



OPEN

## Genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle

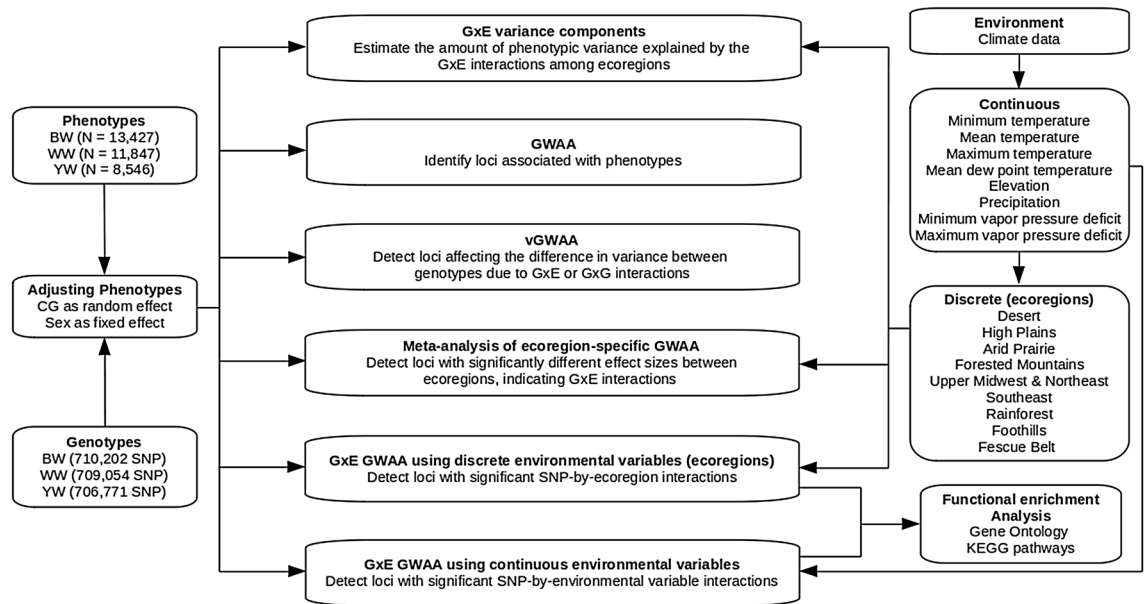
Camila U. Braz<sup>1</sup>, Troy N. Rowan<sup>1,2</sup>, Robert D. Schnabel<sup>1,2,3</sup> & Jared E. Decker<sup>1,2,3</sup>✉

Understanding genotype-by-environment interactions (G × E) is crucial to understand environmental adaptation in mammals and improve the sustainability of agricultural production. Here, we present an extensive study investigating the interaction of genome-wide SNP markers with a vast assortment of environmental variables and searching for SNPs controlling phenotypic variance (vQTL) using a large beef cattle dataset. We showed that G × E contribute 10.1%, 3.8%, and 2.8% of the phenotypic variance of birth weight, weaning weight, and yearling weight, respectively. G × E genome-wide association analysis (GWAA) detected a large number of G × E loci affecting growth traits, which the traditional GWAA did not detect, showing that functional loci may have non-additive genetic effects regardless of differences in genotypic means. Further, variance-heterogeneity GWAA detected loci enriched with G × E effects without requiring prior knowledge of the interacting environmental factors. Functional annotation and pathway analysis of G × E genes revealed biological mechanisms by which cattle respond to changes in their environment, such as neurotransmitter activity, hypoxia-induced processes, keratinization, hormone, thermogenic and immune pathways. We unraveled the relevance and complexity of the genetic basis of G × E underlying growth traits, providing new insights into how different environmental conditions interact with specific genes influencing adaptation and productivity in beef cattle and potentially across mammals.

Genotype-by-environment interactions (G × E) refer to a variable response of genotypes across environments, resulting in different trait values<sup>1,2</sup>. G × E is a central issue in genetics, evolution, and ecology<sup>3</sup>, and cattle represent a unique opportunity to address these questions due to a large number of genotyped animals<sup>4</sup> reared over a wide range of climatic and topographic regions. G × E may contribute to poor adaptation, which negatively affects the profitability and sustainability of the beef cattle production systems<sup>1,5</sup>. Studies using the relationship between breeding values and the environment have reported the existence of G × E for body weight in beef cattle<sup>6,7</sup>; however, limited research has focused on understanding the genetic loci of G × E underlying such traits. This knowledge could be used to better drive selection and mating decisions<sup>8</sup> as well as in statistical models incorporating environmental information to accomplish more accurate genomic predictions<sup>9</sup>. In addition, G × E information could also be useful to predict the vulnerability of populations to climate change<sup>3</sup> and to understand environmental adaptation in other mammalian species.

Here, we present a comprehensive study of G × E for birth weight (BW), weaning weight (WW), and yearling weight (YW) using SNP marker information and over 13,000 Simmental cattle. First, we estimated the contribution of G × E variance on the total phenotypic variances using a restricted maximum likelihood (REML) approach. Then, we searched for SNPs enriched for G × E using four linear mixed model approaches: (i) direct G × E genome-wide association analyses (G × E GWAA) using continuous environmental variables; (ii) direct G × E GWAA using discrete environmental variables (United States ecoregions); (iii) variance-heterogeneity genome-wide association analyses (vGWAA), which enable identification of interactions, using phenotypes adjusted for additive relationship matrix or additive, dominance and epistatic relationship matrices; and (iv) meta-analysis of ecoregion-specific GWAA. Lastly, we performed enrichment analyses using G × E candidate genes to investigate the biological mechanisms by which environmental factors modulate such phenotypes. A flowchart overview of this study is shown in Fig. 1. To our knowledge, this is the most comprehensive study

<sup>1</sup>Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA. <sup>2</sup>Genetics Area Program, University of Missouri, Columbia, MO 65211, USA. <sup>3</sup>Informatics Institute, University of Missouri, Columbia, MO 65211, USA. ✉email: DeckerJE@missouri.edu



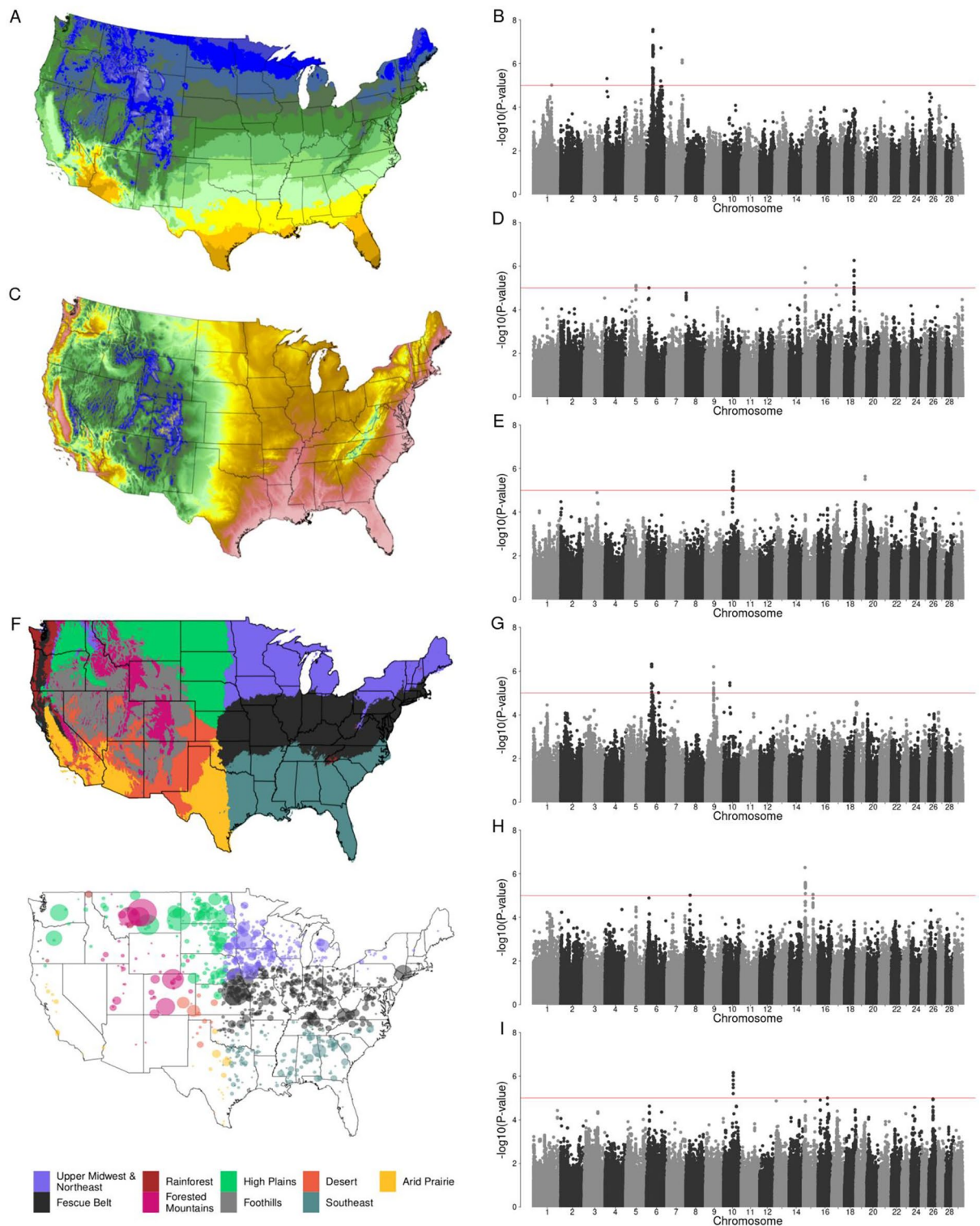
**Figure 1.** A flowchart overview of the entire study. Each topic is discussed in detail in the corresponding sections. *N* sample size, *CG* contemporary group, *BW* birth weight, *WW* weaning weight, *YW* yearling weight, *GWAA* genome-wide association analysis, *G* × *E* genotype-by-environment interactions, *vGWAA* variance-heterogeneity GWAA.

investigating the interactions of genome-wide SNP markers with a vast assortment of environmental variables and searching for SNPs controlling phenotypic variance using a large beef cattle dataset. Our results revealed novel trait associations and alternative biological mechanisms involved in shaping the total phenotypic variance of growth traits providing new insights into how the environment influences growth traits and adaptation in beef cattle and potentially across mammals.

## Results

**Environmental variance component estimates.** Phenotypes were pre-adjusted for fixed effect of sex and random effect of contemporary groups (combination of farm, season and year of birth), which represent differences in management, to remove any confounding between environment and management effects. The amount of *BW*, *WW*, and *YW* variation explained by contemporary group effects was 19%, 53%, and 61%, respectively, showing that the effect of farm management on the animal body weight increases with age. After adjusting the phenotypes, we estimated the contribution of *G* × *E* variance on the total phenotypic variances using additive relationship matrix calculated based on SNP marker information, and United States ecoregions as environmental factor. Ecoregions were determined based on the combination of 30-year normal values of mean temperature, precipitation, and elevation<sup>10</sup>. Each 1 km square of the United States was then assigned to one of 9 resulting ecoregions. We found that *G* × *E* variances contribute in 10.1%, 3.8%, and 2.8% to the *BW*, *WW*, and *YW* total phenotypic variances, respectively, suggesting that there is a decrease in the effect of *G* × *E* on the animal body weight throughout life. These results indicate an inverse relationship between farm management and environmental stress and that management practices (e.g. season of birth, forage quality and quantity, feed supplementation, health programs) are more relevant for later weights than climate environmental factors. One possible explanation for these findings is that birth weight reflects intrauterine growth which is highly affected by environmental conditions<sup>11</sup>; however, as the animal ages physiological, morphological, endocrine, cellular and molecular mechanisms are altered in order to cope up with environmental challenges<sup>12</sup>. These results also suggest that as management inputs are decreased, animals become more sensitive to environmental stress, which has ramifications for animal performance as we seek to decrease inputs to make beef production more sustainable.

