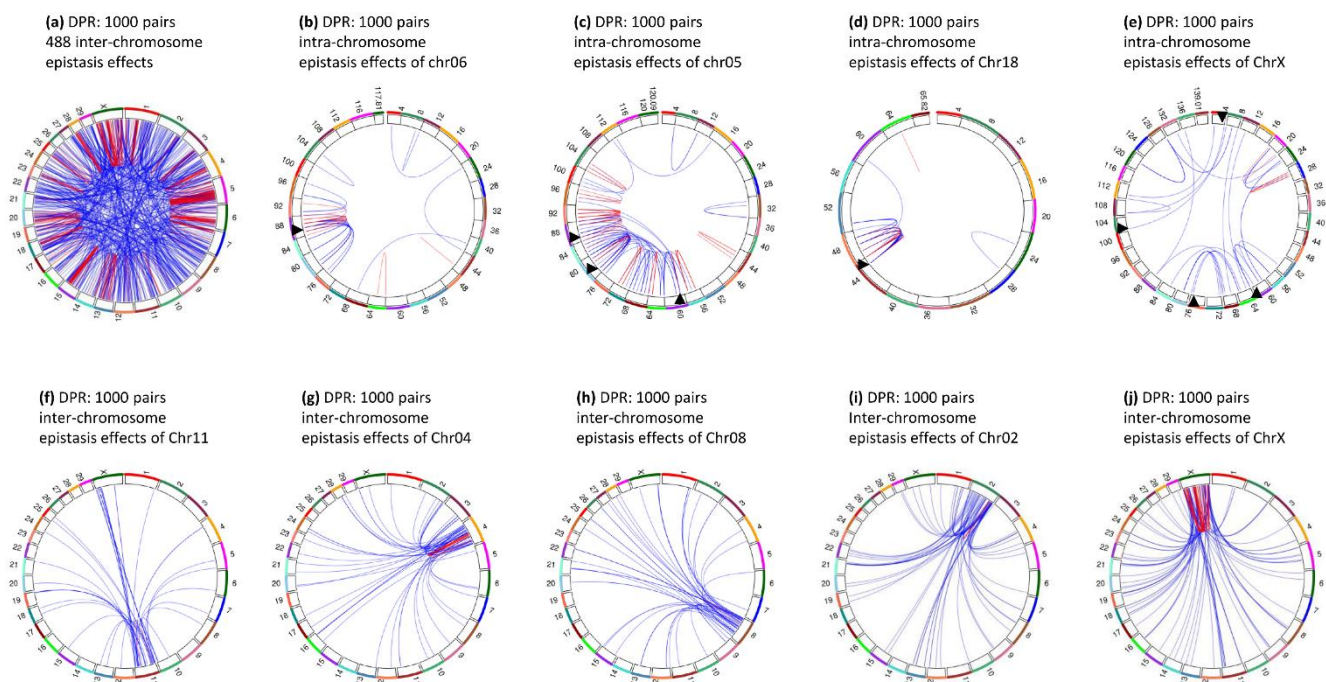


**Figure 10.** Examples of inter-chromosome epistasis effects with  $\log_{10}(1/p) \geq 12$  for protein percentage and the three yield traits. Each number on the outer circle is the chromosome number, red color represents intra-chromosome epistasis effects, and blue color represents inter-chromosome epistasis effects. PPC = protein percentage. PY = protein yield. MY = milk yield. FY = fat yield. (a–e): Examples of inter-chromosome epistasis effects for protein percentage. (f–j): Examples of inter-chromosome epistasis effects for protein yield. (k–o): Examples of inter-chromosome epistasis effects for milk yield. (p–t): Examples of inter-chromosome epistasis effects for fat yield.

