Powerful detection of polygenic selection and environmental adaptation in US beef cattle populations.

Troy N. Rowan, Harly J. Durbin, Sara M. Nilson, Christopher M. Seabury, Robert D. Schnabel, Jared E. Decker

Summary

Selection is a basic principle of evolution. Many approaches have identified mutations rapidly selected to high frequencies, which leave pronounced signatures on the surrounding sequence (selective sweeps). However, for complex, quantitative traits, selection does not leave these intense signatures. Instead, in polygenic selection thousands of loci experience small allele frequency changes. Directional selection and local adaptation are actively changing genomes; however identifying loci responding to this polygenic selection is difficult. We show that using novel dependent variables in linear mixed models (which accounts for confounders) identifies polygenic selection and local adaptation. In cattle, thousands of loci are responding to artificial directional selection and hundreds more show evidence of local adaptation. While advanced technologies are increasing the rate of directional selection, local adaptation is being lost. In changing climates, loss of local adaptation is especially problematic. These selection mapping approaches can be used in evolutionary, model organism, and agriculture contexts.

Main

In a changing climate, organisms are forced to adapt to environmental pressures rapidly or perish. The genomic changes that underlie adaptation are not well understood and have proven difficult to identify, aside from a handful of large-effect selective sweeps [1]. It is becoming increasingly apparent that in adaptation, hard sweeps are likely the exception, not the rule [2,3]. Polygenic selection on complex traits causes significant changes in the mean phenotype while producing only subtle changes in allele frequencies across the genome [2,4–6]. Most methods require discrete grouping of genetically similar individuals, making the identification of selection within a largely panmictic population difficult. American *Bos taurus* beef cattle have been under well-documented strong artificial and environmental selection over the last ~50 years, making them an intriguing model to study how selection acts on the genome over short periods of time and across diverse environments.

Though domesticated, beef cattle are exposed to a host of unique environments and local selection pressures compared to other more intensely managed livestock populations. This suggests that local adaptation and genotype-by-environment interactions likely play an important role in the expression of complex traits. Here, we present two methods for detecting complex polygenic selection and local adaptation using genome-wide linear mixed models (LMM) in conjunction with novel dependent variables in datasets distributed over time and space. These models identify ongoing polygenic selection on complex traits and local adaptation within populations while explicitly controlling for family structure. When applied to three US beef cattle populations, we identify numerous locally adaptive genomic regions. Further, using a meta-analysis approach, we identify loci responding to region-specific selection, largely due to the erosion of local adaptation from artificial insemination-related gene flow in cattle populations. This work provides insight into the biology of adaptation in mammalian species, and will eventually assist producers in identifying and selecting locally adapted individuals to further reduce the industry's environmental footprint.

Results

Detecting ongoing polygenic selection with Generation Proxy Selection Mapping (GPSM)

Though the first cattle genotyping assay was developed just over a decade ago [7], influential individuals from deep within the pedigree have been genotyped, providing a temporal distribution of samples at least ten generations deep for most well-established breeds. This temporal distribution of genotypes allows us to search for allelic associations with generation or some proxy for generation. Using a mixed linear model, we tested for associations between an individual's generation proxy, in our case birth year, and SNP alleles. We control for confounding population structure, relatedness, and inbreeding with a random effect that uses a genomic relationship matrix (GRM)[8]. Significant associations with birth year indicate variants that are undergoing frequency changes greater than expected due to drift over the time span of a sampled dataset. Simulations show that GPSM is able to distinguish selection from drift under a variety of genetic architectures, selection intensities, effective population sizes, and sampled time periods (**Supplementary Text**).

We used birth dates and genotypes from three large beef cattle populations; Red Angus (RAN), Simmental (SIM), and Gelbvieh (GEL) breeds to map polygenic selection (**Supplementary Table 1, Supplementary Figure 1**). When performing GPSM, we estimate that the proportion of additive genetic variance of an individual's birth year is large [Proportion Variance Explained (PVE) = 0.520, 0.588, 0.459 in RAN, SIM, and GEL, respectively], meaning that we can predict an animal's generation (or relevant proxy) with ~70% accuracy if we know its genotype[9]. Obviously, birth year is not a heritable trait, but these PVE estimates indicate significant associations between genotype and birth year across the genome. The large amount of genetic variance explained by birth year persists even when restricting the analysis to individuals born in the last 10 years (~2 generations) or 20 years (~4 generations) (**Supplementary Table 2**), demonstrating that GPSM has power to detect subtle allele frequency changes over extremely short time periods. We removed the link between generation proxy and genotype by randomly permuting the birth years assigned to each animal and observed PVE drop to effectively zero.

GPSM analyses in these three populations identified 268, 548, and 763 statistically significant SNPs (q-value < 0.1), representing at least 52, 85, and 92 genomic loci associated with year-of-birth in RAN, SIM, GEL, respectively (**Fig. 1a-f, Supplementary Table 3**). Rapid shifts in allele frequency create highly significant GPSM signals. For example, *rs1762920*, a SNP on BTA28 is undergoing massive shifts in allele frequency in all three breeds (**Fig. 1g)**, which in turn creates highly significant q-values (2.810×10^{-27}, 2.323×10^{-150}, 2.787×10^{-265} in RAN, SIM, GEL, respectively). The allele frequency changes observed for this GPSM peak are extreme compared to other significant regions, most of which are undergoing only small to moderate shifts in frequency over the last ~10 generations. When we regress allele frequency (0, 0.5, or 1.0 representing AA, AB, BB genotype per individual) on birth year, the average allele frequency change per generation (ΔAF) for GPSM hits is 0.017, 0.024, and 0.022 for RAN, SIM, and GEL respectively (**Supplementary Table 4**). In each dataset, GPSM identifies significant SNPs with $\Delta AF < 1.1 \times 10^{-4}$. The generally small allele frequency changes detected by GPSM are consistent with the magnitude of shifts expected for traits with polygenic trait architectures [6,10]. As a result, we suspect that many of these loci would go largely undetected when using other selection mapping methods.

We performed a genomic REML analysis to identify how much variation in birth year is explained by classes of GPSM SNPs. We built three GRMs with different SNP sets: One with GPSM

genome-wide significant SNPs (q < 0.1), one with the top 2,000 suggestive GPSM SNPs outside of peaks (> 1 Mb from significant SNP), and one with 2,000 non-significant SNPs (p > 0.5) intended to represent loci randomly drifting in the population. In each case, we observed that nearly all of the variation in birth year was accounted for by the significant and suggestive GRMs. Of that variation, the suggestive GRM's PVE was 2-3x greater than the significant GRM's PVE in each of our datasets. This is likely due to the significantly larger number of SNPs used. Still, these suggestive SNPs are tagging small shifts in allele frequency, congruent with directional polygenic selection, but that are too small to detect given the power of these datasets. We expect that more variants will be detectable as these commercially-generated datasets continue to increase in size.

Table 1. Genetic variation in birth year explained by three classes of GPSM significant SNPs. PVE estimates from a GREML analysis of birth year using three GRMs. GRMs were constructed with genome-wide significant SNPs (q < 0.1), the 2,000 next most significant SNPs outside of genome-wide significant peaks, and 2,000 random non-significant SNPs (p>0.5) outside of genome-wide significant peaks.

Breed	**Genome-wide significant SNP PVE (Standard Error)**	**Suggestive significant SNP PVE (Standard Error)**	**Other SNP PVE (Standard Error)**	**Total PVE (Standard Error)**
Red Angus	0.107 (0.017)	0.331 (0.014)	0.025 (0.003)	0.464 (0.015)
Simmental	0.166 (0.016)	0.339 (0.013)	0.028 (0.003)	0.533 (0.012)
Gelbvieh	0.162 (0.017)	0.297 (0.014)	0.000 (0.003)	0.465 (0.015)

GPSM identifies dozens of genomic regions under selection in each population. In Simmental, we identify significant associations at three known Mendelian loci that explain the major differences in appearance between early imported European Simmental and modern American Simmental (**Fig. 1h)**. These loci: *POLLED* (absence of horns [11]), *ERBB3/PMEL* (European Simmental cream color [12]), and *KIT* (piebald coat coloration [13]) have not experienced appreciable changes in allele frequency since 1995, but still generate significant signal in GPSM.

