scientific reports

OPEN

New loci and neuronal pathways for resilience to heat stress in cattle

Evans K. Cheruiyot[1,2], Mekonnen Haile-Mariam[2✉], Benjamin G. Cocks[1,2], Iona M. MacLeod[2], Ruidong Xiang[2,3] & Jennie E. Pryce[1,2]

While understanding the genetic basis of heat tolerance is crucial in the context of global warming's effect on humans, livestock, and wildlife, the specific genetic variants and biological features that confer thermotolerance in animals are still not well characterized. We used dairy cows as a model to study heat tolerance because they are lactating, and therefore often prone to thermal stress. The data comprised almost 0.5 million milk records (milk, fat, and proteins) of 29,107 Australian Holsteins, each having around 15 million imputed sequence variants. Dairy animals often reduce their milk production when temperature and humidity rise; thus, the phenotypes used to measure an individual's heat tolerance were defined as the rate of milk production decline (slope traits) with a rising temperature–humidity index. With these slope traits, we performed a genome-wide association study (GWAS) using different approaches, including conditional analyses, to correct for the relationship between heat tolerance and level of milk production. The results revealed multiple novel loci for heat tolerance, including 61 potential functional variants at sites highly conserved across 100 vertebrate species. Moreover, it was interesting that specific candidate variants and genes are related to the neuronal system (*ITPR1*, *ITPR2*, and *GRIA4*) and neuroactive ligand–receptor interaction functions for heat tolerance (*NPFFR2*, *CALCR*, and *GHR*), providing a novel insight that can help to develop genetic and management approaches to combat heat stress.

Heat stress from rising global temperatures is an issue of growing importance across tropical and temperate zones affecting humans, livestock, wildlife, and plants. A recent study[1] indicates that many people are now exposed to harmful heat, and this has risen by more than twofold when compared to the pre-industrial climates (i.e., 95 vs. 275 million people), with future projections showing that over 1 billion people will experience an even greater impact of heat within the next 50 years[2]. In livestock, the annual temperature–humidity values that rise above thresholds considered to be comfortable have been increasing in many regions including Australia, the USA, Canada, and parts of Europe[3,4], making heat stress a multimillion-dollar issue in the livestock industry that compromises production (reduced growth, milk, eggs, etc.) and reproduction leading to economic losses[5].

The thermoregulatory capacities of mammals and plants to cope with extreme heat have been studied for decades. Genetic variation of thermoregulation during heat stress exists within species, including cattle breeds, with the literature indicating that tropical breeds, such as Zebu (*Bos indicus*), have a better tolerance to temperature and humidity than cattle from temperate zones (e.g., Holsteins), in part, due to the lower productivity of Zebu cattle[6]. Temperate breeds also show genetic variation in heat tolerance; for example, New Zealand Holsteins appear to exhibit higher reductions in milk yield in hotter climates than Jerseys or crossbreds[7]. While it is not fully understood why animals differ in their thermotolerance, it is hypothesised to be due to a myriad of biological mechanisms; including cellular, morphological (coat color, coat length, etc.), behavioural (e.g., feed and water intake, standing and lying time), as well as neuro-endocrine systems. See comprehensive review by[8] for more information. Notably, the molecular basis for differences in these adaptive responses within various mammalian species is still largely unknown.

Dairy cattle particularly Holsteins are excellent and convenient model for enhancing our knowledge on the molecular aspects of heat tolerance in mammals for two main reasons: (1) large phenotype datasets needed to study heat tolerance, as well as extensive genomic information, are available; (2) they have been genetically selected mainly for high milk production over many years, offering an opportunity to understand the genetic basis for coping with both environmental and elevated metabolic-heat stress associated with increased milk production.

[1]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia. [2]Agriculture Victoria Research, Centre for AgriBiosciences, AgriBio, Bundoora, VIC 3083, Australia. [3]Faculty of Veterinary and Agricultural Science, The University of Melbourne, Parkville, VIC 3052, Australia. ✉email: Mekonnen.HaileMariam@agriculture.vic.gov.au

Trait	Milk intercept traits[a]		Heat tolerance milk slope traits[b]	
	AG ± SE	h² ± SE	AG ± SE	h² ± SE
Milk yield (liters)	381.0 (14.36)	0.36 (0.01)	4.80 (0.20)	0.23 (0.01)
Fat yield (kg)	38.78 (1.64)	0.30 (0.01)	0.56 (0.02)	0.21 (0.01)
Protein yield (kg)	18.60 (0.89)	0.24 (0.01)	0.47 (0.02)	0.20 (0.01)

Table 1. Additive genetic variance (AG) and genomic heritability (h^2) for milk intercept and heat tolerance slope traits estimated for 29,107 cows based on 50 k SNP panel. *SE* Standard errors. [a]Represents the level of milk production of cows. [b]Heat tolerance proxy-phenotypes.