### 3.9. Epistasis Effects of Daughter Pregnancy Rate

Daughter pregnancy rate only had one significant epistasis effect with  $\log_{10}(1/p) \geq 12$  (Figure 1f), and this epistasis effect was an intra-chromosome A  $\times$  A effect between

*rs133222820* in *TMPRSS11E* and *rs43470988* (*RUFY3-GRSF1*) of Chr06. Although other effects did not reach the Bonferroni significance, several epistasis effects involved genes with reported fertility effects. These included *EIF5B* (inter-chromosome epistasis effect between *rs43652424* in *EIF5B* of Chr11 and *rs136798451* flanked by *RAP2C* and *MBNL3* of ChrX) associated with female sterility [34], *RAP2C* (inter-chromosome epistasis effect) associated with gestational duration and/or spontaneous preterm birth [35], *PKP2* (intra-chromosome epistasis effect of Chr05) associated with cow conception rate [36], *HSD17B6* (intra-chromosome epistasis effect of Chr05) associated with polycystic ovary syndrome [37], and *SOX5* (intra-chromosome epistasis effect of Chr05) associated with nonobstructive azoospermia susceptibility [38]. Figure 11a shows the top 1000 epistasis effects for daughter pregnancy rate, with 55.2% intra-chromosome and 48.8 inter-chromosome epistasis effects. Figure 11b–e show examples of intra-chromosome and Figure 11f–j shows examples of inter-chromosome epistasis effects.



**Figure 11.** Intra- and inter-chromosome epistasis effects of daughter pregnancy rate (top 1000 pairs). For intra-chromosome epistasis effects, each number on the outer circle is the chromosome number, red color indicates intra-chromosome epistasis effects, and blue color indicates inter-chromosome epistasis effects. For inter-chromosome epistasis effects, each number on the outer circle is the chromosome distance, red color indicating intra-4 Mb epistasis effects and blue color indicating inter-4 Mb epistasis effects. DPR = daughter pregnancy rate. (a): Intra- and inter-chromosome epistasis effects of all chromosomes for daughter pregnancy rate. (b): Intra-chromosome epistasis effects of Chr06 for daughter pregnancy rate. (c): Intra-chromosome epistasis effects of Chr05 for daughter pregnancy rate. (d): Intra-chromosome epistasis effects of Chr18 for daughter pregnancy rate. (e): Intra-chromosome epistasis effects of ChrX for daughter pregnancy rate. (f): Inter-chromosome epistasis effects of Chr11 for daughter pregnancy rate. (g): Inter-chromosome epistasis effects of Chr04 for daughter pregnancy rate. (h): Inter-chromosome epistasis effects of Chr08 for daughter pregnancy rate. (i): Inter-chromosome epistasis effects of Chr02 for daughter pregnancy rate. (j): Inter-chromosome epistasis effects of ChrX for daughter pregnancy rate.

## 4. Discussion

### 4.1. Complex Epistasis Effects Existed in U.S. Holstein Cattle

The epistasis results of the five milk production traits showed that complex epistasis effects existed in U.S. Holstein cattle. Every production trait had intra-chromosome epistasis effects covering large chromosome distances, and inter-chromosome epistasis

effects existed. The Chr05 example for fat yield was the largest chromosome segment (nearly the entire chromosome) covered by intra-chromosome epistasis effects. In addition, inter-chromosome epistasis effects also existed involving most chromosomes for every trait. For fat percentage, a Chr14 region had inter-chromosome epistasis effects with all other chromosomes. The complexity of the epistasis effects in this study was unprecedented.

#### *4.2. Genetic Selection Based on Genome-Wide SNP Additive Effects Likely Accounted for Most Intra-Chromosome $A \times A$ Effects*