In addition to these three known Mendelian loci, we detect dozens of novel targets of selection both within and across breed datasets. While many GPSM signals are breed-specific, we identify seven loci under selection in all three populations, and fifteen under selection in at least two (**Supplementary Table 5**). Of regions identified in multiple breeds, many correspond to candidate genes with known production-related functions in cattle (*DACH1*-Growth QTL[14], *LRP12*-Growth QTL[15], *MYBPH*-Muscle growth[16], *RHOU*-Carcass weight[17], *BIRC5*-Feed Intake[18]). Interestingly, GPSM does not identify any known large-effect growth QTL (i.e. *PLAG1*, *LCORL*). Growth traits are known to be under strong selection in all three analyzed populations [19], but antagonistic pleiotropy created by increased calving difficulty may prevent directional selection from acting on these large-effect QTL. We also identify multiple genes under selection that have known immune functions in cattle (*ARHGAP15* [20–22], *ADORA1* [23,24], *CSF2RA* [25]). While immune function has not been explicitly selected for, it is well documented that healthy cattle perform better than their sick counterparts [26,27]. Many of the signatures identified in multiple breeds have no known functions or associations in cattle, highlighting the ability of GPSM to identify

important loci without any phenotypic records. Further, the majority of genomic regions identified by GPSM are breed-specific (79.8%, 79.8%, and 77.2% of significant regions in RAN, SIM, and GEL respectively).

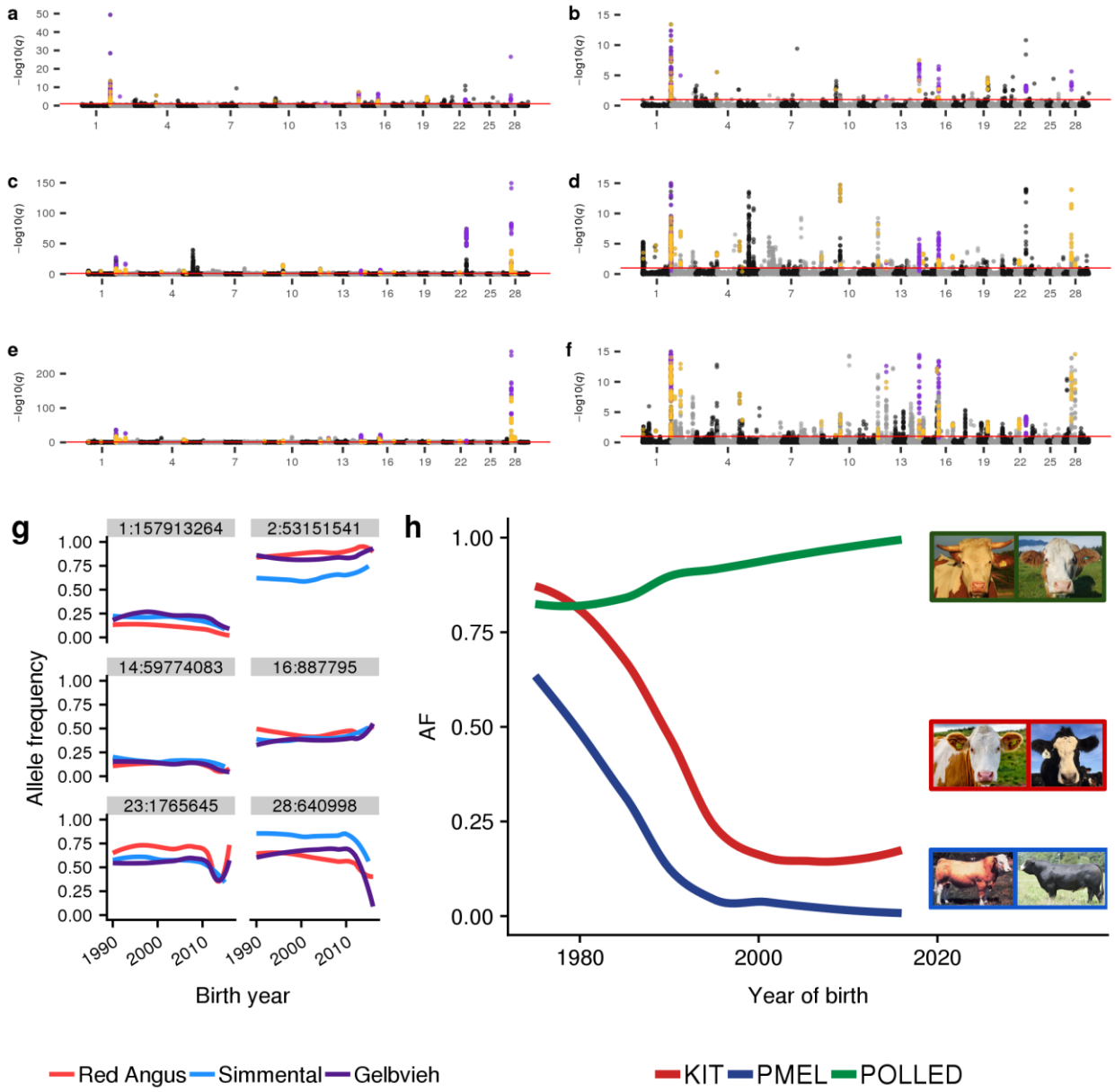

Figure 1. Generation Proxy Selection Mapping (GPSM) maps polygenic selection in three major U.S. cattle populations. Full and truncated ($-\log_{10}(p) < 15$) Manhattan plots for GPSM analysis of Red Angus (a & b), Simmental (c & d), and Gelbvieh (e & f). Purple points indicate SNPs identified in GPSM analysis of all three breeds and orange points indicate SNPs identified in at least two. (g) Smoothened allele frequency trajectories for six loci identified as being under selection in all three datasets. (h) Allele frequency trajectories for three known Mendelian loci that control differences between early European and modern American Simmental cattle.

Biological processes and pathways enriched in GPSM signals point towards selection on drivers of production efficiency and as well as known breed-specific characteristics (**Supplementary Table 6**). In each breed, we identify numerous biological processes involved in cell cycle control under selection, a group of processes directly involved in muscle growth rate in cattle [28]. Additionally in Red Angus and Gelbvieh we identify multiple cancer pathways under selection. This likely represents further evidence of selection on cell cycle regulation and growth rather than on actual cancer pathways [29]. Red Angus cattle are known to be a highly fertile breed with exceptional maternal characteristics [30]. We identify the "ovarian steroidogenesis" pathway under selection, a likely driver of hormone balance and female fertility [31,32]. We also identify numerous other processes involved in the production and metabolism of hormones. Hormone metabolism is a central regulator of growth in cattle [33–36], but could also represent selection for increased female fertility in the breed. Gelbvieh cattle are known for their quick growth rate and carcass quality. These selection criteria likely drive the six biological processes that we identify related to muscle development and function from the Gelbvieh GPSM gene set. A complete list of genomic regions under breed-specific selection and their associated candidate genes can be found in **Supplementary Table 3**.

Detecting environmental adaptation using envGWAS

Using an equivalent model to GPSM, but with continuous environmental variables (30 year normals for temperature, precipitation, and elevation) or statistically-derived discrete ecoregions as the dependent variable allows us to map environmentally adaptive genomic loci. We refer to this method as environmental GWAS (envGWAS). envGWAS extends the theory of the Bayenv approach from Coop et al. (2010) [37,38] that searches for allele frequency correlations along environmental gradients to map potentially adaptive loci, but in biobank-sized datasets. We used k-means clustering with 30-year normal values for temperature, precipitation, and elevation [39] to divide the United States into 9 discrete ecoregions (**Fig. 3a**). These largely agree with previously-published maps in the environmetrics and atmospheric science literature [40,41], and reflect well-known differences in cattle production environments. The resulting ecoregions capture not only combinations of environmental variables, but associated differences in forage type, local pathogens, and ecoregion-wide management differences to which animals are exposed. These three major breeds are not universally present in all ecoregions (**Fig. 3b, Supplementary Figures 2b & 3b, Supplementary Table 7**). This uneven regional distribution suggests that since the introduction of these breeds in the late 1960s and early 1970s, registered seedstock animals have a small footprint in desert regions with extreme temperatures and minimal rainfall. We anticipate that similar selection has occurred, though less extreme at the microenvironment scale, likely resulting in observable allele frequency differences.