The development of methods to describe heat tolerance in cattle has been an active research area for many years. Measuring changes in core body temperature (e.g., rectal, vaginal, rumen temperature, etc.), thermal indices (e.g., temperature–humidity index (THI)) are some of the ways to assess thermal adaptations and performance in animals. Ravagnolo et al.[9] pioneered using daily milk yield and temperature–humidity data to measure variability in the rate of decline in milk yield associated with variability in response to heat stress. This method has been widely adopted due to the availability of large datasets from routine recording in dairy farms, e.g.,[3]. Heat tolerance in dairy cattle measured using rectal temperatures or the rate of milk yield decline is partly under genetic control, having a low (0.1) to moderate heritability (0.30)[3,9,10], which makes it amenable to selection. As such, considerable research has been undertaken to provide breeding solutions for heat stress, which is already a feature of dairy cattle breeding programmes in some parts of the world, e.g., Australia[3]. Identifying specific genetic variants that increase tolerance to heat may help to improve dairy breeding programmes in addition to improving our knowledge of the thermal biology in other mammals. However, except for mutations in the SLICK locus[11], the identification of the specific genetic variants for heat tolerance in cattle and other species has, in most cases, remained elusive, in part due to many reasons, including the sample size used in past studies[12–15].

Having a large sample size is particularly important for identifying rare causal variants with medium-sized effects and common variants with small effects. As sample size increases, the loci significantly associated with complex traits are expected to increase, as demonstrated for the human height[16]. Several selection signature work e.g.,[17,18] and genome-wide association studies (GWAS) using Single nucleotide polymorphisms (SNPs) have been conducted over the last decade to identify candidate causal genes for various heat tolerance traits (rectal temperature, heart rate, sweating rate, rate of milk yield decline, etc.) in dairy cattle[12–15] and pigs[19]. However, these GWAS were underpowered, with the largest sample size to date of around 5000 animals[12,13]. These studies have also used standard industry SNP panels of random genome-wide markers, either 50 k or 600 k SNPs, leading to inconsistencies and poor replication of the results. Although these studies have identified multiple significant variants associated with heat stress in animals, none were established to be causal mutations.

Here, we performed a GWAS using milk production records of 29,107 Holstein cows, each having over 15 million sequence variants that were imputed from various lower density SNP chips to whole-genome sequence using a reference dataset of sequences from the Run7 of 1000 Bull Genome Project[20]. The specific aims of the current study were to: (1) perform single-trait GWAS to identify genomic variants associated with sensitivity of milk traits (milk, protein, and fat) to heat stress; (2) combine single-trait GWAS results in a multi-trait meta-analysis to boost the power and identify pleiotropic variants associated with all the milk traits; and (3) conduct post-GWAS pathway analysis using the list of candidate genes identified in single-trait GWAS and meta-analysis to elucidate biological mechanisms underlying heat tolerance.

Results

Descriptive statistics and genomic heritability of the study phenotypes.
The average yield and their corresponding standard deviation (in brackets) of milk (in liters), fat (kg), and proteins (kg) used in our study was 25.85 (8.19), 0.98 (0.30), and 0.85 (0.26), respectively. The heat tolerance proxy-phenotypes (i.e., slope traits) and intercepts (representing level of milk production) that were derived from the milk traits are in Table 1. The slope traits derived from the milk, fat, and protein yield using reaction norm models on a function of the temperature–humidity index (THI) were defined as follows: heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slope traits, respectively. On the other hand, the intercept solutions from the reaction norm models – representing the level of milk production were defined as milk (MYint), fat (FYint), and protein (PYint) yield intercept traits. The values for slopes (no units) for HTMYslope, HTFYslope, and HTPYslope ranged between [− 36.80 to 27.17], [− 11.39 to 9.0] and [− 8.91 to 9.31], with values at 25% and 75% quartiles of [− 0.98 and 0.90], [− 2.76 and 2.66], and [− 1.03 and 0.95], respectively. Note that the values for milk, fat, and protein yield have been scaled by a factor of 10, 100 and 100, respectively (see "Methods"). The genomic heritability estimates for the intercept traits were high [0.36 ± 0.01 (MYint), 0.30 ± 0.01 (FYint), 0.24 ± 0.01 (PYint)] compared to slope traits [0.23 ± 0.01 (HTMYslope), 0.21 ± 0.01 (HTFYslope), 0.20 ± 0.01 (HTPYslope)] (Table 1). The phenotypic correlations between the intercept and slope traits were high, with values of -0.71 (MYint versus HTMYslope), -0.77 (FYint versus HTFYslope), and -0.83 (PYint versus HTPYslope), suggesting that lower producing cows have a smaller reduction in their yield as the THI increases. The Pearson correlations of slope solutions from the reaction norm model were 0.90 (HTMYslope versus HTPYslope), 0.56 (HTMYslope versus HTFYslope) and 0.62 (HTPYslope versus HTFYslope).