**Genotype-by-environment GWAA.** The *G* × *E* GWAA for *BW*, *WW*, and *YW* were performed by fitting univariate and multivariate linear mixed models considering continuous or discrete environmental variables. The continuous environmental variables used were minimum, mean, maximum, and mean dew point temperatures; elevation; precipitation; and minimum and maximum vapor pressure deficit, separately (Additional file 1: Figures S1–S8). Discrete environmental variables included United States ecoregions (Desert and Arid Prairie, Southeast, High Plains, Forested Mountains, Fescue Belt, and Upper Midwest and Northeast—Fig. 2f), which are combinations of mean temperature, precipitation, and elevation 30-year normal values (Additional file 1: Figures S9–S14). Descriptive information for growth traits and environmental variables are included in Additional file 2: Table S1.



**Figure 2.** Genotype-by-environment interaction genome-wide association analyses ( $G \times E$  GWAA) for growth traits in Simmental cattle. (a) Mean temperature average annual over the most recent three full decades covering the conterminous United States. (b) Manhattan plot of  $G \times E$  GWAA of mean temperature for birth weight. (c) Elevation of the conterminous United States. (d) Manhattan plot of  $G \times E$  GWAA of elevation for weaning weight. (e) Manhattan plot of  $G \times E$  GWAA of elevation for yearling weight. (f) Boundaries for ecoregion assignments in the United States; (top panel) United States partitioned into nine ecoregions based on similar topographic and environmental conditions; (bottom panel) location of beef farms for which data was retrieved. (g) Manhattan plot of  $G \times E$  GWAA of Forested Mountains ecoregion for birth weight. (h) Manhattan plot of  $G \times E$  GWAA of Forested Mountains ecoregion for weaning weight. (i) Manhattan plot of  $G \times E$  GWAA of Forested Mountains ecoregion for yearling weight. In Manhattan plots, horizontal red line indicates a significant threshold ( $P < 1e-5$ ). Environmental continuous variables were drawn from the PRISM climate dataset (<http://prism.oregonstate.edu>). The United States was partitioned into nine regions using k-means clustering. Maps were plotted using the maps R package (version 3.1, <https://cran.r-project.org/web/packages/maps/>), using public domain data from the US Department of Commerce, Census Bureau.

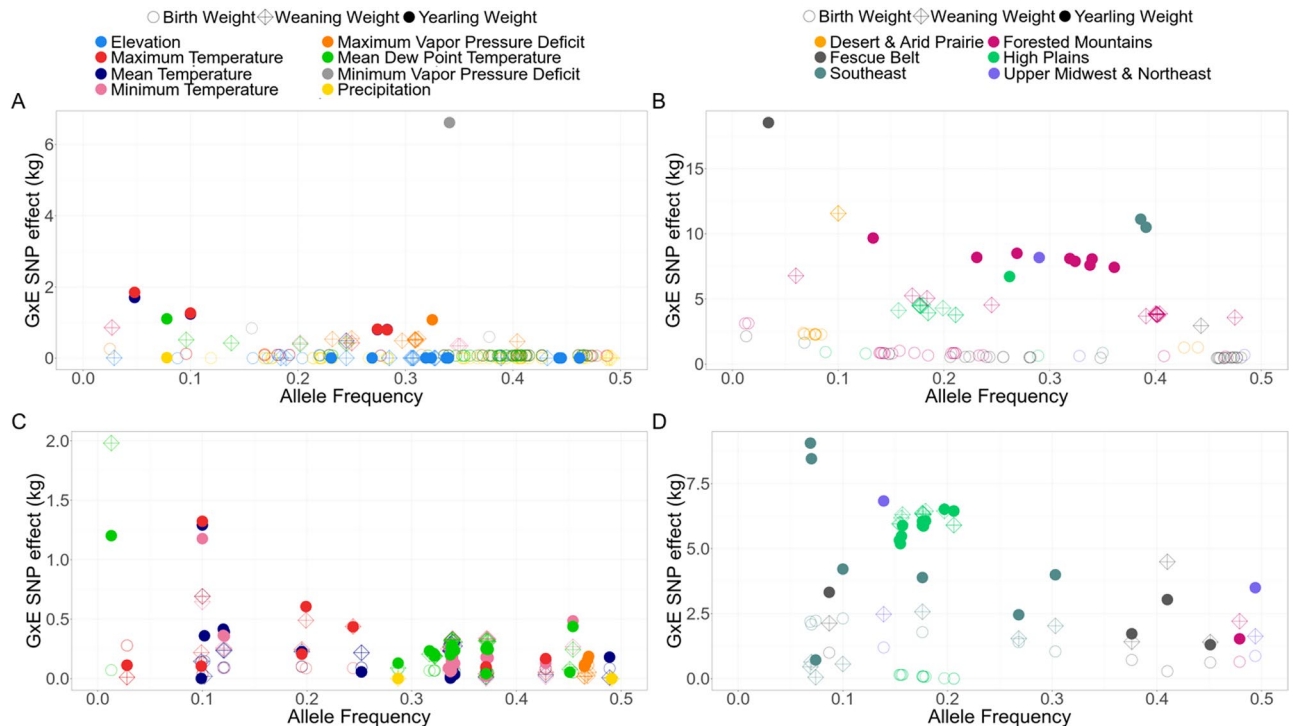
Env	Birth weight			Weaning weight			Yearling weight			Multivariate		
	N <sub>B</sub>	N <sub>A</sub>	QTL <sub>A</sub>	N <sub>B</sub>	N <sub>A</sub>	QTL <sub>A</sub>	N <sub>B</sub>	N <sub>A</sub>	QTL <sub>A</sub>	N <sub>B</sub>	N <sub>A</sub>	QTL <sub>A</sub>
Elev	142	5	5	28	15	7	15	10	3	50	–	–
Precip	338	77	11	4	4	1	1	1	1	12	2	2
MTemp	555	55	18	1	1	1	4	4	4	69	24	10
MinTemp	756	55	14	6	6	4	–	–	–	88	28	6
MaxTemp	408	49	18	1	1	1	4	4	4	23	9	8
MDpTemp	941	63	15	4	4	3	1	1	1	94	25	7
MinVPD	152	2	2	2	–	–	1	1	1	10	–	–
MaxVPD	224	5	4	10	10	4	1	1	1	17	6	1
SE	54	2	2	–	–	–	2	2	1	7	7	5
HP	40	3	3	16	12	2	4	1	1	23	13	1
FM	90	21	9	31	17	5	49	8	1	148	1	1
FB	47	32	9	1	1	1	1	1	1	11	4	4
UN	24	2	2	–	–	–	1	1	1	10	2	2
DA	125	13	3	8	3	1	6	–	–	*	*	*
Total	1923	384	115	101	74	30	78	35	20	353	121	47

**Table 1.** Results of the G × E GWAA before and after adjustment for genomic control using continuous environmental variables or United States ecoregions as environmental factors in univariate and multivariate analysis for birth weight, weaning weight, and yearling weight. *Env* environmental variables, *N<sub>B</sub>* number of significant SNPs prior to genomic control adjustment, *N<sub>A</sub>* number of significant SNPs after genomic control adjustment, *QTL<sub>A</sub>* number of significant G × E QTL based on haplotype blocks of the significant G × E SNP after adjustment for genomic control, *Elev* elevation, *Precip* precipitation, *MTemp* mean temperature, *MinTemp* minimum temperature, *MaxTemp* maximum temperature, *MDpTemp* mean dew point temperature, *MinVPD* minimum vapor pressure deficit, *MaxVPD* maximum vapor pressure deficit, *SE* Southeast, *HP* High Plains, *FM* Forested Mountains, *FB* Fescue Belt, *UN* Upper Midwest and Northeast, *DA* Desert and Arid Prairie. The complete description of the G × E SNPs can be found in the Additional files 3 and 4. \*Multivariate G × E GWAA was not performed for DA due to small sample size.

Univariate and multivariate models using both continuous and ecoregions environmental variables detected a total of 2319 G × E SNPs ( $P < 1e-5$ , Table 1, Additional files 3 and 4). However, G × E GWAA can be affected by population environmental substructure (in addition to population genetic substructure), which may cause spurious signals<sup>13</sup>, therefore for such analyses the *P*-values were adjusted using the genomic control inflation factor (Additional file 2, Table S2). After this adjustment, 614 G × E SNPs remained significant ( $P < 1e-5$ , Table 1, Additional files 3 and 4), not all of which are independent. Figure 3 shows the absolute values of the significant G × E SNP effects and their allele frequency. The majority of the G × E SNP had small effects on the traits; however, large G × E SNP effects were also detected, such as the *rs109164178* (chromosome 2: 81,187,988 bp) with effect of 6.6 kg on YW when interacting with minimum vapor pressure deficit (Fig. 3a); the *rs110995777* (chromosome 13: 16,232,193 bp) with effect of 18.5 kg on YW when comparing animals reared in the Fescue Belt ecoregion and the total dataset (Fig. 3b), among others. The multivariate model also detected G × E SNP with large effects, such as *rs42917868* (chromosome 14: 52,925,690 bp) when interacting with mean dew point temperature has an effect of 0.07 kg, 2 kg, and 1.2 kg on BW, WW, and YW, respectively (Fig. 3c); and *rs385339402* (chromosome 16: 49,283,484 bp) with G × E effect of 2.2 kg, 0.46 kg, and 9.1 kg on BW, WW, and YW, respectively, when comparing animals from the Southeast ecoregion and the total dataset (Fig. 3d).

G × E SNPs were clustered into haplotype blocks based on their linkage disequilibrium, which are referred to as G × E QTL hereafter (Table 1). In total, 212 independent G × E QTL were detected. Few G × E QTL detected by the ecoregion G × E GWAA were also identified by the continuous G × E GWAA. These results indicate that the use of ecoregion information as environmental variable allowed us to capture G × E QTL interacting with region-specific stressors that were not directly measured (e.g. forage differences, water availability, pathogens) and the combination of multiple stressors. There was no overlap between the G × E QTL identified across the traits using univariate models, however, multivariate models detected several G × E QTL associated with BW, WW, or YW ( $P < 1e-5$ , Table 1). In addition, multivariate analyses were able to detect G × E QTL that were not identifiable by standard univariate analyses, indicating advantages of incorporating a multivariate statistical approach to G × E genome-wide association studies of correlated traits.