The chromosome regions with highly significant intra-chromosome  $A \times A$  epistasis effects of the five production traits overlapped the chromosome regions with highly significant SNP additive effects we previously reported from the same data set. Given that genomic prediction using single-SNP additive models for the production traits was successful, we hypothesize that genome-wide SNP additive effects likely accounted for most intra-chromosome  $A \times A$  effects. The linkage between SNPs with intra-chromosome  $A \times A$  effects should have contributed to the success of genomic prediction using SNP additive models for the production traits. As the percentage of inter-chromosome epistasis effects increases, the effectiveness of SNP prediction models is expected to decrease, as discussed next. Based on this assumption, SNP prediction models should be most effective for fat and protein percentages that had the highest frequencies of intra-chromosome  $A \times A$  effects, followed by milk yield, fat yield and protein yield. Statistical significance is another factor affecting the contribution of intra- or inter-chromosome epistasis effects to the trait. For the five production traits, intra-chromosome epistasis effects were more significant than inter-chromosome effects, e.g.,  $\log_{10}(1/p)$  was in the range of 13–537 for intra- and 13–23 for inter-chromosome epistasis effects of fat percentage (Table 1), and the other four production traits had a similar trend. Daughter pregnancy rate was different with similar statistical significance for both intra- and inter-chromosome epistasis effects,  $\log_{10}(1/p)$  was in the range of 6–13 for intra- and 6–11 for inter-chromosome epistasis effects. Combining the frequency and statistical significance of intra-chromosome epistasis effects, genomic selection using SNP additive models is expected to be highly effective for the five production traits.

#### *4.3. Inter-Chromosome Epistasis Effects Could Be a Genetic Mechanism for Lack of Selection Response and Low Heritability*

Daughter pregnancy rate had the highest frequency of inter-chromosome epistasis effects (84.2%, Table 1), and more than half (52.9%) of the top 50,000 epistasis effects were  $A \times A$  effects. Although only one epistasis effect reached the Bonferroni significance, the epistasis results provided the first indications that inter-chromosome  $A \times A$  effects may be the primary genetic effects of daughter pregnancy rate. Assuming this was true, inter-chromosome  $A \times A$  effects would be consistent with the fact that genetic selection for daughter pregnancy rate was less responsive than the yield traits because genetic selection for many independent variants would be more difficult than selection for many lined variants. As discussed later, we hypothesize that marginal effect of many epistasis effects of linked loci is a contributing factor to highly significant SNP effect. However, such marginal effect is unavailable for inter-chromosome epistasis effects of independent SNPs on different chromosomes.

#### *4.4. Chr14-Specific Inter-Chromosome $A \times A$ Epistasis Effects Increase the Statistical Confidence of the Epistasis Results*

Although the statistical significance for declaring significant epistasis effects was the stringent significant level of 5% genome-wide false positives with the Bonferroni multiple testing correction, large numbers of epistasis effects for the five production traits reached this Bonferroni significance,  $\log_{10}(1/p) \geq 12$ , e.g., all the top 50,000 epistasis effects of fat or protein percentage were significant. This raised the question whether false positive effects were substantially more than the expected 5% genome-wide false positives. While a definitive answer to this question was unavailable, the inter-chromosome  $A \times A$  epistasis



effects between the 0.3–10.5 Mb Chr14 region and all chromosomes for fat percentage should increase the statistical confidence of the epistasis results, because the high degree of specificity of all 29 chromosomes interacting with the same chromosome region could not happen by chance.