	Single-trait GWAS for intercept traits			Single-trait GWAS for slope traits			Multi-trait meta-analysis[a]	
	MYint	FYint	PYint	HTMYslope	HTFYslope	HTPYslope	Meta intercept	Meta slope
Significant ($p < 1E–05$)	9344 [72]	7844 [28]	4195 [49]	6061 [37]	8684 [21]	2998 [30]	51,568 [100]	40,220 [65]
FDR < 0.10	16,469 [188] ($p = 1E–04$)	11,469 [98] ($p = 7E–05$)	5285 [78] ($p = 3E–05$)	9172 [109] ($p = 6E–05$)	12,619 [98] ($p = 8E–05$)	3310 [51] ($p = 2E–05$)	108,934 [347] ($p = 7E–04$)	77,499 [293] ($p = 5E–04$)

Table 2. Number of SNPs identified at a p value of < 1E–05 (significant), and false discovery rate (FDR < 0.10) for QTL discovery cows (N = 29,107) based on 15 million imputed-whole genome sequence variants. MYint, Milk yield intercept; FYint, Fat yield intercept; PYint, protein yield intercept; HTMYslope, heat tolerance milk yield slope; HTFYslope, heat tolerance fat yield slope; HTPYslope, heat tolerance protein yield slope. [a]Multi-trait meta-analysis of single-trait GWAS was performed for intercept and slope traits following[21]. Values in square brackets are the number of lead SNPs defined as the top significant SNP within 5 Mb non-overlapping windows across the chromosome. The FDR was calculated following the method described by[21], where the p value in the brackets represents the cut-off threshold equivalent to FDR < 0.10 for each trait.

Single-trait GWAS for intercept and slope traits. The number of significant SNPs was generally higher for intercept than slope traits at the p value thresholds tested (Table 2). At a stringent p value of < 1E–05, the false discovery rate (FDR) varied between 0.02 and 0.03 for intercept and 0.02 and 0.05 for slope traits. The number of significant independent QTL (based on the number of 5 Mb non-overlapping windows across the chromosome with at least one significant SNP) ranged from 28 to 72 for intercept traits and from 21 to 37 for slope traits. At a relaxed cut-off threshold, where the FDR was < 0.10, the number of significant QTLs from single-trait GWAS ranged from 78 to 188 (intercept traits) and from 51 to 109 (slope traits).

The number of significant ($p < 1E–05$) QTLs (i.e., 5 Mb windows) varied across the three slope traits with greater overlap between HTMYslope and HTPYslope (13 QTLs; 20.6%) compared to HTMYslope and HTFYslope (3 QTLs; 4.8%) (Supplementary Fig. S3). The overlaps were based on whether the lead SNPs (most significant) within QTLs between traits were close (within 1 Mb). Surprisingly, none of the candidate QTLs overlapped between HTFYslope and HTPYslope. The effects of the lead SNPs within QTLs that overlapped between HTMYslope and HTPYslope were generally in the same direction.

Multi-trait meta-analysis of GWAS to detect variants with pleiotropic effects. Meta-analysis of GWAS results could increase the power of detecting informative variants[21, 22]. Compared to single-trait GWAS, the number of significant independent QTLs (based on 5 Mb windows with at least one significant SNP) was much higher for a multi-trait meta-analysis (Fig. 1; Table 2). At FDR < 0.10, the number of significant independent QTLs from multi-trait meta-analysis was 347 and 293 for intercept and slope traits, respectively (Table 2). At $p < 1E–05$, the number of significant QTLs was 100 (meta-analysis of intercept traits) and 65 (meta-analysis of slope traits). Of the significant QTLs ($p < 1E–05$; N = 65) for meta-analysis of slope traits, 35% (N = 23) overlapped with the candidate QTLs for single-trait GWAS analysis based on whether the lead SNP (most significant) within overlapping QTLs were close (within 1 Mb).

Lead SNPs detected using single-trait GWAS and meta-analysis of slope traits. The lead SNPs were defined as the most significant SNPs within an independent QTL (i.e., the most significant SNP chosen within 5 Mb windows across the chromosome). Detailed annotation of all the lead SNPs for single slope traits and the meta-analysis (N = 118) detected at the most stringent p value cut-off ($p < 1E–05$) are in the Supplementary Table S2.

About half the lead SNPs (51%) for slopes were in relatively low LD ($r^2 < 0.5$) with nearby (within 1 Mb region) lead SNPs for intercepts, indicating that they are not strongly associated with the level of milk production. Some lead SNPs mapped within or close to several candidate genes, which have been linked to environmental stress or heat tolerance in animals in previous studies, including *REG3A*[23], *NPFFR2*[24], and *CLSTN2*[25]. Several other lead SNPs mapped close to novel candidate genes that, to our knowledge, have not been described for thermotolerance in previous studies.