Consistent with our G × E variance component results, a larger number of G × E QTL was detected for BW than for WW or YW (Table 1). Temperature showed more interactions with QTL affecting BW than the other environmental variables analyzed ( $P < 1e-5$ , Fig. 2a,b). The ecoregions that most interacted with QTL for BW were Fescue Belt and Forested Mountains ( $P < 1e-5$ , Fig. 2g). For WW and YW, elevation (Fig. 2c) was found to be the continuous environmental variable that most interacted with genotype ( $P < 1e-5$ , Fig. 2d,e respectively). In agreement, Forested Mountains, the ecoregion with the highest elevation (Additional file 2: Table S1), showed more genotype-by-environment interaction for WW and YW ( $P < 1e-5$ , Fig. 2h,i, respectively), compared with other ecoregions. No common G × E QTL was identified for two or more different ecoregions indicating that the



**Figure 3.** The absolute values of the significant  $G \times E$  SNP effects and their allele frequency using (a) univariate models with continuous environmental variables, (b) univariate models with ecoregion, (c) multivariate models with continuous environmental variables, and (d) multivariate models with ecoregion. Figure made with R version 3.6.3. (<https://www.r-project.org/>).

loci interacting with abiotic and biotic stressors are different across divergent environments, and consequently, the ecoregions have specific genetic architecture involved in  $G \times E$ .

We explored genomic regions that flanked all significant  $G \times E$  QTL ( $P < 1e-5$ ) to scan for genes in their vicinity and to identify possible regulatory elements (Additional file 5). Using these  $G \times E$  candidate gene sets, we performed functional enrichment analysis ( $FDR \leq 10\%$ ) to get insights into biological mechanisms by which beef cattle respond to changes in their environment (Additional file 2, Table S3). Enriched GO terms revealed molecular functions of  $G \times E$  genes that are involved in neurotransmission processes such as transmitter-gated channel activity, ligand-gated channel activity, ionotropic glutamate receptor activity, GABA-A receptor activity, neurotransmitter receptor activity involved in regulation of postsynaptic membrane potential, neurotransmitter receptor activity, among others. Function analysis also unveiled KEGG pathways involved in endocrine system (estrogen signaling pathway, and steroid hormone biosynthesis), immune response (*Staphylococcus aureus* infection, and viral myocarditis), metabolic processes (metabolism of xenobiotics by cytochrome P450, ascorbate and aldarate metabolism, pentose and glucuronate interconversions, and retinol metabolism), signal transduction (phospholipase D signaling pathway, and VEGF signaling pathway), cellular processes (endocytosis), and nervous system activity (glutamatergic synapse). Reactome pathways such as formation of the cornified envelope and keratinization were also enriched. In addition, many  $G \times E$  genes participate in nitrogen compound metabolic process, response to stimulus, protein metabolic process, developmental process, anatomical structure development, gene expression, nervous system development, immune system, among other biological processes (Additional file 2, Table S4).

The majority of  $G \times E$  SNP associations were identified for BW in this study. Studies have shown that loci identified in GWAA of BW could either represent direct effects of the fetal genotype, indirect effects of maternal genotype (acting via the intrauterine environment), or some combination of the two<sup>14,15</sup>. According to Sharma et al.<sup>16</sup> the uterine environment may even have a greater influence on fetal growth and BW than the parental genome. Thus, we investigated whether  $G \times E$  SNPs affecting BW could be reflecting maternal uterine environmental conditions. To test this, we adjusted BW by including random maternal effects in the model (MAM) and compared the results with those from the  $G \times E$  GWAA without maternal effects (MA). Results showed that 876 SNPs were not associated with  $G \times E$  effects when accounting for the effect of dam. After adjustment for genomic control, 94  $G \times E$  SNP were not detected when accounting for the effect of dam (Additional file 6), suggesting that those SNPs are interacting with the uterine environment. Several maternally-influenced  $G \times E$  genes identified in this study have been previously reported as maternal genes influencing embryonic development and/or response to environmental stress in cows and also in other mammals<sup>17–26</sup>. These results demonstrate that the bovine is an important model to understand the interaction between the physical environment, uterus, and fetus, and could be used in future research.

**Variance-heterogeneity GWAA.** The vGWAA identify loci associated with differences in variance, rather than differences in mean, between genotype classes, called vQTL, which can be explained by  $G \times E$  or  $G \times G$  (epistatic) interactions<sup>27</sup>. Therefore, we performed a REML analysis to estimate how much of the BW, WW, and YW variance is explained by additive, dominance and epistasis effects. We fitted models (M) using only additive relationship matrix (MA) or using additive, dominance, and epistasis relationship matrices (MADE), see Additional file 2: Table S5. Overall, the heritabilities were greater than the ratio of the variance explained by the non-additive genetic effects to the phenotypic variances. Proportions of dominance variances in WW and YW were low (2% and 5%, respectively), whereas no dominance variance was estimated for BW. Epistatic relationship matrix accounted for 25%, 12% and 11% of the phenotypic variances for BW, WW, and YW, respectively, suggesting that epistasis has a substantial contribution to phenotypic variation for body weight traits in Simmental cattle. Thus, to search for vQTL we used squared normalized residuals ( $z^2$ ), which are a measure of phenotypic variation<sup>28</sup>, as dependent variables fitting univariate and multivariate models. Using  $z^2$  from the MA, we expected to detect loci with both  $G \times E$  and  $G \times G$  effects, since this phenotype was adjusted only for the additive relationship matrix. However, using  $z^2$  from the MADE, we should detect exclusively  $G \times E$  loci as the phenotype was adjusted for additive, dominance and epistatic relationship matrices. In total, 44 SNPs with effect on  $z^2$  ( $P < 1e-5$ ) were identified on chromosomes 3, 5, 8, 10, 11, 13, 14, 16, 17, 18, 21, 27, and 29 (Additional file 1: Figures S15 and S16, Additional file 2: Table S5). These 44 SNPs tag 24 putative vQTL (grouped by LD patterns), with seven, four, and eleven vQTL for univariate BW, WW, and YW, respectively. The multivariate analysis identified five vQTL, of which two vQTL (*rs109142386*, chromosome 8:4,822,842 bp; and *rs135379559*, chromosome 29:39,441,489 bp) were not identified by the univariate analyses.

Nine vQTL were only significant using  $z^2$  from MA (three vQTL for univariate BW, three vQTL for univariate WW, two vQTL for univariate YW, and one vQTL using multivariate analysis). Among these vQTL likely involved in  $G \times G$ , one vQTL for BW is located on chromosome 14 at 23 Mb and resides near or within the genes *LYN*, *RPS20*, *MOS*, *PLAG1*, and *CHCHD7*. This region has been reported to have epistatic effects on several chromosomes affecting post-weaning weight in beef cattle<sup>29</sup>. These effects could be due to the gene *PLAG1* which encodes a transcription factor that regulates the expression of several genes involved in a variety of cellular processes<sup>30</sup>, and has been previously associated with stature and growth traits in many beef cattle breeds<sup>31–33</sup>. In addition, we detected a  $G \times G$  vQTL on chromosome 3 at 106 Mb within *TRIT1*, a gene that was also reported to harbor a vQTL involved in epistatic interactions in humans<sup>34</sup>.

Using  $z^2$  from MADE, 15 vQTL were detected (four vQTL for univariate BW; one vQTL for univariate WW; nine vQTL for univariate YW; and four vQTL using multivariate analyses, of which three vQTL were also identified by the univariate analyses), which likely identify  $G \times E$ . Based on GO information, genes near or within these vQTL are involved in response to stimulus (*NOS1*, *FBLN1*, *YWHAH*, and *DEPDC5*), inflammation and immunity mechanisms (*LYZL1*, *KSR2*, *FBXO21*, *IL18RAP*, *XCL2*, *XCL1*, *TTYH1*, *LENG8*, *LENG9*, *CDC42EP5*, *LAI1*, *OSCAR*, and *TARMI*), neurogenesis (*ZHX2*, and *NOS1*), regulation of transcription (*ZHX2*, *YWHAH*, and *CNOT3*), developmental process (*FBXW8*, and *CNOT3*), transmembrane transport (*SLC9A2*, *SLC9A4*, *SLC5A1*, *SLC5A4*, and *TTYH1*), neurotransmission regulation (*SHANK2*), and cell proliferation (*DPT*, *RPS9*, and *TACCI*). Some of those genes were previously reported to be involved in environmental adaptation in cattle (*XCL2*<sup>35</sup>, *SLC9A4*<sup>36</sup>, *ZMAT4*<sup>37</sup>), and heat tolerance in buffalo (*XCL1*<sup>38</sup>) and dairy cattle (*SHANK2*<sup>39</sup>).

Using  $z^2$  as dependent variables in the analyses allowed the estimation of the proportion of residual variance that is explained by the SNPs in GEMMA (PVE). For BW  $z^2$ , WW  $z^2$  and YW  $z^2$ , SNPs explained 4.9% (SD = 0.8), 4.7% (SD = 0.8), and 5.9% (SD = 1.1), respectively, showing that residual variance is heritable and could be changed by selection, supporting previous studies<sup>40–42</sup>. Rönnegard et al.<sup>41</sup> reported advantages in using breeding values explaining differences in residual variance (vEBV) for milk yield uniformity in Holstein cattle. According to Mulder et al.<sup>40</sup> the reduction of phenotypic variance by selecting for reduced residual variance could be more beneficial for low heritability traits.

**Meta-analysis of ecoregion-specific GWAA.** We performed a meta-analysis of ecoregion-specific GWAA (GWAA for each ecoregion separately) to look for significant differences (Cochran's Q statistics) in effect size of SNPs between the ecoregions, which are interpreted as  $G \times E$ <sup>43</sup>. In addition, we calculated the m-value statistic for the significant SNPs (Cochran's Q statistic's  $P$ -value  $< 1e-5$ ), which is the posterior probability of an effect being present in an ecoregion given the observations from all other ecoregions<sup>43</sup> allowing us to identify which ecoregions a given variant has an effect. Southeast and Desert and Arid Prairie ecoregions were not included in the meta-analysis due to their small sample sizes. Meta-analysis of ecoregion-specific GWAA detected 12  $G \times E$  SNPs for BW located on chromosome 6 (Additional file 1: Figure S17a); two  $G \times E$  SNPs for WW on chromosome 15 (Additional file 1: Figure S17b); and two  $G \times E$  SNPs for YW, one located on chromosome 13 and one on chromosome 15 (Additional file 1: Figure S17c, Additional file 7). Analyzing the PM-plot ( $P$ -values against m-values) and PB-plot ( $P$ -values against beta-values, SNP effects estimates obtained from the ecoregion-specific GWAA) of the  $G \times E$  SNPs detected by the meta-analysis, we observed that some  $G \times E$  SNPs on chromosome 6 at 30 Mb have an effect on animal's BW from Forested Mountain and Upper Midwest and Northeast ecoregions, however with different certainty (Additional file 1: Figures S18a and S19a). For WW and YW, despite that the analysis had shown a significant heterogeneity between the ecoregions, they did not show different posterior probabilities of an effect in any ecoregion (m-value), (Additional file 1: Figure S18b,c, respectively). However, large differences in the SNP effect size (beta) were noticed between ecoregions (Additional file 1: Figure S19b,c, respectively).