Although environmental variables and regions are not inherited, the estimated Proportion Variation Explained (PVE) provides a measure of the genome-wide variation associated with environment and the degree to which cattle families are distributed across the United States. The proportion of variation explained by SNPs (PVE) ranged from 0.586 to 0.691 for temperature, 0.526 to 0.677 for precipitation, and 0.585 to 0.644 for elevation (**Supplementary Table 8**). In Red Angus, PVEs for ecoregion identity ranged from 0.463 for the Arid Prairie to 0.673 for the Fescue Belt (**Supplementary Table 9**). These measures suggest that quantifiable genetic similarity exists along both continuous environmental gradients and within discrete ecoregions. Permutation tests that shuffled individual IDs and genotypes, removing the link between environment and genotype, drove PVEs to ~0,

suggesting that observed associations were not spurious. Another permutation test that shuffled zip codes, such that animals from a given zip code were assigned the same "new" zip code provided similar results, indicating that bias due to sampling at certain zip codes was not producing envGWAS signals. In 10 rounds of permutation, we never observed p-values < $1x10^{-5}$. For all envGWAS analyses we use this as an empirically-derived p-value cutoff for significance.

Discrete ecoregion envGWAS

In Red Angus, we identified 54 variants significantly associated (significance threshold p < $1x10^{-5}$ determined from permutation) with membership of an ecoregion in the discrete multivariate envGWAS analysis, encompassing 18 total genomic regions (**Fig. 2c**). Of these regions, only two overlapped with regions identified in continuous envGWAS, suggesting that using different definitions of environment with envGWAS may detect different sources of adaptation. Of the 18 significant regions, 17 were within or near (< 100 kb) candidate genes (**Supplementary Table 10**), many of which have potentially adaptive functions. For example, envGWAS identified SNPs immediately (22.13 kb) upstream of *CUX1* (Cut Like Homeobox 1) gene on chromosome 25. *CUX1* controls hair coat phenotypes in mice [42]. Alleles within *CUX1* can be used to differentiate between breeds of goats raised for meat versus those raised for fiber [43]. The role of *CUX1* in hair coat phenotypes makes it a prime adaptive candidate in environments where animals are under heat, cold, or toxic fescue stress. Other candidate genes identified by envGWAS have been previously identified as targets of selection between breeds of cattle (*MAGI2*[44], *CENPP*[45]), or in other species (*DIRC1*-humans [46–48], *GORASP2*-fish[49], *ADRB1*-dogs[50]). Adaptive signatures shared between cattle and other species may point to shared biological processes that drive environmental adaptation. We also identify four adaptive candidate genes known to be directly involved in immune functions (*RASGEF1B*, *SPN*, *ZMYND8*, *LOC100298064*/*HAVCR1*). envGWAS identified variants within or near immune genes under ongoing selection across all three populations, suggesting that genetic adaptations conferring resistance or tolerance to local pathogens and immune stressors may be as important as adaptations to more explicit environmental stressors.

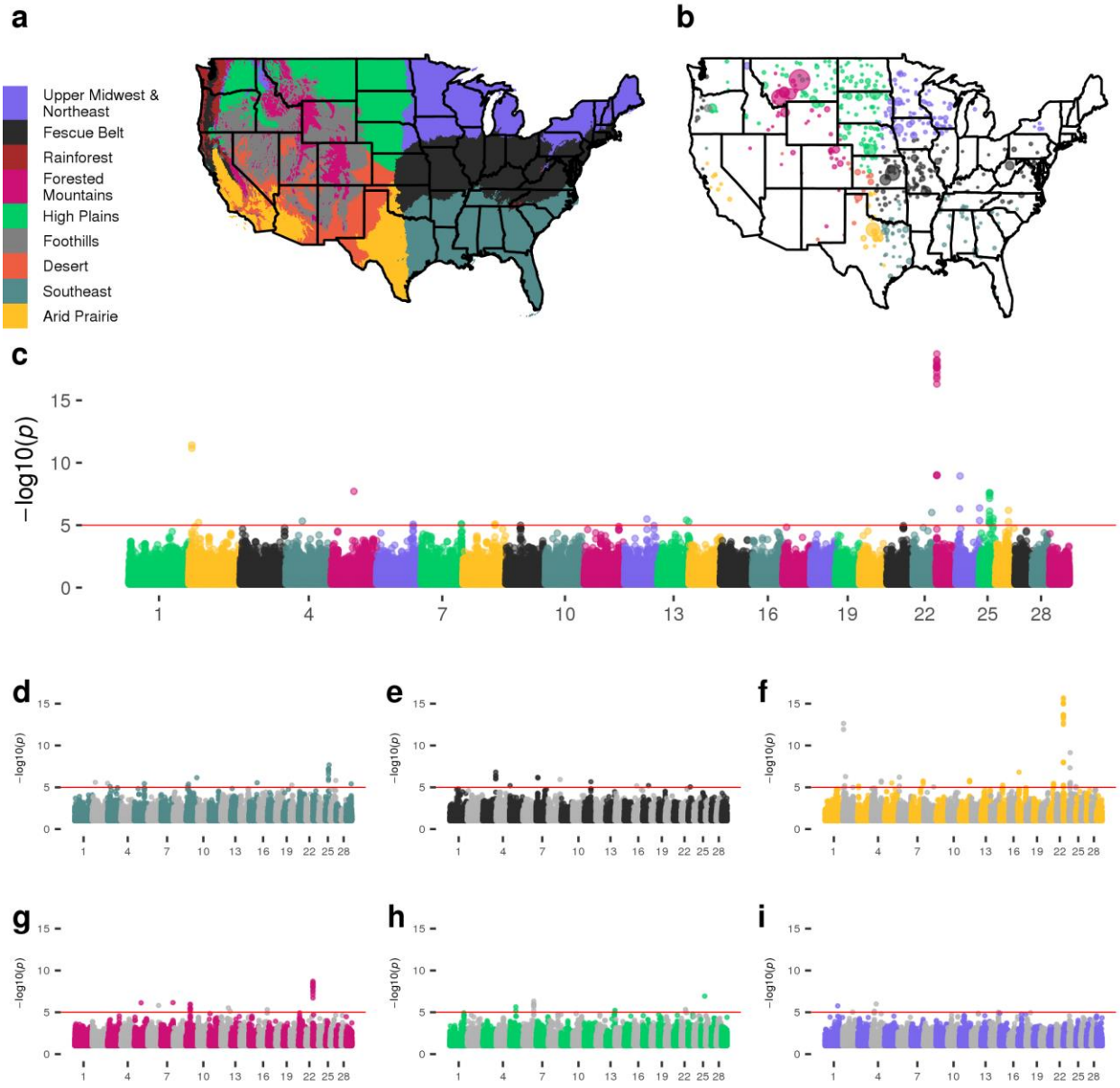

Figure 2. Manhattan plots of discrete envGWAS in Red Angus cattle. (a) Nine ecoregions of the continental United States defined by k-means clustering of 30-year normal temperature, precipitation, and elevation. (b) Locations of Red Angus animals colored by breeder's ecoregion and sized by number of animals at that location. (c) Multivariate envGWAS (case-control for regions with > 600 animals). Univariate discrete envGWAS for (d) Southeast, (e) Fescue Belt, (f) Arid Prairie, (g) Forested Mountains, (h) High Plains, and (i) Upper Midwest & Northeast ecoregions. In all Manhattan plots the red line indicates an empirically-derived p-value significance threshold from permutation testing ($p < 1x10^{-5}$).

In Simmental, we identified 11 total genomic regions tagged by 39 variants significantly associated with membership in an ecoregion multivariate envGWAS analysis (**Supplementary Figure 3**). Further, in Gelbvieh, 66 variants identified 33 local adaptation loci (**Supplementary Figure 4**). In all

three datasets, we identify a common local adaptation signature on chromosome 23 (peak SNP *rs1023574*). Multivariate analyses in all three breeds identify alleles at this SNP to be significantly associated with a region or regions ($q = 1.24x10^{-13}$, $3.15x10^{-12}$, $4.82x10^{-5}$ in RAN, SIM, and GEL, respectively). In all three datasets, we identify this same SNP as a univariate envGWAS association with membership of the Forested Mountains ecoregion. Its most significant univariate association in the Red Angus dataset was with the Arid Prairie region which is excluded from both Simmental and Gelbvieh analyses due to low within-region sample size. In the Red Angus multivariate analysis, this peak spanned 18 SNPs from (1,708,914 bp to 1,780,836 bp) and contained the pseudogene *LOC782044*. The nearest annotated gene, *KHDRBS2* (KH RNA Binding Domain Containing, Signal Transduction Associated 2) has been identified by other adaptation studies in cattle, sheep, and pigs [51–53]. Interestingly, this variant was not significantly associated with a single environmental variable in any analysis, suggesting that its adaptive benefits are unique to a particular combination of ecoregions or environmental variables.