However, the remaining lead SNPs (49%) for slopes were in medium to strong LD ($r^2 > 0.50$) with nearby (within 1 Mb) lead SNPs identified for intercept traits (Supplementary Fig. S4), suggesting that they affect both traits, which was expected due to the strong genetic negative correlation between heat tolerance and milk production, with estimates in this study of around -0.80. The most significant lead SNPs for heat tolerance (slope traits) that were strongly (LD; $r^2 > 0.8$) associated with the level of milk production (intercept traits) mapped close to or are within genomic loci previously reported to have pleiotropic effects on bovine milk production traits, including the *DGAT1*[26,27], *MGST1*[28], and *GHR* gene[29].

Conditional GWAS for slope traits on either the lead SNPs or the intercept traits. We performed two conditional GWAS for slope traits to confirm whether the top hits (lead SNPs) detected in the first-round of GWAS for the slope traits were in fact discoveries of heat tolerance rather than indicators of milk yield (as the intercept and slope traits are genetically correlated). Of interest was the conditional GWAS analysis on chromosome 14, since the highly significant QTL around 0.5 Mb harbours the DGAT1 gene and the HSF1 (heat shock factor 1) gene, for which the latter has been linked to thermotolerance in Holstein cattle in different coun-
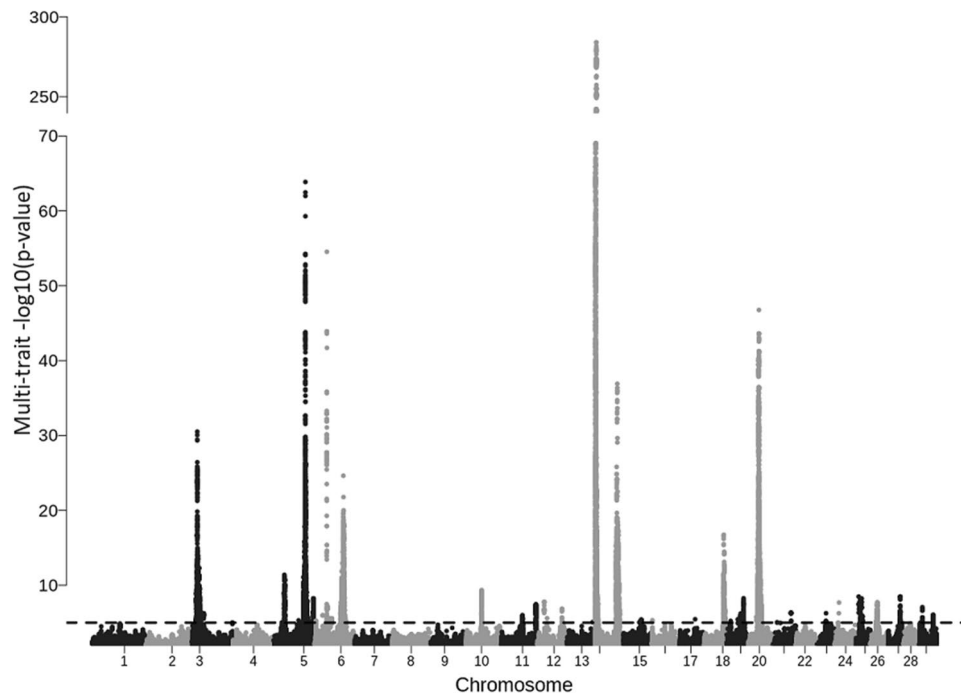
Figure 1. Manhattan plot of *p* values obtained from combining single-trait GWAS results for milk yield slope traits.

tries, including Australia[13], and the USA[14]. Notably, the lead SNPs from the first-round of GWAS for HTMYs-lope and HTFYslope (Chr14:581,569) and HTPYslope (Chr14:555,701) traits were upstream to *SLC52A2* and a synonymous variant in the *CPSF1* genes, respectively.

Figure 2 shows conditional GWAS results for chromosome 14 (around the region which showed the strongest signal in the first-round of GWAS for the slope traits—here, the conditional analyses were for slope traits on either the lead SNP or the intercept trait. In both approaches, we found that most of the SNPs were no longer significant after conditional analysis. This was the case for HTMYslope and HTPYslope slope traits, suggesting that these SNPs were possibly tagging the lead SNPs for slope traits. The lead SNP was in strong LD ($r^2 > 0.8$) with several other variants around this QTL spanning over 10 genes (Fig. 2), including variants in the HSF1 (heat shock factor 1) gene, which implies that any variant (s) around this region are possible causal mutations for heat tolerance. Nonetheless, the complex LD within this QTL region makes it difficult to pinpoint a putative causal variant (s) for heat tolerance.