**Traditional genome-wide association analyses.** We also performed traditional GWAA using univariate and multivariate linear mixed models in order to compare with the  $G \times E$  GWAA results. We detected major

QTL on chromosomes 6 (26–48 Mb), 7 (88–90 Mb), 14 (15–30 Mb), and 20 (41–78 Mb) for univariate BW, WW and YW (Additional file 1: Figure S20a–c and Additional file 10). In addition, significant SNPs ( $P < 1e-5$ ) were identified on chromosomes 3 (53.6 Mb) and 17 (73 Mb) for univariate BW; and on chromosomes 1 (81.1–81.3 Mb), 2 (116.1 Mb), and 5 (96.7 Mb) for univariate WW. The GWAA using multivariate model detected QTL also identified in the univariate GWAA on chromosomes 5, 6, 7, 14, 17, and 20 in addition to on chromosomes 4 (24.5 Mb), 9 (57.9 Mb), 13 (63.3 Mb), 19 (50.5 Mb), and 26 (45.4 Mb) (Additional file 1: Figure S20d and Additional file 9). All genes detected by GWAA are included in the Additional file 10. In total, 240, 92, and 141 haplotype blocks were detected for univariate BW, WW, and YW GWAA, respectively, and 222 haplotype blocks for multivariate analysis.

Out of 948 total significant SNPs using multivariate GWAA analysis for the growth traits, 531 SNPs were located on chromosome 6 (from 26.8 to 48.8 Mb). Likewise, several SNPs interacting with Forested Mountains ecoregion, elevation, precipitation and, mainly, with the temperatures were also detected in the same positions on chromosome 6. Given the importance of chromosome 6 underlying the growth traits, we repeated the multivariate GWAA including six independent significant SNPs (e.g. *rs109849093*, *rs464458177*, *rs110305942*, *rs43459303*, *rs109278547*, and *rs137209027* located at 37,211,057; 37,234,136; 37,418,164; 37,695,352; 38,139,495; and 41,084,765 bp, respectively) as fixed effects in the model. In this way, if those SNPs are causal mutations accounting for QTL effect the loci association signals would be missed. Results showed that even including six significant SNPs (not in LD) simultaneously as fixed effects in the model, a small number of SNPs were still associated (six out of 531 SNPs), indicating that there are possibly multiple causal mutations in this region. Based on the linkage disequilibrium between adjacent significant SNP, we found 166, 49, and 95 haplotype blocks in this region for univariate BW, WW, and YW, respectively, and 131 haplotype blocks for multivariate analysis.

**Comparison of the GWA approaches.** A nominal significant threshold value ( $P < 1e-5$ ) was used to equally interpret the results of all GWAA approaches, including continuous and discrete  $G \times E$  GWAA, vGWAA, ecoregion-specific GWAA, and traditional GWAA. Therefore, the false discovery rate (FDR) was estimated for each analysis (Additional file 2: Table S2). Overall, the FDR showed high values for  $G \times E$  GWAA and vGWAA suggesting that many significant SNPs in this study may be false positives. However, QQ plots from several analyses showed that  $P$ -values lower than  $1e-5$  may show deviation from the null hypothesis, therefore, can be considered significant (Additional file 1: Figures S21–S36). The high FDR for the analyses detecting  $G \times E$  may be due to the requirement of larger sample size to detect such interactions and the small amount of variation explained by  $G \times E$  effects. Bolormaa et al.<sup>29</sup> also reported high FDR for non-additive SNP effects on growth traits in beef cattle. The FDR estimation for traditional GWAA, which detect SNPs with additive effects, was very low varying from 0.007 to 0.014. QQ plots of GWAA are displayed in Additional file 1: Figure S37.

Comparing traditional GWAA with all  $G \times E$  GWAA results, we verified that 312 SNPs with additive effects also have  $G \times E$  effects (mainly on chromosome 6, but also on chromosomes 7 and 20). After adjusting the  $G \times E$  GWAA for genomic control, 68 SNP on chromosome 6 remained significant in both analyses. Some SNPs showed  $G \times E$  effects even greater than their additive effects on the traits. For example, the SNP *rs136523580* (chromosome 6 at 30,566,685) has an additive effect of 0.41 kg for BW in the total dataset, but its  $G \times E$  effects is even more impactful (0.67 kg) on the BW of animals from the Forested Mountains ecoregion. These findings show the importance of  $G \times E$  effects on quantitative traits and the need for precision, ecoregion-specific genomic predictions to optimize animal performance.

Comparing vGWAA with the traditional GWAA results, only one vQTL, on chromosome 14 at 23.1–23.3 Mb harboring the genes *LYN*, *MOS*, *RPS20*, *PLAG1*, and *CHCHD7*, was also found as a direct effect QTL for BW, meaning that this locus affects both mean and variance of BW in this population. Five vQTL were also identified by the  $G \times E$  GWAA, which were found to interact with multiple continuous environmental variables as well as interacting with ecoregions. However, after genomic control adjustment on the  $G \times E$  GWAA, only one vQTL was also identified by the  $G \times E$  GWAA, which was found to interact with mean and minimum temperature.  $G \times E$  GWAA detected a greater number of  $G \times E$  loci when compared with the vGWAA results.

Only few  $G \times E$  SNPs were detected using the meta-analysis of ecoregion-specific GWAA (Cochran's  $Q$  statistic's  $P$ -value  $< 1e-5$ ), which were also identified by the  $G \times E$  GWAA. In addition, using the meta-analysis approach the Southeast and Desert and Arid Prairie ecoregions could not be included in the analysis due to their small sample sizes, suggesting that this was not an optimal approach to detect  $G \times E$  associations for our dataset. However, creating PM-plots (contrast between environment-specific  $P$ -value and posterior probability of an effect,  $m$ -value)<sup>44</sup>, and PB-plots (environment-specific  $P$ -value and SNP effect size) helped interpret results of our  $G \times E$  GWAA and vGWAA. For significant  $G \times E$  SNPs detected by the  $G \times E$  GWAA, there were clear differences in the SNP effect sizes between the significant ecoregion and the remaining ecoregions (Additional file 1: Figures S38 and S39). For SNPs with significant variance-heterogeneity effects (detected by vGWAA), patterns were not as consistent. While some vQTL had differences in mean effects, others did not (Additional file 1: Figures S40 and S41).

## Discussion

This  $G \times E$  study provides a more comprehensive understanding of how  $G \times E$  modulate growth traits in beef cattle and possibly in other mammalian species. We observed that  $G \times E$  explain substantial proportions of growth traits phenotypic variances, and a great number of  $G \times E$  SNPs, some with large effects, were detected affecting such traits. Detection of large  $G \times E$  SNP effects indicates that there is considerable opportunity to improve environmental resistance and performance in cattle. Our study also revealed the first evidence of vQTL controlling phenotypic variation in body weights due to  $G \times G$  or  $G \times E$  effects in beef cattle. In addition, we uncovered several biological pathways affected by  $G \times E$  that likely drive environmental adaptation.

Among the environmental variables analyzed, birth weight was most affected by temperature. Also, a greater number of  $G \times E$  QTL for birth weight were detected in animals born in the Fescue Belt and Forested Mountain ecoregions, which have high and low temperatures, respectively. The effect of higher temperature on calf birth weight may be due to the reduced feed intake or the decreased uterine blood flow and placental function often found in heat-stressed cows, factors that contribute to impaired cow-to-fetal nutrient exchange, negatively affecting fetal growth<sup>45–47</sup>. On the other hand, low temperatures can increase calf birth weight due to the fact that the cow's body directs blood flow toward internal organs, including the uterus, to conserve heat and maintain body temperature, increasing the nutrients being carried to the calf<sup>48</sup>. Weaning weight and yearling weight were more affected by elevation and Forested Mountain, followed by the High Plains, ecoregions with high elevation. Higher altitude regions are associated with reduced atmospheric oxygen (hypoxia), which causes pulmonary hypertension and vasoconstriction leading to right-side heart failure in cattle<sup>49</sup>. This condition is known as brisket disease and accounts for significant losses in growth and reproductive performance, being one of the top causes of mortality in cattle residing at high altitudes<sup>50</sup>. Therefore, the genomic regions detected interacting with Forested Mountain and High Plains ecoregions may be related to adaptation to oxygen deprivation in beef cattle.

Molecular functions involved in neurotransmission processes were enriched by  $G \times E$  genes. The release of specific neurotransmitters in the brain, such as dopamine, acetylcholine, glutamate, serotonin and GABA, is the main physiological mechanism that organizes the response to a stressful situation<sup>51–52</sup>. Specifically, GABA receptor activity and glutamatergic synapse pathway were enriched. Genes with  $G \times E$  associations with roles in the activity of the GABA receptor (*GABRA4*, and *GABRB1*) were found interacting with Fescue Belt ecoregion. This may be due to the consumption of tall fescue, a grass commonly used in grazing cattle systems in the Southeastern United States<sup>53</sup>, which usually contains an endophytic fungus that produces ergot alkaloids<sup>54</sup>. Ergot alkaloids have vasoconstrictive effects and interact with GABAergic receptor affecting multiple aspects of animal's physiology<sup>55–57</sup>. Animals consuming the endophyte-infected tall fescue forage often develop fescue toxicosis, which is characterized by reduced feed intake, weight gain, and blood flow, as well as elevated body temperature associated with heat stress<sup>58–60</sup>. The glutamatergic synapse pathway was enriched mainly by  $G \times E$  genes interacting with Forested Mountains and/or elevation (*PLA2G4B*, *PLA2G4E*, *GRIN2D*, and *GRIK2*), supporting that neurotransmitters respond to the effects of hypoxia<sup>61</sup>.

Endocytosis and VEGF signaling pathways were also enriched by candidate  $G \times E$  genes interacting with elevation or ecoregions with high altitudes. Endocytosis pathway plays an important role in cell survival under hypoxia condition<sup>62</sup>. VEGF signaling is involved in growth of blood vessels promoting angiogenesis and have an important role in response to hypoxia<sup>63</sup>. Both endocytosis and VEGF signaling pathways have previously been related to high altitude adaptation in cattle<sup>64</sup> and other species<sup>65,66</sup>.

Genes of the keratin (KRT) family (*KRT15*, *KRT31*, *KRT32*, *KRT33A*, *KRT34*, and *KRT36*) are candidates to interact with Fescue Belt ecoregion. Such  $G \times E$  genes enriched pathways involved in the keratinocyte differentiation (formation of the cornified envelope and keratinization). Such process is involved in the formation of cornified cell envelope (keratin filaments covalently attached to the cornified envelope) which provides the barrier function of epidermis helping to protect the body against environmental stimuli<sup>67</sup>. Keratin proteins are also involved in the hair-fiber formation cyclical process, which involves growth, regression, resting phases and shedding of the hair shaft<sup>68</sup>. In cattle, hair shedding is an adaptive process that help the animal to cope with heat stress, which is aggravated by fescue toxicosis condition<sup>69</sup>. Durbin et al.<sup>70</sup> detected genes of the KRT family associated with hair shedding in Angus cattle reared in the Fescue Belt ecoregion. Interestingly, the KRT family genes also enriched the estrogen signaling pathway, which is involved in fertility in cattle<sup>71</sup>. This result may explain, at least in part, fertility problems in heifers grazing endophyte-infected tall fescue<sup>72</sup>.

Genes that are candidates for  $G \times E$  associations detected in the Upper Midwest & Northeast ecoregion may be involved in metabolic adjustment to adapt to cold weather. Steroid hormone biosynthesis, retinol metabolism and bile secretion are involved in the regulation of thermogenic processes<sup>73–75</sup>. In addition, ascorbate and aldarate metabolism, and pentose and glucuronate interconversions have previously been involved in molecular mechanisms responsive to cold stress<sup>76</sup>.