Continuous environmental variable envGWAS

Using continuous temperature, precipitation, and elevation as quantitative phenotypes in a multivariate envGWAS analysis of Red Angus animals, we identify 46 loci with p-values below our empirically-derived cutoff ($p < 1 \times 10^{-5}$) (**Fig 3a**). These loci tag 17 genomic regions, many of which have clear candidate genes nearby (< 100 kb). Univariate envGWAS identifies 23, 17, and 10 variants associated with temperature, precipitation, and elevation, respectively (**Fig. 3b-g**). The most significant locus identified is located on chromosome 29 is located within *BBS1* (Bardet-Biedl syndrome 1), a gene directly involved in energy homeostasis in mice and humans [54,55]. *BBS1* mutant knock-in mice show irregularities in photoreceptors and olfactory sensory cilia [56] functions that are likely important to sensing an individual's local environment. Interestingly, this region is not significantly associated with a single environmental variable, nor is it identified in any discrete ecoregion envGWAS. Of the candidate genes identified in this Red Angus analysis, 9 have been implicated in adaptive functions in humans or cattle (*DIRC1*[46–48], *ABCB1*[57–59], *TBC1D1*[60–66], *AP5M1*[67], *GRIA4*[68], *LRRC4C*[69], *RBMS3*[47,70,71], *GADL1*[72–74], *ADCYAP1*[75–77], *CUX1* [42,43], *PLA2G12B*[78,79]). A complete accounting of significant SNPs and their corresponding candidate genes for all three datasets are reported in **Supplementary Table 12**.

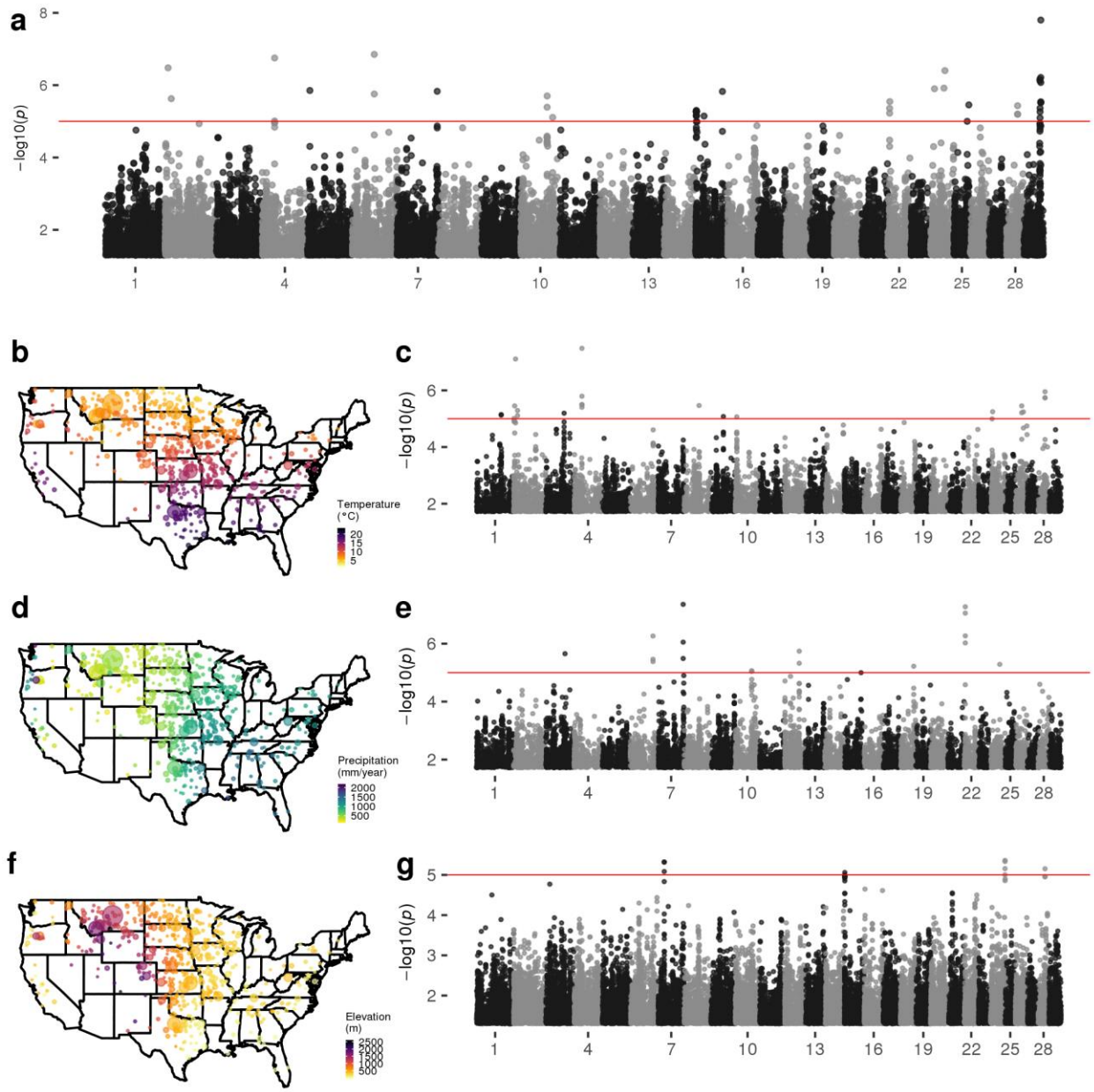

Figure 3. Continuous environmental variable envGWAS in Red Angus cattle. (a) Multivariate envGWAS of temperature, precipitation, and elevation for Red Angus cattle. Geographic distributions colored by (b) temperature, (d) precipitation, (f) elevation. Manhattan plots for univariate envGWAS analysis of (c) temperature, (e) precipitation, (g) elevation. Red lines indicate permutation-derived p-value cutoff of $1x10^{-5}$.

While minimal candidate gene overlap existed between breeds, we identified multiple shared biological pathways and processes (**Table 2**) from lists of envGWAS candidate genes. Pathways in common between breeds were driven by largely different gene sets, suggesting genomic control accessible through multiple genes. Across all three breeds, we consistently identified the "axon guidance" pathway, and numerous GO terms relating to axon development and guidance under region-specific selection. Ai et al. (2015) suggested that axon development and migration in the central nervous system is essential for the maintenance of homeostatic temperatures by modulating heat loss or production [80,81].

Further, direction and organization of axons is an essential component of the olfactory system which is frequently implicated in environmental adaptation through the recognition of local environmental cues [47,82]. In addition to axonal development, host of other neural signalling pathways are identified in multiple breeds. A genome-wide association study for gene-by-environment interactions in Simmental cattle by Braz et al. (2019) identified a similar set of enriched pathways[83]. These common neural signalling pathways identified by envGWAS are known regulators of stress response, temperature homeostasis, and vasoconstriction [84]. We identify other shared pathways involved in the control of vasodilation and vasoconstriction (relaxin signalling, renin secretion, and insulin secretion). Vasodilation and vasoconstriction are essential to physiological temperature control in cattle and other species [85–87]. The ability to mount a physiological response to temperature stress has a direct impact on cattle performance, making it a prime candidate for environment-specific selection[88]. A complete accounting of pathways and processes identified by envGWAS signals are reported in **Supplementary Table 13.**

Table 2. Shared KEGG pathways identified by envGWAS. Enrichment analyses were performed in each breed using candidate genes within 10 kb of significant envGWAS SNPs. Pathways were significant when Benjamini-Hochberg FDR-corrected p-value < 0.1.