Notably, even after fitting the lead SNP in a conditional GWAS analysis, there were still other somewhat significant ($p < 1E–05$; Fig. 2) SNPs remaining for the HTFYslope trait (though not very strong signals; Fig. 2), suggesting that they could be other QTLs for heat tolerance, which were not captured by the lead SNPs identified in the study.

Although the two conditional GWAS strategies (i.e., conditioning slopes on either lead SNP or intercept traits) were generally comparable regarding the strength of the GWAS signals (Fig. 2), we observed a significant (Student's *t* test; $p < 0.001$) difference in the distribution of the GWAS *p* values across slope traits. This is, in part, due to the difference in the two conditional GWAS approaches regarding the covariate fitted in the linear model. We also observed similar findings for the conditional GWAS analysis on chromosome 20 (Supplementary Fig. S5).

By conducting a conditional analysis of slope traits on the intercepts, we detected multiple additional QTL signals (lead SNPs) across the genome at $p < 1E–05$ (Supplementary Fig. S6). However, most of these lead SNPs were associated with a large FDR > 0.10—FDR for each SNP computed following Storey and Tibshirani[30]. Of the few candidate variants (all of which were detected from HTFYslope traits) with FDR < 0.10, the strongest GWAS signal was in BTA 14–1.7 Mb, of which the lead SNP (Chr14:1,726,184) mapped to the downstream region of JRK (Jrk helix-turn-helix protein). Notably, this gene was found to regulate behavioural rhythms in *Drosophila* flies, which is crucial for adaptive response to environmental changes such as temperature variations[31].

When combining conditional GWAS results for slope traits (conditioning on the intercept traits) in the meta-analysis approach, we detected 40 lead SNPs ($p < 1E–05$), all of which associated with low FDR < 0.10 (Supplementary Fig. S7 and Table S3). The mean LD between these 40 lead SNPs and the lead SNPs detected for intercept traits was very low ($r^2 < 0.20$), confirming that the conditional analysis was successful in identifying additional candidate variants for heat tolerance (besides the QTL detected from the first-round of GWAS) that are not strongly associated with the level of milk production. The most significant lead SNP (Chr14:531,267; $p = 9.04E–12$) mapped to the upstream region of the SLC39A4 gene, a member of the solute carrier family, required for intestinal zinc uptake.
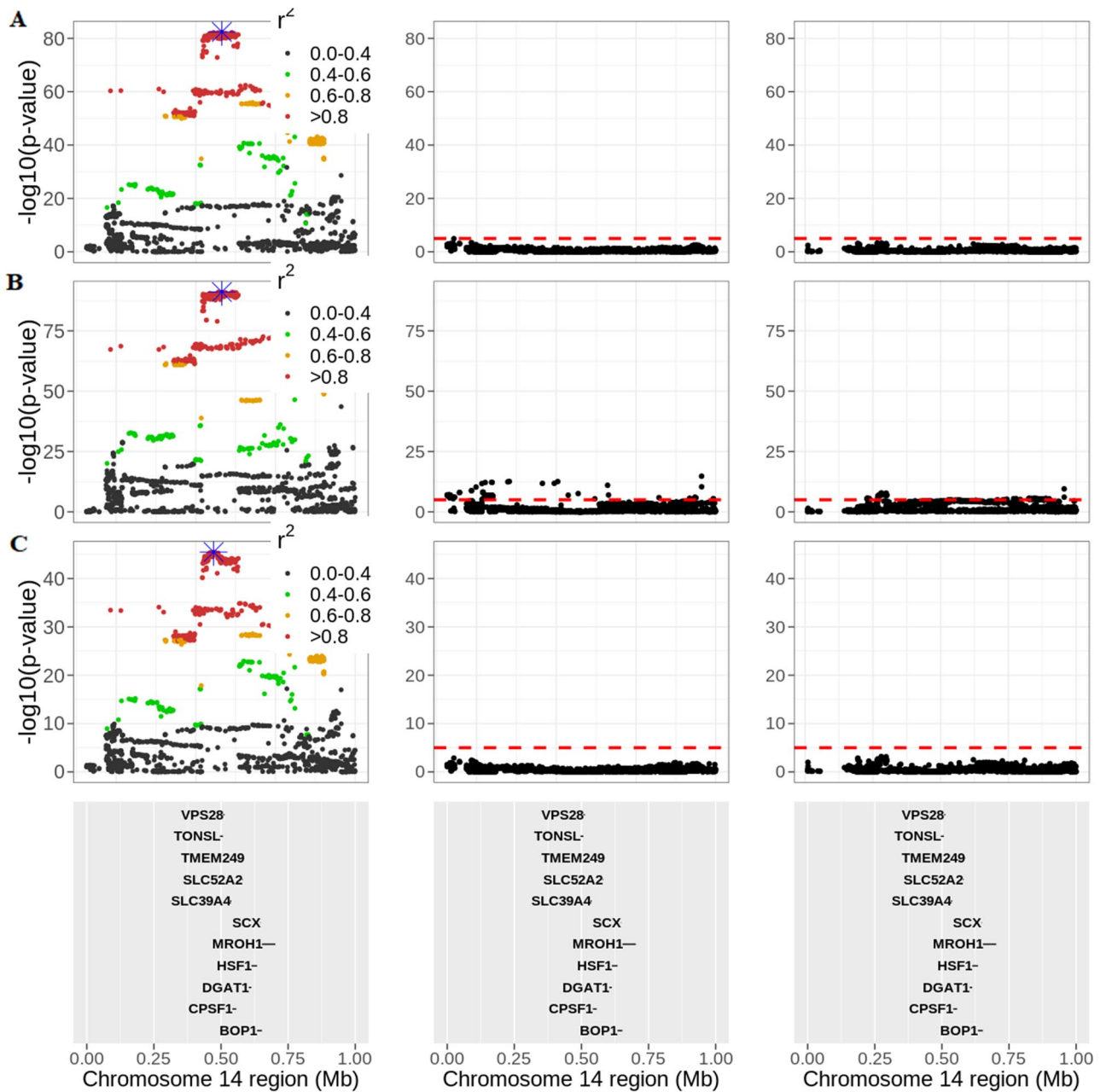
Figure 2. QTL discovery on chromosome 14 at 0 to 1 Mb for heat tolerance milk (HTMYslope; **A**), fat (HTFYslope; **B**), and protein (HTPYslope; **C**) yield slope traits. The three panels represent the GWAS *p* values before conditional analysis (right panel), after conditioning slope traits on the lead SNP (highlighted in blue) defined as the most significant SNP (middle panel), and after conditioning slope traits on the intercept traits (left panel), respectively. The red horizontal dashed line is the GWAS cut-off of $p < 1E-05$. The strength of LD (r^2) between the lead SNP (blue color) and all the other SNPs are color-coded accordingly.