Pathways related to immune responses, as *Staphylococcus aureus* infection and viral myocarditis were also enriched. Immune functions reflect the involvement of innate and adaptive immune systems in the activation of host defense reactions to deal with pathogens<sup>77</sup>. Many of these immune  $G \times E$  genes have been associated with diseases caused by pathogens leading to reduced productivity in cattle<sup>78–88</sup>. This result is in agreement that heterogeneity in the immune response to infectious diseases across populations living in different environmental conditions is under genetic control<sup>89</sup>. In addition, pathogen resistance increases host fitness and has driven the evolution of plant and animals through time<sup>90</sup>.

One of our major findings is that the ecoregions showed specific  $G \times E$  with large effects on the traits. In addition, ecoregion  $G \times E$  GWAA detected  $G \times E$  loci that were not identified by the continuous  $G \times E$  GWAA. One possible explanation for these results is that when analyzing ecoregions, we take into account not only climate and topography, but also unmeasured variables like forage quality and local pathogens, as well as the combination of them. This information allowed us to identify environmentally sensitive genotypes. It indicates that the same trait can be influenced by different genes in the various ecoregions which may affect particular physiological and behavioral functions leading to local adaptation. This assumption is supported by the GO and pathways enrichment analysis of the  $G \times E$  genes identified for each ecoregion showing that the diverse ecoregion environment affects different biological mechanisms. We also found that major additive QTL also carry  $G \times E$  effects on growth traits and that some of their  $G \times E$  effects are greater than their additive effects. Additionally, novel  $G \times E$  associations were detected in multiple genomic regions. These findings support the importance of creating ecoregion-specific genomic predictions to identify local resilient or adapted animals, using the  $G \times E$  SNPs detected in this study to assist selection decisions which would lead to an increase in animal performance



within an ecoregion. However, the lack of overlap of  $G \times E$  QTL between the ecoregions may reflect a lack of statistical power to detect such interactions due to the different sample size within the ecoregions.

We showed that vGWAA may be able to detect loci enriched with  $G \times E$  effects without requiring prior knowledge of the interacting environmental factors. A vGWAA may capture  $G \times E$  loci that interact with multiple environmental variables, possibly even those not included in the  $G \times E$  GWAA model (e.g. water availability, pasture condition, pathogens). When multiple interacting factors induce variance heterogeneity, the power of identification of any single one of them or all together may be lower than the power of the variance heterogeneity test<sup>91</sup>, which may explain why vGWAA identified  $G \times E$  loci that were not detected by direct  $G \times E$  GWAA. We also observed that some vQTL have differences in means across ecoregions, however, not all do. This suggests that some vQTL could be the difference between robust genotypes (little variation, flat reaction norms) across environments while other genotypes at the same locus are more variable across environments, indicating that selection for low variance vQTL genotypes may provide sustainability advantages. Therefore, genetic variation in residual variance may be utilized to increase uniformity of production in livestock populations<sup>40</sup>. Improvement of uniformity through selection is particularly interesting for traits with an intermediate optimum value (e.g. litter size in sheep, egg weight in laying hens, carcass weight and quality traits in pigs and broilers, and maternal growth in beef), or for decreasing variation in performance within sire's daughter group<sup>40,41</sup>. Nonetheless, many genes harboring the vQTL identified here were previously reported to be involved in environmental adaptation in cattle<sup>35–39</sup> supporting that phenotypic variability may be an adaptive evolutionary solution to environmental changes<sup>92</sup>. However, our results support Paré et al.<sup>93</sup> which reported that when an environmental factor of interest has been measured on all the genotyped individuals, the direct  $G \times E$  GWAA will be able to detect a larger number of  $G \times E$  loci.

Our study also demonstrated the challenges of detecting  $G \times E$  effects. Although we have used a relatively large beef cattle population to date, it still seems that a greater number of animals representing all possible environments are needed to detect such interactions. Also, several  $G \times E$  QTL have small effect sizes on quantitative traits increasing the difficulty to differentiate them from potential bias. These limitations directly impact the statistical power to detect  $G \times E$  as previously reported. For example,  $G \times E$  studies observed unclear evidence of physical activity modulating the effects of *FTO* gene variants on obesity in humans<sup>94–96</sup> until a study with large sample size (~218,000 individuals) confirmed this interaction<sup>97</sup>. Therefore, studies using a larger population size could validate our results.

The environmentally affected biological mechanisms that modulate growth traits identified in this study have been reported to be involved in resilience<sup>12,52,77</sup>, and adaptation processes<sup>10,49,64,98,99</sup>, revealing the role that  $G \times E$  plays in adaptation and evolution at the level of the genome in cattle and potentially across mammals. It is important to understand such biological mechanisms since the inability of an animal to cope with its environment results in failure to produce optimally leading to lost potential profitability of livestock operations<sup>12,50,100,101</sup>. Further, these biological insights may help us understand how adaptation occurs in wild populations and the potential effect of climate change on animal physiological functions.

## Conclusions

$G \times E$  GWAA and vGWAA detected several loci affecting growth traits that the traditional GWAA did not, showing that functional loci may have non-additive genetic effects between genotype classes regardless of differences in genotypic means. We revealed biological mechanisms by which beef cattle respond to changes in their environment. Neurotransmitter activity, hypoxia-induced processes, keratinization, hormone, thermogenic and immune pathways were highlighted, indicating that climate change may become a burden on animal health and productivity. In addition, ecoregion-specific  $G \times E$  SNPs detected in this study may play a crucial role in resilience and adaptation of beef cattle to divergent environments. Although conventional genomic selection has shown great promise in improving genetic gain, it only considers additive effects and, according to our findings, ecoregion-specific genomic predictions should be created to identify animals best suited for given environments. This could help producers in making optimal decisions given their geographical location. Our results revealed novel trait associations and alternative genetic mechanisms involved in shaping the total phenotypic variance of growth traits providing new insights into how the environment influences such traits and adaptation in beef cattle and other mammalian species.

## Methods

**Phenotype, genotype and environmental data.** Data from the American Simmental Association males and females born between 1975 and 2016 and genotyped for genomic-enhanced expected progeny differences (EPDs, i.e. genomic predictions) were transferred to the University of Missouri. Bulls who were born decades ago were genotyped from cryopreserved semen used for artificial insemination. The traits studied were birth weight (BW), 205-d adjusted weaning weight (WW) and 365-d adjusted yearling weight (YW), with the traits also adjusted for age of dam as per breed association practice. The animals were assigned to contemporary groups (CG), defined by the combination of farm, season (spring vs. fall) and year of birth.

The DNA samples were genotyped with various low-density assays (GeneSeek GGP-LDv3, GeneSeek GGP-LDv4, GeneSeek GGP-90KT, GeneSeek GGP-HDv3, GeneSeek GGP-F250, IlluminaSNP50, and IlluminaHD) and were imputed to the combination of Illumina BovineHD (Illumina, San Diego, CA) and the GeneSeek Genomic Profiler F250 (GeneSeek, Lincoln, NE) according to Rowan et al.<sup>102</sup>. Briefly, the imputation pipeline begins with genotype filtering to remove non-autosomal variants, as well as both variants and individuals with call rates less than 0.90, using PLINK 1.9<sup>103</sup>. SNP positions were based on the ARS-UCD1.2 Bovine reference genome assembly<sup>104</sup> and all genotypes were coded as reference versus alternate alleles as A, C, G, and T bases. The genotypes were then phased using Eagle 2.4<sup>105</sup> and imputed using Minimac3<sup>106</sup>. The final imputed genotype

data contained a total of 835,947 bi-allelic variants with imputation accuracy of 99.6%<sup>102</sup>. Variants with minor allele frequency less than 0.01 were removed using PLINK 1.9 for further analysis, leaving 710,202, 709,054, and 706,771 SNP markers for 13,427, 11,847, and 8546 animals for BW, WW and YW, respectively.

Phenotypes were adjusted for fixed effect of sex and random effect of CG. A random maternal permanent environmental effect was also included in the WW adjustment to account for differences in milk production and maternal care of the animal's dam. For G × E GWAS models, birth weights were analyzed twice, once adjusted for a maternal effect and once without. Fixed effect of sex was estimated using '--reml-est-fix' option, and random effects were predicted by the BLUP method using '--reml-pred-rand' flag while controlling for population stratification using single-trait animal model implemented in GCTA software<sup>107</sup>. Contemporary groups were included in the univariate model as random effect since 19% of individuals analyzed belong to CG with only a single animal<sup>108</sup>. After removing confounding management effects, the pre-adjusted phenotypes were used in the downstream analyses.

Eight continuous environmental variables were used for the G × E GWAA, including information of 30-year normal of minimum, mean, maximum temperatures and mean dew point temperature (°C); elevation (m); precipitation (ml); and minimum and maximum vapor pressure deficit (hPs) data, which were drawn from the PRISM climate dataset<sup>109</sup>. The United States was partitioned into nine regions (Fig. 2f, top) based on similar topographic and environmental conditions (ecoregions), including mean temperature, elevation and precipitation information, using k-means clustering implemented in 'RStoolbox' R package<sup>110,111</sup>. The optimal ecoregions were identified using 'pamk' function from the R package 'fpc'<sup>112</sup>. The ecoregions were named Desert (D), High Plains (HP), Arid Prairie (AP), Forested Mountains (FM), Upper Midwest and Northeast (UN), Southeast (SE), Rainforest (R), Foothills (FH), and Fescue Belt (FB)<sup>10</sup>. Animals were assigned to the ecoregions based on the zip code of their breeder (Fig. 2f, bottom), and only ecoregions with more than 200 records were analyzed. Thus, animals located in the Rainforest ecoregion were removed from the dataset, and those that were reared in the Desert and Arid Prairie ecoregions were analyzed together (in a group referred to as DA) due to their similar environmental conditions. No animal record was assigned to the Foothills ecoregion. A complete description of the dataset, including weight records and environmental variable measurements is shown in Additional file 2: Table S2.

**Estimation of G × E variance components.** In order to estimate the amount of phenotypic variance explained by the G × E interactions among the ecoregions, G × E variance component for pre-adjusted BW, WW, and YW were estimated. This analysis was performed using multi-component restricted maximum likelihood (REML) approach, implemented in GCTA software, considering the ecoregions as an environmental factor in the univariate linear mixed model as follows:

$$y^* = \mu + eco + u + gxe + e$$

where  $y^*$  is the vector of pre-adjusted phenotypes;  $\mu$  is the overall mean;  $eco$  is the fixed effects of the environmental factor (ecoregion);  $u$  is the vector of random additive genetic effects;  $G \times E$  is the vector of random G × E effects; and  $e$  is the vector of random residuals. Assuming that  $u \sim N(0, G\sigma_a^2)$ ,  $gxe \sim N(0, G_{gxe}\sigma_{gxe}^2)$ ,  $e \sim N(0, I\sigma_e^2)$ , where  $\sigma_a^2$ ,  $\sigma_{gxe}^2$ , and  $\sigma_e^2$  are the additive, genotype-by-environment, and residual variances, respectively.  $I$  is the identity matrix and  $G$  is the genomic relationship matrix. The  $G$  matrix was calculated using SNP marker information according to<sup>107</sup> as:  $G = W_G W_G' / N$ , where  $w_{G(i)} = (x_{G(i)} - 2p_i) / \sqrt{2p_i(1-p_i)}$ , with  $x_{G(i)}$  being 0, 1, or 2 for genotypes AA, AB or BB, respectively;  $p_i$  is the reference allele frequency at SNP  $i$ ; and  $N$  is the number of SNPs.  $G_{G \times E[i,j]}$  is  $G_{[i,j]}$  if  $i$  and  $j$  are from the same ecoregion and zero otherwise<sup>113</sup>.