Pathway	Num. Genes	% genes	Term p-value (FDR Corrected)	Gene Names	Dataset
Axon guidance	6	3.409	0.028	[EFNA5, EFNB2, FYN, LRRC4C, SEMA6A, UNC5C]	RAN
Axon guidance	3	1.705	0.073	[PLXNA4, SEMA3D, SEMA3E]	SIM
Axon guidance	4	2.273	0.085	[CHL1, FOXP1, NFASC, ROBO1]	
Cholinergic synapse	4	3.54	0.05	[FYN, LOC529425, PLCB3, PRKACB]	RAN
Cholinergic synapse	3	2.655	0.028	[CACNA1A, KCNQ5, PRKACA]	SIM
Cholinergic synapse	3	2.655	0.037	[GNG13, ITPR2, KCNJ12]	GEL
Dopaminergic synapse	4	2.985	0.065	[GRIA4, LOC529425, PLCB3, PRKACB]	RAN
Dopaminergic synapse	3	2.239	0.031	[CACNA1A, MAPK10, PRKACA]	SIM
GABAergic synapse	2	1.667	0.068	[CACNA1A, PRKACA]	SIM
GABAergic synapse	2	2.247	0.082	[ABAT, GNG13]	GEL
Glutamatergic synapse	4	3.54	0.05	[GRIA4, LOC529425, PLCB3, PRKACB]	RAN
Glutamatergic synapse	2	1.77	0.066	[CACNA1A, PRKACA]	SIM
Glutamatergic synapse	2	1.77	0.104	[GNG13, ITPR2]	GEL
Platelet activation	5	4.167	0.027	[FYN, LYN, PLCB3, PRKACB, PRKG2]	RAN
Platelet activation	2	1.667	0.068	[COL1A1, PRKACA]	SIM
Platelet activation	2	1.667	0.11	[ITPR2, PPP1R12A]	GEL
Relaxin signaling pathway	4	2.963	0.066	[LOC529425, PLCB3, PRKACB, VEGFB]	RAN
Relaxin signaling pathway	3	2.222	0.029	[COL1A1, MAPK10, PRKACA]	SIM
Retrograde endocannabinoid signaling	5	3.226	0.04	[GRIA4, LOC529425, NDUFA12, PLCB3, PRKACB]	RAN
Retrograde endocannabinoid signaling	3	1.935	0.033	[CACNA1A, MAPK10, PRKACA]	SIM

Identifying loci undergoing region-specific selection with GPSM ecoregion meta-analysis

envGWAS detects allelic associations with continuous and discrete environments, but does not address whether local adaptation is being positively selected for, or being eroded by the increased exchange of germplasm between ecoregions via artificial insemination. We used the spatiotemporal stratification of genotyped animals to identify genomic regions undergoing ecoregion-specific selection. We perform GPSM within each sufficiently populated ecoregion and identify variants with high p-value heterogeneity (Cocharn's Q statistic). Variants with significant Cochran's Q statistics ($p < 1x10^{-5}$) that were also significant in at least one within-region GPSM analysis ($p < 1x10^{-5}$) imply region-specific allele frequency change. These changes can be due either to selection for local adaptation, or locally different allele frequencies moving towards the population mean. We identified 59, 38, and 46 significant SNPs in Red Angus, Simmental, and Gelbvieh datasets, respectively representing 15, 21, and 26 genomic regions (> 1 Mb to nearest other significant SNP) (**Fig 4a**). In most cases, these variants have an effect (posterior probability of an effect: m-value > 0.9) in only one or two regions (**Fig 4b**). Of these variants, nearly all represent the decay of ecoregion-specific allele frequency differences towards the population average (**Fig. 4c**) as opposed to positive selection for beneficial adaptations (**Supplementary File 8**).

Despite the apparent ongoing decay of local adaptation signatures, this meta-analysis on ecoregion-specific GPSM identifies multiple intriguing adaptive candidate genes. A peak on BTA 1 at ~73.8 Mb (lead snp rs254372) lies within the *OPA1* (OPA1 mitochondrial dynamin like GTPase) gene, and is under selection in the Fescue Belt ecoregion. *OPA1* has been implicated in circadian rhythm control in mice [89,90], and is a known regulator of metabolic and cardiac adaptations [91,92] through its interactions with mitochondria. We also identify variants located within *ADAMTS16* (ADAM Metallopeptidase With Thrombospondin Type 1 Motif 16), a gene that has been identified by human adaptation studies, and that regulates blood pressure in mice [93,94]. These genes are especially interesting because the main symptoms of fescue toxicosis are due to vasoconstriction caused by ergot alkaloids synthesized by endophytes in Fescue[95]. These regions under ecoregion-specific selection are being driven towards the population average allele frequency, and in many cases fixation (**Supplementary Figures 8-10**). While these regions may be artifacts due to our sparse sampling of early animals, they may also be intriguing candidates for follow up studies on their potentially adaptive functions. Further, maintaining this ecoregion-specific variation will be essential for breeding more adapted animals.

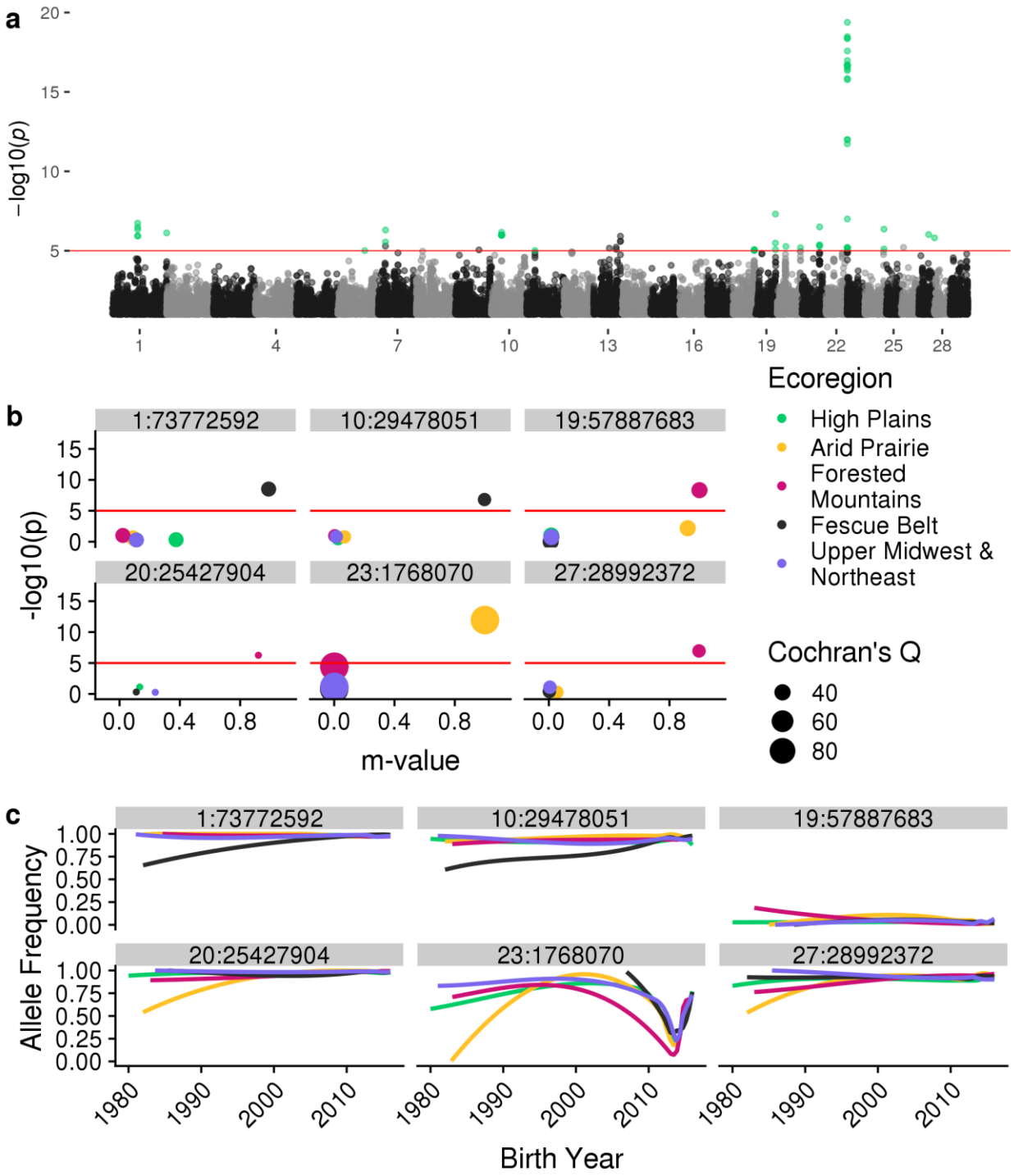

Figure 4. Meta-analysis of within-ecoregion GPSM for Red Angus cattle. A) Manhattan plot of per-variant Cochran's Q p-values. Points colored green had significant Cochran's Q ($p < 1x10^{-5}$) and were significant in at least one within-region GPSM analysis ($p < 1x10^{-5}$). B) Ecoregion effect plots for lead SNPs in six peaks from (A). Points are colored by ecoregion and sized based on Cochran's Q value. C) Region-specific allele frequency trajectories for SNPs from (B), colored by ecoregion.