#### 4.5. An Intergenic Variant May Have an Important Role for Inter-Chromosome Epistasis Effects of Fat Percentage

The 0.9 Mb intergenic region at 2.33–2.42 Mb of Chr14 had six SNPs and five of these SNPs had significant inter-chromosome epistasis effects including the most significant effect of *rs134537992* at 242.1119 Kb for fat percentage. This 0.9 Mb intergenic region was 113.637 Kb from *LOC101901918* or 311.437 Kb from *TSNARE1* upstream, and 88.851 Kb from *SLC45A4* downstream. These three flanking genes had eight SNPs, one in *LOC101901918*, five in *TSNARE1* and two in *SLC45A4*, but none of these eight SNPs had a significant inter-chromosome effect. These results would be consistent with the hypothesis that the 0.9 Mb intergenic region at 2.33–2.42 Mb had its own biological functions affecting many chromosomes for fat percentage rather than linked effects due to linkage disequilibrium with adjacent coding genes. The 2.24 Mb gap in inter-chromosome epistasis effects at 87.65 Kb–2.30 Mb should further support this hypothesis. This gap region had many intra-chromosome epistasis effects (Figure 3a, Table S1), many SNP effects [19], and at least 44 coding genes, but did not have significant inter-chromosome epistasis effects. Given that this 2.24 Mb gap region at 87.65 Kb–2.30 Mb was so close to the *PPP1R16A-CYHR1* region at 0.46–0.50 Mb and the 2.33–3.34 Mb intergenic region with significant inter-chromosome epistasis effects, the coding genes in this 2.24 Mb gap region did not have biological functions for significant inter-chromosome epistasis effects. These results indicated that the 0.9 Mb segment of noncoding DNA at 2.33–3.34 Mb harboring *rs134537992* had potential regulatory functions affecting many chromosomes for fat percentage.

#### 4.6. Causal or Linked Epistasis Effects

The results in this study did not provide evidence about potential causal variants with epistasis effects, but had indications about potential involvement of linkage disequilibrium in clusters of epistasis effects of the production traits, and an approximate chromosome region with causal effects for inter-chromosome epistasis effects. Fat percentage about is an ideal trait to discuss the issue of causal and linked effects due to the unique statistical significance and patterns of the epistasis effects. The chromosome regions with large clusters of intra-chromosome epistasis effects mostly contained the chromosome regions with large clusters of SNP effects including the Chr14 region with intra-chromosome epistasis effects. We previously reported that the region containing *DGAT1* about 6.79 Mb in size had strong linkage disequilibrium, and the removal of the *DGAT1* effect from the phenotypic values eliminated 63% of the top 1% effects in the region, showing that the top 1% SNP effects in that 6.79 Mb region involved both causal and linked effects. Since  $A \times A$  epistasis effect of a SNP pair in this study was the contrast of the four haplotype average values and each haplotype frequency was affected by linkage disequilibrium between the two SNPs [30],  $A \times A$  intra-chromosome epistasis effects due to linkage should have existed and contributed to the large clusters of intra-chromosome epistasis effects for the five production traits. Inter-chromosome epistasis effects are expected to be unaffected by linkage disequilibrium between the two SNPs of each SNP pair but some effects still could have contributions from linkage disequilibrium between a SNP adjacent to another SNP that had a significant inter-chromosome epistasis effect. An example of this possibility is the 0.9 Mb intergenic region at 2.33–2.42 Mb of Chr14 where five of the six SNPs had significant inter-chromosome epistasis effects including the most significant effect of *rs134537992* at 242.1119 Kb for fat percentage. Some of these five significant epistasis effects could be due to linkage to a causal effect nearby. However, inter-chromosome epistasis effects clearly were less affected by linkage, e.g., the 2.24 Mb gap region at 87.65 Kb–2.30 Mb without significant inter-chromosome epistasis effects was very close to the *PPP1R16A-*

CYHR1 region at 0.46–0.50 Mb and the 2.33–3.34 Mb intergenic region with significant inter-chromosome epistasis effects.

## 5. Conclusions

The epistasis results showed that complex intra- and inter-chromosome epistasis effects existed in U.S. Holstein cattle. Intra-chromosome A × A effects were the main epistasis effects for the five production traits, whereas inter-chromosome A × A effects were the main epistasis effects for daughter pregnancy rate. Chromosome regions with highly significant intra-chromosome epistasis effects mostly overlapped the chromosome regions with highly significant SNP effects we previously reported using the same data. A chr14 region widely reported to have highly significant SNP effects for fat percentage interacted with all other chromosomes, and this result provided evidence for a potential global regulatory role of this Chr14 region.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes12071089/s1>, Figure S1: Intra-chromosome epistasis effects with  $\log(1/p) \geq 12$ , Figure S2: Inter-chromosome epistasis effects with  $\log(1/p) \geq 12$ , Table S1: Significant intra-chromosome epistasis effects, Table S2: Significant inter-chromosome epistasis effects.