**Genome-wide association analyses (GWAA).** Univariate linear mixed model analyses were performed for the pre-adjusted BW, WW or YW using GEMMA software<sup>114</sup> as follows:

$$y^* = 1\mu + X\beta + Zu + e$$

where  $y^*$  is an  $n$ -vector of pre-adjusted phenotypes;  $\mu$  is the overall mean;  $X$  is the incidence matrix of genotypes;  $\beta$  is the genotype effects;  $Z$  is the design matrix for random animal effects,  $u$  is an  $n$ -vector of random additive genetic effects; and  $e$  is an  $n$ -vector of random residual. Assuming that  $u \sim N(0, G\sigma_a^2)$  and  $e \sim N(0, I\sigma_e^2)$ ,  $G$  is the  $n \times n$  genomic relationship matrix;  $\sigma_a^2$  is the additive genetic variance;  $\sigma_e^2$  is the residual variance component;  $I$  is an  $n \times n$  identity matrix. Genomic relationship matrices were estimated using the standardized genotypes of 710,202, 709,054, and 706,771 SNP markers for 13,427, 11,847, and 8546 animals for BW, WW and YW, respectively.

We also fitted a multivariate linear mixed model for the pre-adjusted BW, WW and YW using GEMMA<sup>115</sup> in the following form:

$$Y^* = 1\mu + X\beta + ZU + E$$

where  $Y^*$  is an  $n \times d$  matrix of  $d$  pre-adjusted phenotypes for  $n$  individuals;  $\mu$  is the overall mean;  $X$  is the incidence matrix of genotypes;  $\beta$  is a  $d$  vector of the genotype effects for the  $d$  phenotypes;  $Z$  is the design matrix for random animal effects;  $U$  is an  $n \times d$  matrix of random additive genetic effects; and  $E$  is an  $n \times d$  matrix of random residuals.  $U \sim N(0, G \otimes V_a)$  and  $E \sim N(0, I \otimes V_e)$ ,  $G$  is the  $n \times n$  genomic relationship matrix;  $V_a$  is the  $d \times d$  additive genetic (co)variance matrix;  $V_e$  is a  $d \times d$  residual (co)variance matrix;  $I$  is an  $n \times n$  identity matrix.  $\otimes$  denotes Kronecker product operation. For all GWAA performed, single-marker  $P$ -values were used to generate Manhattan and QQ plots using 'manhattan' and 'qq' functions, respectively, implemented in 'qqman' R package<sup>116</sup>.

**Genotype-by-environment GWAA.** The  $G \times E$  GWAA were performed using GEMMA through two approaches: using eight continuous environmental variables (i.e. minimum, mean, maximum and mean dew point temperatures; elevation; precipitation; and minimum and maximum vapor pressure deficit) separately; and using the ecoregions as discrete environmental variable, where each ecoregion was compared against the total dataset using 0 and 1 dummy coding. In the  $G \times E$  GWAA, for each SNP in turn, GEMMA fits a linear mixed model that controls both the SNP main effect and environmental main effect, while testing for the interaction effect and controlling for population stratification, evaluating the alternative hypothesis ( $H_1: x_i * x_j \beta_{ij} \neq 0$ ) against the null hypothesis ( $H_0: x_i * x_j \beta_{ij} = 0$ ) for each interaction, therefore the resulting  $P$ -values correspond to the significance of the  $G \times E$  interaction<sup>31</sup>. This model can be described as:

$$y^* = 1\mu + x_i\beta_i + x_j\beta_j + x_i * x_j\beta_{ij} + u + e$$

where  $y^*$  is an  $n$ -vector of pre-adjusted phenotypes;  $x_i$  is an  $n$ -vector of the genotypes of the SNP  $i$ ;  $\beta_i$  is the additive effect of the SNP  $i$ ;  $x_j$  is an  $n$ -vector of environmental variable  $j$ ;  $\beta_j$  is the fixed effect of the environmental variable  $j$ ;  $\beta_{ij}$  is the interaction effect between the genotypes of the SNP  $i$  and the environmental variable  $j$ ;  $u$  is an  $n$ -vector of random additive genetic effects [ $u \sim N(0, G\sigma_a^2)$ ]; and  $e$  is an  $n$ -vector of random residual [ $e \sim N(0, I\sigma_e^2)$ ].  $G \times E$  GWAA using multivariate linear mixed model were also performed for all ecoregions but Desert and Arid Prairie due to small sample size.

**Variance-heterogeneity GWAA.** Variance-heterogeneity GWAA (vGWAA) can detect loci affecting the difference in variance between genotypes (vQTL), which can be explained by  $G \times E$  or  $G \times G$  (epistatic) interactions<sup>27</sup>. Therefore, other possible effects on the phenotypes should be removed in the model in order to detect  $G \times E$  using vGWAA approach. To do this, we estimated additive, dominance and epistasis variances for BW, WW, and YW using REML approach in univariate linear mixed models, using GCTA software. We fitted two different models (M): including only additive effects (MA), or using additive, dominance, and epistasis effects (MADE) as follows:

$$y^* = 1\mu + Z_u u + e_A \quad (\text{MA})$$

$$y^* = 1\mu + Z_u u + Z_d d + Z_{ep} ep + e_{ADE} \quad (\text{MADE})$$

where  $y^*$  is the vector of pre-adjusted phenotypes;  $\mu$  is the overall mean;  $u$  is the vector of random additive genetic effects;  $d$  is the vector of random dominance effects;  $ep$  is the vector of random epistatic effects; and  $e_A$  and  $e_{ADE}$  are vectors of random residuals for MA and MADE models, respectively.  $Z$  matrices are incidence matrices relating observations to animals. Assuming that  $u \sim N(0, G\sigma_a^2)$ ,  $d \sim N(0, D\sigma_d^2)$ ,  $ep \sim N(0, E\sigma_{ep}^2)$ ,  $e_A \sim N(0, I\sigma_{e_A}^2)$ ,  $e_{ADE} \sim N(0, I\sigma_{e_{ADE}}^2)$ , where  $\sigma_a^2$ ,  $\sigma_d^2$ ,  $\sigma_{ep}^2$ ,  $\sigma_{e_A}^2$ , and  $\sigma_{e_{ADE}}^2$  are the additive, dominance, epistatic, MA residual, and MADE residual variances, respectively.  $I$  is the identity matrix; and  $G$ ,  $D$ , and  $E$  are the additive (genomic), dominance and epistatic genetic relationship matrices, respectively. The  $G$  and  $D$  matrices were calculated using SNP marker information<sup>107,117</sup> as:  $G = W_G W_G' / N$ , and  $D = W_D W_D' / N$ , where  $w_{G(i)} = (x_{G(i)} - 2p_i) / \sqrt{2p_i(1-p_i)}$  and  $w_{D(i)} = (x_{D(i)}^2 - 2p_i^2) / [2p_i(1-p_i)]$  with  $x_G$  being 0, 1, or 2 and  $x_D$  being 0,  $2p$ , or  $(4p-2)$  for genotypes AA, AB or BB, respectively;  $p_i$  is the reference allele frequency at SNP  $i$ ; and  $N$  is the number of SNPs. According to Henderson<sup>118</sup>  $E$  can be derived from the additive genomic relationship matrix as:  $E = G \# G$ , where  $\#$  denotes the Hadamard product operation.

Residual effects ( $e_A$  and  $e_{ADE}$ ) from the MA and MADE models, respectively, were standardized to z score by the rank-based inverse-normal transformation and squared ( $z^2$ ), which is a measure of phenotypic variance<sup>119,120</sup>. The vGWAA were performed for BW  $z^2$ , WW  $z^2$ , and YW  $z^2$  using univariate and multivariate linear mixed models implemented in GEMMA software as denoted in “Genome-wide association analyses (GWAA)” section, where  $y^*$  is an  $n$ -vector of  $z^2$ ; and  $\beta_i$  is the effect size of the SNP on  $z^2$ .

**Meta-analysis of ecoregion-specific GWAA.** In order to evaluate how significantly the effect sizes of SNPs (Cochran’s Q test) varied between the ecoregions (indicative of  $G \times E$ )<sup>43</sup> we combined the outputs from univariate within-ecoregion GWAA (GWAA for each ecoregion separately) into a single meta-analysis for each phenotype using METASOFT software<sup>121</sup>. Only ecoregions with more than 1000 animals were analyzed. In addition, the statistic m-value (posterior probability of an effect) for the significant SNPs (Cochran’s Q statistic’s  $P$ -value  $< 1e-5$ ) were calculated, which is an estimate of the posterior probability of a locus having an effect in a particular ecoregion<sup>122</sup>. The results were visualized through PM-plots and PB-plots, in which ecoregion-specific  $P$ -values are simultaneously visualized with the m-values and beta-values (SNP effects estimates obtained from the ecoregion-specific GWAA) at each tested locus, respectively.

**Genomic control and false discovery rate estimation.** In  $G \times E$  GWAA samples can have shared environmental effects outside the environmental variable used in the  $G \times E$  GWAA which may cause spurious signals<sup>13</sup> similar to cryptic relatedness or population structure. Thus, we calculated the genomic control inflation factor ( $\lambda_{GC}$ ) for all analyses<sup>123</sup> to improve the calibration of the  $G \times E$  GWAS  $P$ -values. For those  $G \times E$  GWAA with  $\lambda_{GC} > 1$ ,  $P$ -values were divided by the  $\lambda_{GC}$ .

A nominal significant threshold value ( $P < 1e-5$ ) was used to equally compare the results of all GWAA approaches, including continuous and discrete  $G \times E$  GWAA, vGWAA, ecoregion-specific GWAA, and traditional GWAA. Therefore, false discovery rate (FDR) of the nominal significant threshold ( $P < 1e-5$ ) was estimated for all analyses performed (Additional file 2: Table S2), based on Bolormaa et al.<sup>124</sup>, as:

$FDR = P(1 - \frac{A}{T}) / (\frac{A}{T})(1 - P)$ , where  $P$  is the nominal significant threshold (e.g.,  $1e-5$ );  $A$  is the number of SNP that were significant at the nominal significant threshold; and  $T$  is the total number of SNPs tested. Although the true FDR cannot exceed 1, the estimated FDR can exceed 1 if the number of significant SNPs is smaller than expected by chance, meaning that all SNPs that met the nominal significant threshold are likely false positives<sup>29</sup>.

**QTL definition.** For all analysis done in this study, QTL were defined based on haplotype blocks<sup>125</sup> using HaploView software, version 4.2<sup>126</sup>. This methodology defines that pairs of SNP markers are in strong linkage disequilibrium (LD) if the confidence interval is equal to or greater than 95% with  $D'$  equal to 0.98 and the lower limit above 0.70.