DISCUSSION

We leveraged massive commercially-generated genomic datasets[96] from three major beef cattle populations to map polygenic selection and environmental adaptation with novel applications of GWAS models. Using temporally-stratified genotype data allows us to detect incredibly small shifts in allele frequency across the genome. These small ongoing shifts in allele frequency are consistent with our expectations of polygenic selection. While the selection history of cattle is well-known, it remains difficult to say which trait or traits are driving the allele frequency shifts at a particular locus. GPSM is agnostic to the traits under selection in a population, but merely points to regions changing in frequency greater than expected due to drift in the absence of population structure, relatedness, and inbreeding. This provides a more complete picture of the ongoing genomic changes underlying selection than a trait-centric approach. Further, because of our large sample sizes and unique analytical frameworks, we solve the long standing problem in population genetics of identifying polygenic selection[97].

Future studies exploring selection in the context of complex trait networks could explain how hundreds or thousands of genes act together on one or many loosely related traits when under directional selection. Candidate genes identified by GPSM point towards pathways and processes likely involved in increased production efficiency (growth, digestion, muscle development, and fertility). In addition to a handful of sites of known function previously identified in studies of cattle, we identify dozens of novel signatures of ongoing selection.

In each population our envGWAS method identifies dozens of genomic regions associated with both continuous and discrete environments. We identify numerous candidate genes with likely adaptive functions. Using these environmentally-associated candidate genes we identify an enrichment of pathways involved in neural development and signalling. These adaptations further underline the role that the nervous system plays in regulating temperature homeostasis in mammals. Other environmentally-associated pathways highlight mechanisms involved in regulating vasoconstriction and vasodilation, both of which are essential components of responses to heat, cold, and toxic fescue stressors in cattle. We believe the unique statistical power and widespread geographical distributions of these cattle datasets can be leveraged to understand the genomic basis of adaptation in a host of other species. Because different genes in different populations are identifying the same pathways, it is likely that these pathways influence local adaptation across mammalian species. This basic knowledge will be valuable as we deal with the effects of climate change.

Our results suggest that environmental associations are widespread in cattle populations, but that these populations are losing locally adaptive variants through the increased use of artificial insemination. Artificial Insemination has allowed elite germplasm to be used ubiquitously, resulting in substantial genetic improvements at the expense of eroding local adaptation. We identify dozens of genomic regions undergoing region specific selection. In almost every case this selection is driving allele frequencies in a single region back towards the population mean. In three independent datasets, we identify only a handful of common environmentally-associated genomic regions, but significant overlap in associated pathways. This suggests that adaptive networks are complex and can be altered by multiple modes of genetic adaptation. As we work to breed more environmentally-adapted cattle, there will likely be sufficient standing variation that affects these adaptive networks and pathways to make genetic progress. However, what is needed next is selection tools to harness these locally adapted variants.

Methods

Genotype Data:

SNP assays for three populations of genotyped *Bos taurus* beef cattle ranging in density from ~25,000 SNPs to ~770,000 SNPs were imputed to a common set of 830K SNPs using the large multi-breed imputation reference panel described in Rowan et al. 2019 [98]. Map positions originated from the ARS-UCD1.2 reference genome[99]. Genotypes underwent quality control in PLINK (v1.9)[100], reference-based phasing with Eagle (v2.4) [101], and imputation with Minimac3 (v2.0.1) [102]. Following imputation, all three datasets contained 836,118 autosomal variants. All downstream analyses used only variants with minor allele frequencies > 0.01. Upon filtering we performed a principal component analysis for each breed in PLINK.

Generation Proxy Selection Mapping (GPSM):

To identify alleles changing in frequency over time we fit a univariate genome-wide linear mixed model (LMM) using GEMMA (Version 0.98.1) [103]. Here, we used the model:

EQUATION 1:

$$y = Xg + Zu + e$$
$$g \sim N(0, G\sigma_a{}^2)$$
$$e \sim N(0, \sigma^2{}_e I)$$

where **y** is an individual's generation proxy, in our case birth year, and **X** was an incidence matrix that related SNPs to individuals and **g** was the estimated effect size for each marker. We control for confounding population structure, relatedness, and inbreeding with a polygenic term **u** that uses a standardized genomic relationship matrix (GRM) **G** [8]and estimates of **λ** (the ratio of variance components) and **τ** (the residual variance). Here, continuous age served as a proxy for generation number from the beginning of the pedigree. Other than tested SNP effects, no fixed effects were used in our model. Using this model, we tested each SNP for an association with continuous birth year. We converted p-values to FDR corrected q-values and used a significance threshold cutoff of q < 0.1. We performed multiple additional negative-control analyses in each dataset by shuffling the age associated with each genotype to ensure that observed GPSM signals were real. This was performed ten times for each breed dataset. To visualize the change in allele frequency of loci undergoing the strongest selection, we divided additive genotypes (coded as 0, 1, or 2) of significant variants by 2 (to reflect allele frequencies as 0, 0.5, or 1), then fit a loess and simple linear regressions for age and allele frequency in R [104]. These results and others were visualized using ggplot2 [105].

Birth year variance component analysis

To estimate the amount of variation in birth year explained by SNPs significant in GPSM, we performed multi-GRM GREML analyses for birth year in GCTA (v1.92.4) [106]. We built separate GRMs using genome-wide significant markers and all remaining makers outside of GPSM peaks (> 1Mb from significant GPSM markers). To further partition the variance in birth year explained by subsets of markers, we performed a GREML analysis using three GRMs created with genome-wide significant ($p < 1x10^{-5}$), the next 2,000 lowest p-values, and 2,000 random unassociated (p > 0.5) markers with minor allele frequencies > 0.15.

Environmental Data:

Thirty-year normals (1981-2010) for mean temperature ((average daily high (°C)+ average daily low (°C))/2), precipitation (mm/year), and elevation (m above sea level) for each 4 km square of the continental United States were extracted from the PRISM Climate Dataset [39], and used as continuous

dependent variables in envGWAS analysis. Optimal k-means clustering of these three variables grouped each 8km square of the continental US into 9 distinct ecoregions. Using reported breeder zip code for each individual, we linked continuous environmental variables to animals and partitioned them into discrete environmental cohorts for downstream analysis. For ecoregion assignments, latitude and longitude were rounded to the nearest 0.1 degrees. As a result some zip codes were assigned to multiple ecoregions. Animals from these zip codes were excluded from the discrete region envGWAS but remained in analyses that used continuous measures as dependent variables.

Environmental Genome-wide Association Studies (envGWAS):

To identify loci segregating at different frequencies within discrete ecoregions or along continuous climate gradients, we used longitudinal environmental data for the zip codes attached to our study individuals as dependent variables in univariate and multivariate genome-wide LMMs implemented in GEMMA (Version 0.98.1). We fit three univariate envGWAS models that used 30-year normal temperature, precipitation, and elevation as dependent variables. These used an identical model to **EQUATION 1**, but used environmental values as the dependent variable (**y**) instead of birth year. We also fit a combined multivariate model using all three environmental variables to account for shared genetic associations between temperature, elevation, and precipitation. To identify loci associated with entire climates as opposed to only continuous variables, we fit univariate and multivariate case-control envGWAS analyses using an individual's region assignment described in the "Environmental Data" section as binary phenotypes. Proportion variation explained (PVE), phenotypic correlations, and genetic correlations were estimated for continuous environmental variables and discrete environmental regions using GEMMA's implementation of REML.