**Author Contributions:** Conceptualization, Y.D.; methodology, Y.D., D.P., Z.L., L.M. and J.J.; software, D.P., Z.L., L.M.; validation, D.P. and Z.L.; formal analysis, D.P., Z.L. and Y.D.; resources, D.P., Z.L., J.J. and L.M.; data curation, D.P.; writing—original draft preparation, Y.D.; writing—review and editing, D.P., Z.L., J.J. and L.M.; visualization, Z.L., D.P. and Y.D.; supervision, Y.D.; project administration, Y.D.; funding acquisition, Y.D. and L.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by grants 2018-67015-28128, 2020-67015-31133 and 2020-67015-31398 from the USDA National Institute of Food and Agriculture, and by project MIN-16-124 of the Agricultural Experiment Station at the University of Minnesota.

**Institutional Review Board Statement:** Ethical review and approval were waived because this study used existing data only and did not involve the use of live animals.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original genotype data are owned by third parties and maintained by the Council on Dairy Cattle Breeding (CDCB). A request to CDCB is necessary for getting data access on research, which may be sent to: João Dürr, CDCB Chief Executive Officer (joao.durr@cdcb.us). All other relevant data are available in the manuscript, Supplementary Materials.

**Acknowledgments:** Members of the Council on Dairy Cattle Breeding (CDCB) and the Cooperative Dairy DNA Repository (CDDR) are acknowledged for providing the data for the GWAS analysis. The Ceres high performance computing system of USDA-ARS provided computing time and storage for the data analysis. Data access and the use of USDA-ARS computing facilities by this research were supported by USDA-ARS project 8042-31000-002-00-D, “Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals”, and USDA-ARS project 8042-31000-001-00-D, “Enhancing Genetic Merit of Ruminants Through Improved Genome Assembly, Annotation, and Selection”. The USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this manuscript is solely for the purpose of providing specific information and does not imply recommendation or endorsement by USDA. The authors thank Paul VanRaden for help with data preparation, and John Cole, Steven Schroeder and Ransom Baldwin for help with using the USDA-ARS computing facilities.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

1. Mackay, T.F. Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nat. Rev. Genet.* **2014**, *15*, 22–33. [[CrossRef](#)]
2. Carlborg, Ö.; Haley, C.S. Epistasis: Too often neglected in complex trait studies? *Nat. Rev. Genet.* **2004**, *5*, 618–625. [[CrossRef](#)]