**Functional enrichment analyses.** We explored 100 kb sequence windows that flanked the significant SNPs ( $P < 1e-5$ ) to scan for putative G × E genes located in their vicinity and to identify possible regulatory elements based on the ARS-UCD1.2 bovine reference genomic positions. Enrichment analysis using the G × E candidate genes based on Gene Ontology (GO), and KEGG<sup>127</sup> and Reactome pathways were performed using g:Profiler<sup>128</sup>, in which the analyses were adjusted for FDR of 10%. GO annotations were also retrieved from the AmiGO browser<sup>129</sup>.

**Ethics approval and consent to participate.** Animal Care and Use Committee approval was not required or obtained for data that were extracted from the existing American Simmental Association database.

### Data availability

The raw data used in this research are available from the American Simmental Association under a Data Use Agreement, but are not publicly available. Derived data (analytical results) are however available as supplementary files associated with this publication.

Received: 14 September 2020; Accepted: 7 June 2021

Published online: 25 June 2021

### References

- Kolmodin, R., Strandberg, E., Madsen, P., Jensen, J. & Jorjani, H. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. *Acta Agric. Scand. Sect. A Anim. Sci.* **52**, 11–24 (2002).
- Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* 4th edn. (Longmans Green, 1996).
- Saltz, J. B. *et al.* Why does the magnitude of genotype-by-environment interaction vary?. *Ecol. Evol.* **8**, 6342–6353 (2018).
- Decker, J. E. Agricultural genomics: Commercial applications bring increased basic research power. *PLoS Genet.* **11**, e1005621 (2015).
- Howard, J. T. *et al.* Beef cattle body temperature during climatic stress: A genome-wide association study. *Int. J. Biometeorol.* **58**, 1665–1672 (2014).
- Williams, J. L., Bertrand, J. K., Misztal, I. & Lukaszewicz, M. Genotype by environment interaction for growth due to altitude in United States Angus cattle. *J. Anim. Sci.* **90**, 2152–2158 (2012).
- Bradford, H. L., Fragomeni, B. O., Bertrand, J. K., Lourenco, D. A. L. & Misztal, I. Genetic evaluations for growth heat tolerance in Angus cattle. *J. Anim. Sci.* **94**, 4143–4150 (2016).
- Guidolin, D. G. F. *et al.* Genotype-environment interaction for post-weaning traits in Nelore beef cattle. *Anim. Prod. Sci.* **52**, 975–980 (2012).
- MacLeod, I. M. *et al.* Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144 (2016).
- Rowan, T. N. *et al.* Powerful detection of polygenic selection and environmental adaptation in US beef cattle populations. *bioRxiv* <https://doi.org/10.1101/2020.03.11.988121> (2020).
- Ferrell, C. L. *Factors Influencing Fetal Growth and Birth Weight in Cattle*, 132 (Rom. L. Hruska U.S. Meat Anim. Res. Cent., 1993).
- Collier, R. J., Baumgard, L. H., Zimbelman, R. B. & Xiao, Y. Heat stress: Physiology of acclimation and adaptation. *Anim Front.* **9**, 12–19 (2019).
- Sul, J. H. *et al.* Accounting for population structure in gene-by-environment interactions in genome-wide association studies using mixed models. *PLoS Genet.* **12**, e1005849 (2016).
- Beaumont, R. N. *et al.* Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27**, 742–756 (2018).
- Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
- Sharma, R. K. *et al.* Uterine environment as a regulator of birth weight and body dimensions of newborn lambs1. *J. Anim. Sci.* **90**, 1338–1348 (2012).
- Cross, A. J., Keel, B. N., Brown-Brandl, T. M., Cassady, J. P. & Rohrer, G. A. Genome-wide association of changes in swine feeding behaviour due to heat stress. *Genet. Sel. Evol.* **50**, 11 (2018).
- Halstead, M. M., Ma, X., Schultz, R. M. & Ross, P. J. Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. *bioRxiv* 2019.12.12.874479 (2019).
- Moore, T. & Haig, D. Genomic imprinting in mammalian development: A parental tug-of-war. *Trends Genet.* **7**, 45–49 (1991).
- Peretz, H. *et al.* Cys 618 Arg mutation in the RET proto-oncogene associated with familial medullary thyroid carcinoma and maternally transmitted Hirschsprung's disease suggesting a role for imprinting. *Hum. Mutat.* **10**, 155–159 (1997).
- Takacova, M. *et al.* Expression pattern of carbonic anhydrase IX in medullary thyroid carcinoma supports a role for RET-mediated activation of the HIF pathway. *Am. J. Pathol.* **184**, 953–965 (2014).
- Leiser, R., Krebs, C., Ebert, B. & Dantzer, V. Placental vascular corrosion cast studies: A comparison between ruminants and humans. *Microsc. Res. Tech.* **38**, 76–87 (1997).
- Pfarrer, C. D. *et al.* Localization of vascular endothelial growth factor (VEGF) and its receptors VEGFR-1 and VEGFR-2 in bovine placentomes from implantation until term. *Placenta* **27**, 889–898 (2006).
- Hoffert-Goeres, K. A. *et al.* Angiogenesis in day-30 bovine pregnancies derived from nuclear transfer. *Cloning Stem Cells* **9**, 595–607 (2007).

25. Goyal, R., Van Wickle, J., Goyal, D., Matei, N. & Longo, L. D. Antenatal maternal long-term hypoxia: Acclimatization responses with altered gene expression in ovine fetal carotid arteries. *PLoS One* **8**, e82200 (2013).
26. Stallings, J. D. *et al.* Patterns of gene expression associated with recovery and injury in heat-stressed rats. *BMC Genomics* **15**, 1058 (2014).
27. Rönnegård, L. & Valdar, W. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet.* **13**, 63 (2012).
28. Visscher, P. M. & Posthuma, D. Statistical power to detect genetic loci affecting environmental sensitivity. *Behav. Genet.* **40**, 728–733 (2010).
29. Bolormaa, S. *et al.* Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. *Genet. Sel. Evol.* **47**, 26 (2015).
30. Voz, M. L. *et al.* Microarray screening for target genes of the proto-oncogene PLAG1. *Oncogene* **23**, 179–191 (2004).
31. Smith, J. L. *et al.* Genome-wide association and genotype by environment interactions for growth traits in U.S. Gelbvieh cattle. *BMC Genomics* **20**, 926 (2019).
32. Saatchi, M., Schnabel, R. D., Taylor, J. F. & Garrick, D. J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* **15**, 442 (2014).
33. Seabury, C. M. *et al.* Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* **18**, 386 (2017).
34. Brown, A. A. *et al.* Genetic interactions affecting human gene expression identified by variance association mapping. *Elife* **2014**, e01381 (2014).
35. Porto-Neto, L. R. *et al.* Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* **14**, 876 (2013).
36. Taye, M. *et al.* Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mamm. Genome* **28**, 528–541 (2017).
37. Pitt, D. *et al.* Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evol. Appl.* **12**, 105–122 (2019).
38. Liu, S. *et al.* Identifying hub genes for heat tolerance in water buffalo (*Bubalus bubalis*) using transcriptome data. *Front. Genet.* **10**, 209 (2019).
39. Dikmen, S., Cole, J. B., Null, D. J. & Hansen, P. J. Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. *PLoS One* **8**, e69202 (2013).
40. Mulder, H. A., Bijma, P. & Hill, W. G. Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genet. Sel. Evol.* **40**, 37–59 (2008).
41. Rönnegård, L., Felleki, M., Fikse, W. F., Mulder, H. A. & Strandberg, E. Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *J. Dairy Sci.* **96**, 2627–2636 (2013).
42. Gonzalez, P. N. *et al.* Genetic structure of phenotypic robustness in the collaborative cross mouse diallel panel. *J. Evol. Biol.* **29**, 1737–1751 (2016).
43. Kang, E. Y., Han, B., Furlotte, N., Joo, J. W. & Shih, D. Meta-analysis identifies gene-by-environment interactions as demonstrated in a study of 4,965 mice. *PLoS Genet.* **10**, e1004022 (2014).
44. Kang, E. Y. *et al.* ForestPMPlot: A flexible tool for visualizing heterogeneity between studies in meta-analysis. *G3 Genes Genomes Genet.* **6**, 1793–1798 (2016).
45. Reynolds, L. P., Ferrell, C. L., Nienaber, J. A. & Ford, S. P. Effects of chronic environmental heat stress on blood flow and nutrient uptake of the gravid bovine uterus and foetus. *J. Agric. Sci.* **104**, 289–297 (2020).
46. Jones, H. N., Powell, T. L. & Jansson, T. Regulation of placental nutrient transport—A review. *Placenta* **28**, 763–774 (2007).
47. Dunlap, K. A., Brown, J. D., Keith, A. B. & Satterfield, M. C. Factors controlling nutrient availability to the developing fetus in ruminants. *J. Anim. Sci. Biotechnol.* **6**, 1–10 (2015).
48. Deutscher, G. H., Colburn, D. & Davis, R. *Climate Affects Calf Birth Weights and Calving Difficulty*, 400 (Nebraska Beef Cattle Reports, 1999).
49. Newman, J. H. *et al.* High-altitude pulmonary hypertension in cattle (Brisket disease): Candidate genes and gene expression profiling of peripheral blood mononuclear cells. *Pulm. Circ.* **1**, 462–469 (2011).
50. Holt, T. N. & Callan, R. J. Pulmonary arterial pressure testing for high mountain disease in cattle. *Vet. Clin. North Am. Food Anim. Pract.* **23**, 575–596 (2007).
51. Joëls, M. & Baram, T. Z. The neuro-symphony of stress. *Nat. Rev. Neurosci.* **10**, 459–466 (2009).
52. Mora, F., Segovia, G., Del Arco, A., De Blas, M. & Garrido, P. Stress, neurotransmitters, corticosterone and body-brain integration. *Brain Res.* **1476**, 71–85 (2012).
53. Stuedemann JA, Seman DH. *Integrating Genetics, Environment, and Management to Minimize Animal Toxicoses. Neotyphodium Cool Grasses*, 305–324 (2008).
54. McCarty, K. *et al.* Exposure to toxic fescue during late gestation on beef cows: Effects on cow performance and offspring performance from birth to weaning. *EC Vet. Sci.* **5**, 01–15 (2020).
55. Haddjeri, N., Seletti, B., Gilbert, F., De Montigny, C. & Blier, P. Effect of ergotamine on serotonin-mediated responses in the rodent and human brain. *Neuropsychopharmacology* **19**, 365–380 (1998).
56. Xue, Y., Liao, S. F., Strickland, J. R., Boling, J. A. & Matthews, J. C. Bovine neuronal vesicular glutamate transporter activity is inhibited by ergovaline and other ergopeptines. *J. Dairy Sci.* **94**, 3331–3341 (2011).
57. Foote, A. P. *et al.* Constriction of bovine vasculature caused by endophyte-infected tall fescue seed extract is similar to pure ergovaline. *J. Anim. Sci.* **90**, 1603–1609 (2012).
58. Jackson, J. A. *et al.* Summer fescue toxicity in dairy steers fed tall fescue seed. *J. Anim. Sci.* **58**, 1057–1061 (1984).
59. Hoveland, C. S. Importance and economic significance of the Acremonium endophytes to performance of animals and grass plant. *Agric. Ecosyst. Environ.* **44**, 3–12 (1993).
60. Hemken, R. W. *et al.* Interaction of environmental temperature and anti-quality factors on the severity of summer fescue toxicosis. *J. Anim. Sci.* **52**, 710–714 (1981).
61. Davis, J. N. & Carlsson, A. The effect of hypoxia on monoamine synthesis, levels and metabolism in rat brain. *J. Neurochem.* **21**, 783–790 (1973).
62. López-Hernández, T., Haucke, V. & Maritzen, T. Endocytosis in the adaptation to cellular stress. *Cell Stress* **4**, 230–247 (2020).
63. Nagy, J. A., Benjamin, L., Zeng, H., Dvorak, A. M. & Dvorak, H. F. Vascular permeability, vascular hyperpermeability and angiogenesis. *Angiogenesis* **11**, 109–119 (2008).
64. Verma, P. *et al.* Transcriptome analysis of circulating PBMCs to understand mechanism of high altitude adaptation in native cattle of Ladakh region. *Sci. Rep.* **8**, 7681 (2018).
65. Sun, Y. B. *et al.* Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. *Proc. Natl. Acad. Sci.* **115**, E10634–E10641 (2018).
66. Hu, Y., Su, J., Cheng, L., Lan, D. & Li, D. Pectoral muscle transcriptome analyses reveal high-altitude adaptations in Tibetan chickens. *Anim. Biol.* **70**, 385–400 (2020).
67. Kalinin, A. E., Kajava, A. V. & Steinert, P. M. Epithelial barrier function: Assembly and structural features of the cornified cell envelope. *BioEssays* **24**, 789–800 (2002).