To ensure that envGWAS signals were not driven by spurious associations, we performed two separate permuted analyses. In the first, we shuffled the environmental variables and regions associated with an individual prior to performing each envGWAS analysis, completely detaching the link between an individual's genotype and their environment. In the second, to ensure that envGWAS signals were not driven by the over-sampling of individuals at particular zip codes, we shuffled the environmental variables associated with each zip code prior to envGWAS analysis. These two types of shuffled analysis were performed in each dataset and for each type of univariate and multivariate envGWAS analysis.

GPSM meta-analyses:

To identify variants undergoing ecoregion-specific allele frequency change, we performed GPSM analyses within each region with more than 700 individuals The outputs from within-region GPSM analyses were combined into a single meta-analysis for each breed using METASOFT [107]. We identified loci with high heterogeneity in allele effect size, suggesting region-specific selection. An m-value indicating the posterior-probability of a locus having an effect in a particular ecoregion was calculated for each of these loci[108].

Pathway and biological process enrichment analysis:

Using the NCBI annotations for the ARS-UCD1.2 *Bos taurus* reference assembly, we located proximal candidate genes near significant SNPs from each of our analyses. We identified candidate gene lists for all genes located +/- 10 kb from significant envGWAS SNPs. We consolidated significant SNPs from all envGWAS analyses into a single candidate gene list for each breed. Using these candidate gene lists, we performed gene ontology (GO) enrichment analysis using ClueGO (v2.5.5) [109] implemented in Cytoscape (v3.7.2). [110]. Using a *Bos taurus* gene list, we identified GO terms where at least two members of our candidate gene list made up at least 1.5% of the term's total genes. We applied a Benjamini-

Hochberg multiple-testing correction to reported p-values. GO terms with BenjaminFDR-corrected p-values < 0.1 were considered significant.

References

1. Gutiérrez-Gil, B., Arranz, J. J. & Wiener, P. An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. *Front. Genet.* **6**, 167 (2015).
2. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–15 (2010).
3. Hernandez, R. D. *et al.* Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
4. Beissinger, T. M. *et al.* A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. *Genetics* **196**, 829–840 (2014).
5. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* **8**, (2019).
6. Höllinger, I., Pennings, P. S. & Hermisson, J. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genet.* **15**, e1008035 (2019).
7. Matukumalli, L. K. *et al.* Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* **4**, e5350 (2009).
8. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
9. Lynch, M. & Walsh, B. Genetics and analysis of quantitative traits. *invemar.org.co*.
10. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
11. Wiedemar, N. *et al.* Independent polled mutations leading to complex gene expression differences in cattle. *PLoS One* **9**, e93435 (2014).
12. Mészáros, G., Petautschnig, E., Schwarzenbacher, H. & Sölkner, J. Genomic regions influencing coat color saturation and facial markings in Fleckvieh cattle. *Anim. Genet.* **46**, 65–68 (2015).
13. Fontanesi, L., Tazzoli, M., Russo, V. & Beever, J. Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus. *Anim. Genet.* **41**, 295–303

(2010).

14. Serão, N. V. *et al.* Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. *BMC Genet.* **14**, 94 (2013).
15. 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).
16. Chelh, I., Picard, B., Hocquette, J.-F. & Cassar-Malek, I. Myostatin inactivation induces a similar muscle molecular signature in double-muscled cattle as in mice. *Animal* **5**, 278–286 (2011).
17. Espigolan, R. *et al.* Associations between single nucleotide polymorphisms and carcass traits in Nellore cattle using high-density panels. *Genet. Mol. Res.* **14**, 11133–11144 (2015).
18. Chen, Y. *et al.* Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. *Anim. Genet.* **42**, 475–490 (2011).
19. Kuehn, L. A. & Thallman, R. M. Across-Breed EPD Tables For The Year 2016 Adjusted To Breed Differences For Birth Year Of 2014. (2016).
20. Noyes, H. *et al.* Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9304–9309 (2011).
21. Smetko, A. *et al.* Trypanosomosis: potential driver of selection in African cattle. *Front. Genet.* **6**, 137 (2015).
22. Álvarez, I., Pérez-Pardal, L., Traoré, A., Fernández, I. & Goyache, F. African Cattle do not Carry Unique Mutations on the Exon 9 of the ARHGAP15 Gene. *Anim. Biotechnol.* **27**, 9–12 (2016).
23. Seo, J., Osorio, J. S. & Loor, J. J. Purinergic signaling gene network expression in bovine polymorphonuclear neutrophils during the peripartal period. *J. Dairy Sci.* **96**, 7675–7683 (2013).
24. Seo, J. *et al.* Hepatic purinergic signaling gene network expression and its relationship with inflammation and oxidative stress biomarkers in blood from peripartal dairy cattle. *J. Dairy Sci.* **97**, 861–873 (2014).
25. Meade, K. G. *et al.* Antigen stimulation of peripheral blood mononuclear cells from Mycobacterium

bovis infected cattle yields evidence for a novel gene expression program. *BMC Genomics* **9**, 447 (2008).

26. Weber, K. L. *et al.* Identification of Gene Networks for Residual Feed Intake in Angus Cattle Using Genomic Prediction and RNA-seq. *PLoS One* **11**, e0152274 (2016).
27. Alexandre, P. A. *et al.* Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC Genomics* **16**, 1073 (2015).
28. Guo, B. *et al.* Transcriptome analysis of cattle muscle identifies potential markers for skeletal muscle growth rate and major cell types. *BMC Genomics* **16**, 177 (2015).
29. Rolf, M. M. *et al.* Genome-wide association analysis for feed efficiency in Angus cattle. *Anim. Genet.* **43**, 367–374 (2012).
30. Breeds - Red Angus. *The Cattle Site* https://www.thecattlesite.com/breeds/beef/99/red-angus/.
31. Spicer, L. J. & Echternkamp, S. E. Ovarian follicular growth, function and turnover in cattle: a review. *J. Anim. Sci.* **62**, 428–451 (1986).
32. Gareis, N. C. *et al.* Impaired insulin signaling pathways affect ovarian steroidogenesis in cows with COD. *Anim. Reprod. Sci.* **192**, 298–312 (2018).
33. Davis, S. L., Hossner, K. L. & Ohlson, D. L. Endocrine Regulation of Growth in Ruminants. in *Manipulation of Growth in Farm Animals: A Seminar in the CEC Programme of Coordination of Research on Beef Production, held in Brussels December 13–14, 1982* (eds. Roche, J. F. & O'Callaghan, D.) 151–178 (Springer Netherlands, 1984).
34. Brockman, R. P. & Laarveld, B. Hormonal regulation of metabolism in ruminants; a review. *Livestock Production Science* **14**, 313–334 (1986).
35. Hocquette, J. F., Ortigues-Marty, I., Pethick, D., Herpin, P. & Fernandez, X. Nutritional and hormonal regulation of energy metabolism in skeletal muscles of meat-producing animals. *Livestock Production Science* **56**, 115–143 (1998).
36. Hocquette, J. F. Endocrine and metabolic regulation of muscle growth and body composition in cattle. *Animal* **4**, 1797–1809 (2010).

37. Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J. K. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411–1423 (2010).

38. Günther, T. & Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205–220 (2013).

39. PRISM Climate Group. PRISM 30-year Normal Climate Data.

40. Fovell, R. G. & Fovell, M.-Y. C. Climate Zones of the Conterminous United States Defined Using Cluster Analysis. *J. Clim.* **6**, 2103–2135 (1993).

41. Sathiaraj, D., Huang, X. & Chen, J. Predicting climate types for the Continental United States using unsupervised clustering techniques. *Environmetrics* vol. 30 e2524 (2019).

42. Sansregret, L. & Nepveu, A. The multiple roles of CUX1: insights from mouse models and cell-based assays. *Gene* **412**, 84–94 (2008).

43. Bertolini, F. *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* **50**, 57 (2018).

44. Boitard, S., Boussaha, M., Capitan, A., Rocha, D. & Servin, B. Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics* **203**, 433–450 (2016).

45. Wang, M. D., Dzama, K., Hefer, C. A. & Muchadeyi, F. C. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC Genomics* **16**, 894 (2015).

46. Hussin, J., Nadeau, P., Lefebvre, J.-F. & Labuda, D. Haplotype allelic classes for detecting ongoing positive selection. *BMC Bioinformatics* **11**, 65 (2010).

47. Williamson, S. H. *et al.* Localizing recent adaptive evolution in the human genome. *PLoS Genet.* **3**, e90 (2007).

48. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).

49. Guo, B., DeFaveri, J., Sotelo, G., Nair, A. & Merilä, J. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol.* **13**, 19 (2015).

50. Kim, J. *et al.* Genetic selection of athletic success in sport-hunting dogs. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7212–E7221 (2018).

51. León, C. D., De León, C., Manrique, C., Martínez, R. & Rocha, J. F. Research Article Genomic association study for adaptability traits in four Colombian cattle breeds. *Genetics and Molecular Research* vol. 18 (2019).

52. Guo, J. *et al.* Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Sci. Rep.* **8**, 10405 (2018).

53. Gurgul, A. *et al.* A genome-wide detection of selection signatures in conserved and commercial pig breeds maintained in Poland. *BMC Genet.* **19**, 95 (2018).

54. Guo, D.-F. *et al.* The BBSome Controls Energy Homeostasis by Mediating the Transport of the Leptin Receptor to the Plasma Membrane. *PLoS Genet.* **12**, e1005890 (2016).

55. Rouabhi, M., Guo, D. F. & Rahmouni, K. Bardet-Biedl Syndrome 1 Gene in the Ventromedial Hypothalamus is Required for Energy Homeostasis. *The FASEB Journal* **30**, 750.4–750.4 (2016).

56. Davis, R. E. *et al.* A knockin mouse model of the Bardet–Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19422–19427 (2007).

57. Wang, H. *et al.* Comparative and evolutionary pharmacogenetics of ABCB1: complex signatures of positive selection on coding and regulatory regions. *Pharmacogenet. Genomics* **17**, 667–678 (2007).

58. Wang, Z. *et al.* Signatures of recent positive selection at the ATP-binding cassette drug transporter superfamily gene loci. *Hum. Mol. Genet.* **16**, 1367–1380 (2007).

59. López Herráez, D. *et al.* Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS One* **4**, e7888 (2009).

60. Zhao, F., McParland, S., Kearney, F., Du, L. & Berry, D. P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.* **47**, 49 (2015).

61. Cardoso, D. F. *et al.* Genome-wide scan reveals population stratification and footprints of recent selection in Nelore cattle. *Genet. Sel. Evol.* **50**, 22 (2018).

62. Rubin, C.-J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591 (2010).

63. Stone, S. *et al.* TBC1D1 is a candidate for a severe obesity gene and evidence for a gene/gene interaction in obesity predisposition. *Hum. Mol. Genet.* **15**, 2709–2720 (2006).

64. Chadt, A. *et al.* Tbc1d1 mutation in lean mouse strain confers leanness and protects from diet-induced obesity. *Nat. Genet.* **40**, 1354–1359 (2008).

65. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19 (2016).

66. Northrup, J. M., Shafer, A. B. A., Anderson, C. R., Coltman, D. W. & Wittemyer, G. Fine-scale genetic correlates to condition and migration in a wild cervid. *Evol. Appl.* **7**, 937–948 (2014).

67. Duforet-Frebourg, N., Bazin, E. & Blum, M. G. B. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Mol. Biol. Evol.* **31**, 2483–2495 (2014).

68. Igoshin, A. V. *et al.* Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. *BMC Genet.* **20**, 26 (2019).

69. Wang, B. *et al.* On the origin of Tibetans and their genetic basis in adapting high-altitude environments. *PLoS One* **6**, e17002 (2011).

70. Amorim, C. E. G., Daub, J. T., Salzano, F. M., Foll, M. & Excoffier, L. Detection of convergent genome-wide signals of adaptation to tropical forests in humans. *PLoS One* **10**, e0121557 (2015).

71. Granka, J. M. *et al.* Limited evidence for classic selective sweeps in African populations. *Genetics* **192**, 1049–1064 (2012).

72. Flori, L. *et al.* A genomic map of climate adaptation in Mediterranean cattle breeds. *Mol. Ecol.* **28**, 1009–1029 (2019).

73. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).

74. Key, F. M., Fu, Q., Romagné, F., Lachmann, M. & Andrés, A. M. Human adaptation and population

differentiation in the light of ancient genomes. *Nat. Commun.* **7**, 10775 (2016).

75. Mueller, J. C., Pulido, F. & Kempenaers, B. Identification of a gene associated with avian migratory behaviour. *Proc. Biol. Sci.* **278**, 2848–2856 (2011).
76. Bazzi, G. *et al.* Adcyap1 polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson's warbler (Cardellina pusilla). *Ecol. Evol.* **6**, 3226–3239 (2016).
77. Jones, S. E. *et al.* Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* **10**, 343 (2019).
78. Sjöstrand, A. E., Sjödin, P. & Jakobsson, M. Private haplotypes can reveal local adaptation. *BMC Genet.* **15**, 61 (2014).
79. Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).
80. Ai, H. *et al.* Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217–225 (2015).
81. Boulant, J. A. & Dean, J. B. Temperature receptors in the central nervous system. *Annu. Rev. Physiol.* **48**, 639–654 (1986).
82. Mombaerts, P. Axonal wiring in the mouse olfactory system. *Annu. Rev. Cell Dev. Biol.* **22**, 713–737 (2006).
83. Braz, C. U., Rowan, T. N., Schnabel, R. D. & Decker, J. E. Extensive genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. doi:10.1101/2020.01.09.900902.
84. Morrison, S. F. Central control of body temperature. *F1000Res.* **5**, (2016).
85. Daanen, H. A. M. & Van Marken Lichtenbelt, W. D. Human whole body cold adaptation. *Temperature (Austin)* **3**, 104–118 (2016).
86. Choshniak, I., McEwan-Jenkinson, D., Blatchford, D. R. & Peaker, M. Blood flow and catecholamine concentration in bovine and caprine skin during thermal sweating. *Comp. Biochem. Physiol. C* **71C**, 37–42 (1982).

87. Garner, J. B. *et al.* Genomic Selection Improves Heat Tolerance in Dairy Cattle. *Sci. Rep.* **6**, 34114 (2016).

88. Rhoads, M. L. *et al.* Effects of heat stress and plane of nutrition on lactating Holstein cows: I. Production, metabolism, and aspects of circulating somatotropin. *J. Dairy Sci.* **92**, 1986–1997 (2009).

89. Perganta, G. *et al.* Non-image-forming light driven functions are preserved in a mouse model of autosomal dominant optic atrophy. *PLoS One* **8**, e56350 (2013).

90. Davies, V. J. *et al.* Opa1 deficiency in a mouse model of autosomal dominant optic atrophy impairs mitochondrial morphology, optic nerve structure and visual function. *Hum. Mol. Genet.* **16**, 1307–1318 (2007).

91. Patten, D. A. *et al.* OPA1-dependent cristae modulation is essential for cellular adaptation to metabolic demand. *EMBO J.* **33**, 2676–2691 (2014).

92. Piquereau, J. *et al.* Down-regulation of OPA1 alters mouse mitochondrial morphology, PTP function, and cardiac adaptation to pressure overload. *Cardiovasc. Res.* **94**, 408–417 (2012).

93. Dong, K. *et al.* Genomic scan reveals loci under altitude adaptation in Tibetan and Dahe pigs. *PLoS One* **9**, e110520 (2014).

94. Gopalakrishnan, K. *et al.* Targeted disruption of Adamts16 gene in a rat genetic model of hypertension. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20555–20559 (2012).

95. Strickland, J. R., Aiken, G. E., Spiers, D. E., Fletcher, L. R. & Oliver, J. W. Physiological Basis of Fescue Toxicosis. in *Tall Fescue for the Twenty-first Century* (eds. Fribourg, H. A., Hannaway, D. B. & West, C. P.) 203–227 (American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 2009).

96. Decker, J. E. Agricultural Genomics: Commercial Applications Bring Increased Basic Research Power. *PLoS Genet.* **11**, e1005621 (2015).

97. Johnsson, M. Integrating Selection Mapping With Genetic Mapping and Functional Genomics. *Front. Genet.* **9**, 603 (2018).

98. 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).

99. Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**, (2020).

100. 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).

101. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443 (2016).

102. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

103. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).

104. R Core Team, R. & Others. R: A language and environment for statistical computing. (2013).

105. Wickham, H. ggplot2. *Wiley Interdisciplinary Reviews: Computational* (2011).

106. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

107. 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).

108. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* **8**, e1002555 (2012).

109. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).

110. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).