3. Cordell, H.J. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **2002**, *11*, 2463–2468. [[CrossRef](#)]
4. Phillips, P.C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **2008**, *9*, 855–867. [[CrossRef](#)] [[PubMed](#)]
5. Ritchie, M.D.; Van Steen, K. The search for gene-gene interactions in genome-wide association studies: Challenges in abundance of methods, practical considerations, and biological interpretation. *Ann. Transl. Med.* **2018**, *6*, 157. [[CrossRef](#)] [[PubMed](#)]
6. Pryce, J.; Bolormaa, S.; Chamberlain, A.; Bowman, P.; Savin, K.; Goddard, M.; Hayes, B. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J. Dairy Sci.* **2010**, *93*, 3331–3345. [[CrossRef](#)]
7. Cole, J.B.; Wiggans, G.R.; Ma, L.; Sonstegard, T.S.; Lawlor, T.J.; Crooker, B.A.; Van Tassell, C.P.; Yang, J.; Wang, S.; Matukumalli, L.K.; et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genom.* **2011**, *12*, 408. [[CrossRef](#)]
8. Jiang, J.; Shen, B.; O'Connell, J.R.; VanRaden, P.M.; Cole, J.B.; Ma, L. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. *BMC Genom.* **2017**, *18*, 425. [[CrossRef](#)] [[PubMed](#)]
9. Raven, L.-A.; Cocks, B.G.; Hayes, B.J. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genom.* **2014**, *15*, 62. [[CrossRef](#)] [[PubMed](#)]
10. Ma, L.; Wiggans, G.; Wang, S.; Sonstegard, T.; Yang, J.; Crooker, B.; Cole, J.; Van Tassell, C.; Lawlor, T.; Da, Y. Effect of sample stratification on dairy GWAS results. *BMC Genom.* **2012**, *13*, 536. [[CrossRef](#)] [[PubMed](#)]
11. Bouwman, A.C.; Daetwyler, H.D.; Chamberlain, A.J.; Ponce, C.H.; Sargolzaei, M.; Schenkel, F.S.; Sahana, G.; Govignon-Gion, A.; Boitard, S.; Dolezal, M. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* **2018**, *50*, 362–367. [[CrossRef](#)] [[PubMed](#)]
12. Rothhammer, S.; Seichter, D.; Förster, M.; Medugorac, I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genom.* **2013**, *14*, 908. [[CrossRef](#)] [[PubMed](#)]
13. Guo, J.; Jorjani, H.; Carlborg, Ö. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BMC Genom.* **2012**, *13*, 82. [[CrossRef](#)]
14. Bolormaa, S.; Pryce, J.; Hayes, B.; Goddard, M. Multivariate analysis of a genome-wide association study in dairy cattle. *J. Dairy Sci.* **2010**, *93*, 3818–3833. [[CrossRef](#)]
15. Pryce, J.E.; Haile-Mariam, M.; Goddard, M.E.; Hayes, B.J. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet. Sel. Evol.* **2014**, *46*, 71. [[CrossRef](#)]
16. Sanchez, M.-P.; Govignon-Gion, A.; Croiseau, P.; Fritz, S.; Hozé, C.; Miranda, G.; Martin, P.; Barbat-Leterrier, A.; Letaïef, R.; Rocha, D. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. *Genet. Sel. Evol.* **2017**, *49*, 68. [[CrossRef](#)] [[PubMed](#)]
17. Weller, J.; Bickhart, D.; Wiggans, G.; Tooker, M.; O'Connell, J.; Jiang, J.; Ron, M.; VanRaden, P. Determination of quantitative trait nucleotides by concordance analysis between quantitative trait loci and marker genotypes of US Holsteins. *J. Dairy Sci.* **2018**, *101*, 9089–9107. [[CrossRef](#)] [[PubMed](#)]
18. Littlejohn, M.D.; Tiplady, K.; Fink, T.A.; Lehnert, K.; Lopdell, T.; Johnson, T.; Couldrey, C.; Keehan, M.; Sherlock, R.G.; Harland, C. Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. *Sci. Rep.* **2016**, *6*, 25376. [[CrossRef](#)]
19. Jiang, J.; Ma, L.; Prakapenka, D.; VanRaden, P.M.; Cole, J.B.; Da, Y. A large-scale genome-wide association study in US Holstein cattle. *Front. Genet.* **2019**, *10*, 412. [[CrossRef](#)]
20. Jiang, J.; Cole, J.B.; Freebern, E.; Da, Y.; VanRaden, P.M.; Ma, L. Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. *Commun. Biol.* **2019**, *2*, 212. [[CrossRef](#)]
21. Barendse, W.; Harrison, B.E.; Hawken, R.J.; Ferguson, D.M.; Thompson, J.M.; Thomas, M.B.; Bunch, R.J. Epistasis between calpain 1 and its inhibitor calpastatin within breeds of cattle. *Genetics* **2007**, *176*, 2601–2610. [[CrossRef](#)] [[PubMed](#)]
22. Knaust, J.; Hadlich, F.; Weikard, R.; Kuehn, C. Epistatic interactions between at least three loci determine the “rat-tail” phenotype in cattle. *Genet. Sel. Evol.* **2016**, *48*, 1–12. [[CrossRef](#)] [[PubMed](#)]
23. Suchocki, T.; Komisarek, J.; Szyda, J. Testing candidate gene effects on milk production traits in dairy cattle under various parameterizations and modes of inheritance. *J. Dairy Sci.* **2010**, *93*, 2703–2717. [[CrossRef](#)]
24. Yao, C.; Spurlock, D.; Armentano, L.; Page Jr, C.; VandeHaar, M.; Bickhart, D.; Weigel, K. Random Forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. *J. Dairy Sci.* **2013**, *96*, 6716–6729. [[CrossRef](#)] [[PubMed](#)]
25. Wiggans, G.R.; Cole, J.B.; Hubbard, S.M.; Sonstegard, T.S. Genomic selection in dairy cattle: The USDA experience. *Annu. Rev. Anim. Biosci.* **2017**, *5*, 309–327. [[CrossRef](#)] [[PubMed](#)]
26. VanRaden, P. Practical implications for genetic modeling in the genomics era. *J. Dairy Sci.* **2016**, *99*, 2405–2412. [[CrossRef](#)]
27. Ma, L.; Jiang, J.; Prakapenka, D.; Cole, J.; Da, Y. Approximate generalized least squares method for large-scale genome-wide association study. *J. Dairy Sci.* **2019**, *102*, 30.
28. Ma, L.; Runesha, H.B.; Dvorkin, D.; Garbe, J.; Da, Y. Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies. *BMC Bioinform.* **2008**, *9*, 315. [[CrossRef](#)]