68. Khan, I. *et al.* Mammalian keratin associated proteins (KRTAPs) subgenomes: Disentangling hair diversity and adaptation to terrestrial and aquatic environments. *BMC Genomics* **15**, 779 (2014).
69. Aiken, G. E., Klotz, J. L., Looper, M. L., Tabler, S. F. & Schrick, F. N. Disrupted hair follicle activity in cattle grazing endophyte-infected tall fescue in the summer insulates core body temperatures. *Prof. Anim. Sci.* **27**, 336–343 (2011).
70. Durbin, H. J., Lu, D., Yampara-Iquise, H., Miller, S. P. & Decker, J. E. Development of a genetic evaluation for hair shedding in American Angus cattle to improve thermotolerance. *Genet. Sel. Evol.* **52**, 63 (2020).
71. Moraes, J. G. N. *et al.* Uterine influences on conceptus development in fertility-classified animals. *Proc. Natl. Acad. Sci.* **115**, E1749–E1758 (2018).
72. Porter, J. K. & Thompson, F. N. Effects of fescue toxicosis on reproduction in livestock. *J. Anim. Sci.* **70**, 1594–1603 (1992).
73. Sanin, Y. L., Cabrera, A. M. Z. & Morales, A. M. T. Adaptive responses to thermal stress in mammals. *Rev. Med. Vet.* **31**, 121–135 (2015).
74. Herz, C. T. & Kiefer, F. W. The transcriptional role of vitamin A and the retinoid axis in brown fat function. *Front. Endocrinol.* **11**, 608 (2020).
75. Wua, D. *et al.* Effects and transcriptional responses in the hepatopancreas of red claw crayfish *Cherax quadricarinatus* under cold stress. *J. Therm. Biol.* **85**, 102404 (2019).
76. Liu, L., Zhang, R., Wang, X., Zhu, H. & Tian, Z. Transcriptome analysis reveals molecular mechanisms responsive to acute cold stress in the tropical stenothermal fish tiger barb (*Puntius tetrazona*). *BMC Genomics* **21**, 737 (2020).
77. Colditz, I. G. & Hine, B. C. Resilience in farm animals: Biology, management, breeding and implications for animal welfare. *Anim. Prod. Sci.* **56**, 1961–1983 (2016).
78. Li, R. W., Li, C. & Gasbarre, L. C. The vitamin D receptor and inducible nitric oxide synthase associated pathways in acquired resistance to *Cooperia oncophora* infection in cattle. *Vet. Res.* **42**, 48 (2011).
79. Neupane, M., Kiser, J. N. & Neiberger, H. L. Gene set enrichment analysis of SNP data in dairy and beef cattle with bovine respiratory disease. *Anim. Genet.* **49**, 527–538 (2018).
80. Foley, C. *et al.* Integrated analysis of the local and systemic changes preceding the development of post-partum cytological endometritis. *BMC Genomics* **16**, 811 (2015).
81. Murakami, H., Kuroiwa, T., Suzuki, K., Miura, Y. & Sentsui, H. Analysis of syk expression in bovine lymphoma and persistent lymphocytosis induced by bovine leukemia virus. *J. Vet. Med. Sci.* **73**, 41–45 (2011).
82. Tizioto, P. C. *et al.* Immunological response to single pathogen challenge with agents of the bovine respiratory disease complex: An RNA-sequence analysis of the bronchial lymph node transcriptome. *PLoS One* **10**, e0131459 (2015).
83. Coussens, P. M. *et al.* John's disease in cattle is associated with enhanced expression of genes encoding IL-5, GATA-3, tissue inhibitors of matrix metalloproteinases 1 and 2, and factors promoting apoptosis in peripheral blood mononuclear cells. *Vet. Immunol. Immunopathol.* **105**, 221–234 (2005).
84. Verschoor, C. P. *et al.* Polymorphisms in the gene encoding bovine interleukin-10 receptor alpha are associated with *Mycobacterium avium* ssp. paratuberculosis infection status. *BMC Genet.* **11**, 23 (2010).
85. Khalifeh, M. S. & Stabel, J. R. Clinical disease upregulates expression of CD40 and CD40 ligand on peripheral blood mononuclear cells from cattle naturally infected with *Mycobacterium avium* subsp. *paratuberculosis*. *Clin. Vaccine Immunol.* **20**, 1274–1282 (2013).
86. Pant, S. D. *et al.* Bovine CLEC7A genetic variants and their association with seropositivity in John's disease ELISA. *Gene* **537**, 302–307 (2014).
87. Jin, X. *et al.* Activation of a *Neospora caninum* EGFR-like kinase facilitates intracellular parasite proliferation. *Front. Microbiol.* **8**, 1980 (2017).
88. Brym, P. & Kamiński, S. Microarray analysis of differential gene expression profiles in blood cells of naturally BLV-infected and uninfected Holstein–Friesian cows. *Mol. Biol. Rep.* **44**, 109–127 (2017).
89. Domínguez-André, J. & Netea, M. G. Impact of historic migrations and evolutionary processes on human immunity examples of evolutionary pressures providing protection in African populations. *Trends Immunol.* **40**, 1105–1119 (2019).
90. Burdick, N. C., Randel, R. D., Carroll, J. A. & Welsh, T. H. Interactions between temperament, stress, and immune function in cattle. *Int. J. Zool.* **2011**, 373197 (2011).
91. Struchalin, M. V., Dehghan, A., Witteman, J. C. M., van Duijn, C. & Aulchenko, Y. S. Variance heterogeneity analysis for detection of potentially interacting genetic loci: Method and its limitations. *BMC Genet.* **11**, 92 (2010).
92. Dumitrascu, B., Darnell, G., Ayroles, J. & Engelhardt, B. E. Statistical tests for detecting variance effects in quantitative trait studies. *Bioinformatics* **35**, 200–210 (2019).
93. Paré, G., Cook, N. R., Ridker, P. M. & Chasman, D. I. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: A report from the women's genome health study. *PLoS Genet.* **6**, e1000981 (2010).
94. Demerath, E. W. *et al.* Interaction of FTO and physical activity level on adiposity in African-American and European-American adults: The ARIC study. *Obesity* **19**, 1866–1872 (2011).
95. Hubacek, J. A., Pikhart, H., Peasey, A., Kubinova, R. & Bobak, M. FTO variant, energy intake, physical activity and basal metabolic rate in Caucasians. The HAPIEE study. *Physiol. Res.* **60**, 175–183 (2011).
96. Van Vliet-Ostapchouk, J. V., Snieder, H. & Lagou, V. Gene-lifestyle interactions in obesity. *Curr. Nutr. Rep.* **1**, 184–196 (2012).
97. Kilpelainen, T. O. *et al.* Physical activity attenuates the influence of FTO variants on obesity risk: A meta-analysis of 218,166 adults and 19,268 children. *PLoS Med.* **8**, e1001116 (2011).
98. Kong, Z. *et al.* Multi-omics analysis reveals up-regulation of APR signaling, LXR/RXR and FXR/RXR activation pathways in holstein dairy cows exposed to high-altitude hypoxia. *Animals* **9**, 406 (2019).
99. Carvalho, R. *et al.* Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. *Genet. Sel. Evol.* **51**, 29 (2019).
100. Etim, N. N. & Oguike, M. A. Environmental and management stressors: Implications for reproductive and productive performances of farm animals in the tropics. *J. Agric. Sustain.* **5**, 153–170 (2014).
101. Bova, T. L. *et al.* Environmental stressors influencing hormones and systems physiology in cattle. *Reprod. Biol. Endocrinol.* **12**, 58 (2014).
102. Rowan, T. N. *et al.* A multi-breed reference panel and additional rare variants maximize imputation accuracy in cattle. *Genet. Sel. Evol.* **51**, 77 (2019).
103. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
104. Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**, 1–9 (2020).
105. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
106. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
107. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
108. Ugarte, E., Alenda, R. & Carabaño, M. J. Fixed or random contemporary groups in genetic evaluations. *J. Dairy Sci.* **75**, 269–278 (1992).
109. PRISM Climate Group, Oregon State University, <http://prism.oregonstate.edu> (2004).
110. Leutner, B., Horning, N., Schwalb-Willmann, J. & Hijmans, R. J. Package “RStoolbox” Title Tools for Remote Sensing Data Analysis (2019).

111. R Core Team. R: A language and environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org>. (2020).
112. Henning C. Package “fpc” Title Flexible Procedures for Clustering (2019).
113. Vinkhuyzen, A. A. E. & Wray, N. R. Novel directions for G × E analysis in psychiatry. *Epidemiol. Psychiatr. Sci.* **24**, 12–19 (2015).
114. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–826 (2012).
115. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
116. Turner, S. D. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).
117. Zhu, Z. *et al.* Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* **96**, 377–385 (2015).
118. Henderson, C. R. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *J. Anim. Sci.* **60**, 111–117 (1985).
119. Wang, H. *et al.* Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci. Adv.* **5**, eaaw3538 (2019).
120. Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–273 (2012).
121. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
122. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* **8**, e1002555 (2012).
123. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
124. Bolormaa, S. *et al.* Genome-wide association studies for feedlot and growth traits in cattle. *J. Anim. Sci.* **89**, 1684–1697 (2011).
125. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **29**, 2225–2229 (2002).
126. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
127. Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
128. Raudvere, U. *et al.* g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
129. Carbon, S. *et al.* AmiGO: Online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).

## Acknowledgements

The authors would like to acknowledge the American Simmental Association for providing phenotypic and genotypic data. We also appreciate the farmers and ranchers who collected this data.

## Author contributions

C.U.B. and J.E.D. conceived the study. C.U.B. analyzed the data and interpreted the results and wrote the manuscript. T.N.R. and R.D.S. managed data and imputed variants. J.E.D. assisted with analyses, preparation of the manuscript and interpretation of results. All authors read and approved the final manuscript.

## Funding

This study was supported by Agriculture and Food Research Initiative Competitive, Grant No. 2016-68004-24827, from the USDA National Institute of Food and Agriculture.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92455-x>.

**Correspondence** and requests for materials should be addressed to J.E.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021