Candidate causal variants for heat tolerance across all analyses. The candidate causal variants for heat tolerance were defined as the lead SNP (most significant SNP within 5 Mb QTL window) plus other significant SNPs in strong LD ($r^2 > 0.8$) with the lead SNP, 500 kb up or downstream of the chromosome. We identified a total of 3010 candidate causal variants for heat tolerance (slope traits) across all the analyses: single-trait GWAS; a meta-analysis of single-trait GWAS results; and meta-analysis of conditional GWAS results for slope traits, most of which were intergenic (N = 1545; 51%) followed by intronic (N = 947; 32%) and upstream (N = 277, 9%) variants (Fig. 3 and Table S1). At least 25 candidate SNPs were missense variants, most (N = 13) of which were in chromosome 14, including two variants (Chr14:615,597 and Chr14:616,087) mapping to HSF1 (heat shock factor 1) gene.

The candidate causal variants for heat tolerance are highly enriched ($p = 8.54E-25$) in the upstream gene regions (Fig. 4), which agrees with GWAS for quantitative traits in humans[32], suggesting that they perhaps play
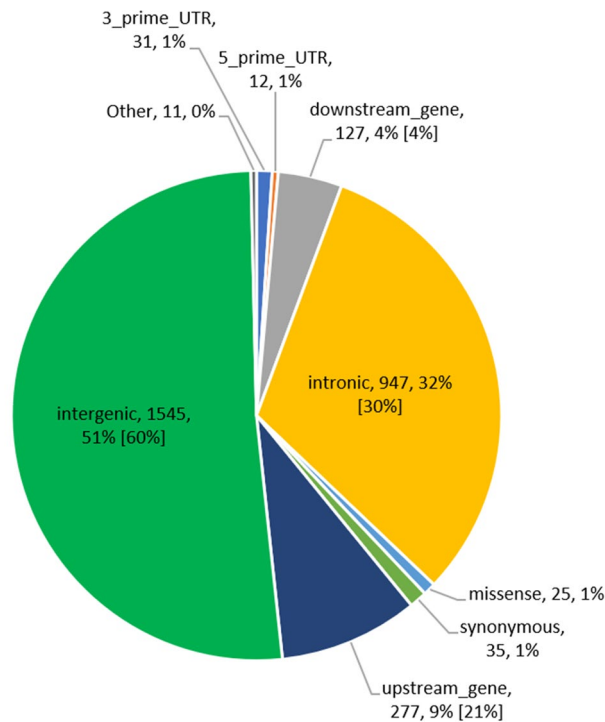
Figure 3. Proportion of candidate causal variants for heat tolerance within different functional classes identified from (**a**) single-trait GWAS, (**b**) meta-analysis, and (**c**) meta-analysis of conditional GWAS results for slope traits. Values in brackets are the proportions of all variants used in the study (~ 15 million SNPs). Functional classes without values in brackets were represented by a small (< 1%) proportion of SNPs in the study dataset.
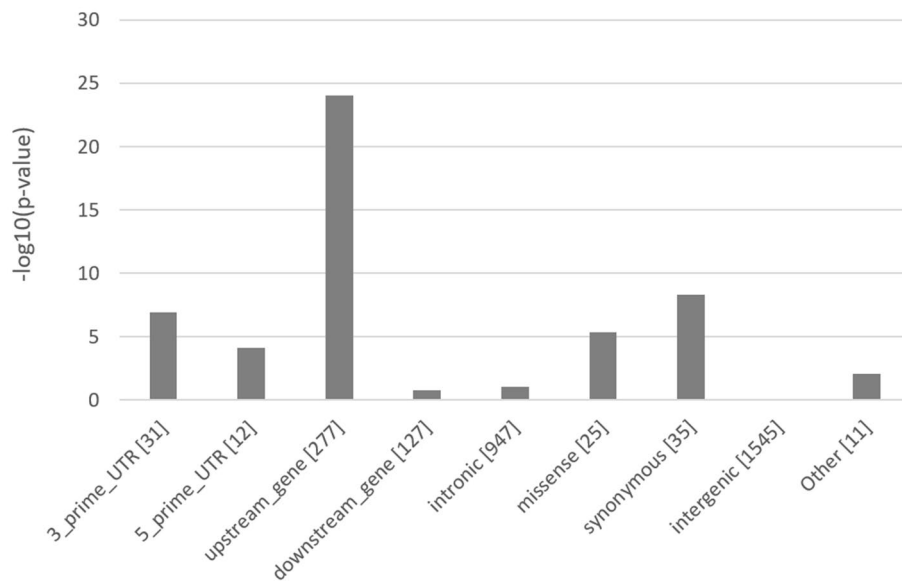

Figure 4. Enrichment of the candidate causal variants for heat tolerance across functional classes. The values in brackets are the number of variants within each class. The class "Other" includes variants with very small proportions of candidate variants (frameshift, stop-codon, splice variants, etc.).