29. Weeks, N.T.; Luecke, G.R.; Groth, B.M.; Kraeva, M.; Ma, L.; Kramer, L.M.; Koltjes, J.E.; Reecy, J.M. High-performance epistasis detection in quantitative trait GWAS. *Int. J. High Perform. Comput. Appl.* **2016**, *32*, 321–336. [[CrossRef](#)]
30. Mao, Y.; London, N.R.; Ma, L.; Dvorkin, D.; Da, Y. Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. *Physiol. Genom.* **2007**, *28*, 46–52. [[CrossRef](#)]
31. Wang, S.; Dvorkin, D.; Da, Y. SNPEVG: A graphical tool for GWAS graphing with mouse clicks. *BMC Bioinform.* **2012**, *13*, 319. [[CrossRef](#)] [[PubMed](#)]
32. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, H.; Meltzer, P.; Davis, S. RCircos: An R package for Circos 2D track plots. *BMC Bioinform.* **2013**, *14*, 244. [[CrossRef](#)] [[PubMed](#)]
34. Johnstone, O.; Lasko, P. Interaction with eIF5B is essential for Vasa function during development. *Development* **2004**, *131*, 4167–4178. [[CrossRef](#)] [[PubMed](#)]
35. Zhou, G.; Holzman, C.; Heng, Y.J.; Kibschull, M.; Lye, S.J.; Vazquez, A. EBF1 gene mRNA levels in maternal blood and spontaneous preterm birth. *Reprod. Sci.* **2020**, *27*, 316–324. [[CrossRef](#)] [[PubMed](#)]
36. Sugimoto, M.; Sasaki, S.; Gotoh, Y.; Nakamura, Y.; Aoyagi, Y.; Kawahara, T.; Sugimoto, Y. Genetic variants related to gap junctions and hormone secretion influence conception rates in cows. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19495–19500. [[CrossRef](#)]
37. Cochran, S.D.; Cole, J.B.; Null, D.J.; Hansen, P.J. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genom.* **2013**, *14*, 49. [[CrossRef](#)]
38. Gu, X.; Li, H.; Chen, X.; Zhang, X.; Mei, F.; Jia, M.; Xiong, C. PEX10, SIRPA-SIRPG, and SOX5 gene polymorphisms are strongly associated with nonobstructive azoospermia susceptibility. *J. Assist. Reprod. Genet.* **2019**, *36*, 759–768. [[CrossRef](#)]