a functional role in regulating gene expression. As expected, most candidate variants have modifier SnpEff[33] predicted impact (Table S5). Two candidate causal mutations detected from the meta-analysis of slope traits have a high SnpEff predicted impact: (a) a stop-gain mutation (Chr5:31,184,185) causing a premature stop codon in the *LALBA* (lactalbumin alpha) gene and (b) a frameshift mutation (Chr29:41,139,622) in *STX5* (syntaxin-5) gene. The two candidate mutations appear to have a stronger effect on milk production compared to heat tolerance. This

Figure 5. Enriched Kyoto Encyclopedia of Gene and Genomes (KEGG) pathways obtained from candidate gene-list for slope traits detected at false discovery rate (FDR < 0.10). SS-slope genes–gene-list from single-trait GWAS; Meta-slope genes–gene-list from multi-trait meta-analysis of slope traits; All-slope genes–combined gene-list from single-trait and meta-analysis. Cells are color-coded according to the strength of the significance for each pathway. Values in brackets are the number of genes within each pathway.

is evidenced by a smaller ($p = 1.39E{-}19$) p value for the stop-gain mutation (Chr5:31,184,185) observed in the meta-analysis of intercept traits compared to the meta-analysis of slope traits ($p = 4.08{-}12$). Similarly, the p value for the frameshift mutation (Chr29:41,139,622) in the *STX5* gene was smaller ($p = 2.06E{-}16$) for the meta-analysis of intercept traits than the meta-analysis of slope traits ($p = 5.06E{-}06$). None of these two candidate stop-gain mutations were significant ($p < 0.05$) following conditional GWAS for slope traits on intercept traits (Table S5).

Using data from[34], which documented over 300 k sequence variants in cattle at highly evolutionarily conserved genome regions across 100 vertebrates (conservation/PhastCon scores > 0.9; see "Methods"), we identified 61 potential functional variants for heat tolerance at these conserved sites in our study (Table S4). However, the candidate causal mutations for heat tolerance are not enriched ($p = 1.0$) in the conserved regions of the genome.

Table 3 provides a short list of putative causal variants (upstream and missense) for heat tolerance that overlap at genomic sites highly conserved across vertebrates. Some of the candidate genes flanking these variants have been reported to be involved with cell survival under stress in animals, e.g., *SCD*[35], *KIAA1324*[36], and *TONSL*[14]. The *SCD* (stearoyl-CoA desaturase) gene encode fatty acid metabolic enzyme and perhaps is required for metabolic homeostasis during heat stress in mammals. Other putative candidate genes for heat tolerance include *KIFC2*, *VPS13B*, and *USP3*. For example, Fang et al.[37] demonstrated that the *USP3* gene, a member of the ubiquitin-specific proteases (USPs) family, is required for eliminating misfolded proteins under heat stress conditions in Yeast.

Pathway enrichment analysis. We generated a list of candidate genes mapping within or near lead SNPs detected at FDR < 0.10 for each trait for the pathway enrichment analyses. We found that the candidate gene-list for slope traits were highly enriched for the KEGG pathways related to the neuronal system (neuroactive ligand–receptor interaction and glutamatergic synapse) and metabolism system (citrate cycle) (Fig. 5). Interestingly, the heat tolerance candidate gene-list (N = ~400 genes) identified from various analyses (single-trait GWAS, meta-analysis, and conditional analysis) were consistently significantly enriched for a neuroactive ligand–receptor interaction pathway comprising of 15 genes (*CALCR*, *PTGER2*, *THRB*, *GRIK2*, *NPY2R*, *F2RL1*, *GRIN2A*, *NR3C1*, *CHRM3*, *GRM8*, *GRM7*, *GRID2*, *NPFFR2*, *MC4R*, *GHR*). A total of 8 genes were enriched ($p = 4.0E{-}03$) in the glutamatergic synapse pathway (*GRIN2A*, *GRM7*, *GRM8*, *ITPR1*, *ITPR2*, *SLC17A6*, *GRIK2*, *GRIA4*). The citrate cycle pathway was also enriched ($p = 1.87E{-}03$), comprising of 5 candidate genes for heat tolerance (*ACLY*, *PDHA2*, *MDH1*, *SUCLG2*, *PCK1*).

We also analysed a smaller set of genes (N = ~230) with the strongest ($p < 1E{-}05$) evidence of association for heat tolerance, separately (that is, the gene-list underlying the candidate causal variants defined as the lead SNP (most significant) within an independent QTL plus other significant SNPs in strong LD ($r^2 > 0.80$) with the lead SNP, 500 kb up or downstream), to see enriched biological pathways. Interestingly, we observed enrichment ($p = 0.02$) of the genes in the neuroactive ligand–receptor interaction pathway, which provides strong support that this neuronal pathway is relevant for heat tolerance comprising of 8 genes (*GHR*, *NPFFR2*, *P2RY8*, *GRIN2A*, *CHRM1*, *THRB*, *CALCR*, *F2RL1*).

When examining the candidate gene-list from single-trait GWAS analyses for slope traits separately, the neuroactive ligand–receptor interaction pathway was overrepresented for candidate gene-list for HTMYslope ($p = 3.19E{-}04$) and HTPYslope ($p = 7.79E{-}03$) traits (Fig. 6). On the other hand, gene-list for HTFYslope were enriched ($p = 1.55E{-}02$) for the axon guidance pathway comprising four genes (*ABLIM2*, *ABLIM3*, *NTN1*, *ROBO1*) and metabolic ($p = 0.06$) pathways.

To further test whether the neuronal pathways are real and not an artifact of our analyses for heat tolerance traits (slopes), we performed enrichment analyses for the significant candidate gene-list for intercepts traits (level of milk production traits). In the candidate gene-list for intercept traits, we found no evidence for enrichment ($p < 0.05$) in any neuronal pathways; thus, providing further favourable support that neuronal pathways are relevant for heat tolerance in mammals.

SNP	Chr	BP	Annotation	VEP_impact	Gene	HTMYslope	HTFYslope	HTPYslope	Meta-analysis[a]	Conditional analysis[b]	p value[c]
rs209684414	3	34,215,670	Upstream	MODIFIER	KIAA1324				✓		5.98E–07
rs207668220	3	34,273,899	Upstream	MODIFIER	C3H1orf194				✓		6.5E–07
rs210324395	10	46,503,113	Upstream	MODIFIER	USP3				✓		8.66E–09
rs210468775	10	46,505,212	Upstream	MODIFIER	USP3				✓		1.89E–09
rs207681599	14	432,274	Upstream	MODIFIER	LRRC14	✓	✓	✓	✓	✓	1.46E–52|1.13E–63|1.13E–28|8.8E–243|7.92E–08
rs136474298	14	471,951	Missense	MODERATE	KIFC2	✓	✓	✓	✓	✓	6.11E–53|4.83E–63|5.59E–29|2.5E–242|7.77E–08
rs207886320	14	479,761	Upstream	MODIFIER	KIFC2	✓	✓	✓	✓	✓	6.18E–52|4.17E–64|4.55E–28|3.9E–242|7.85E–08
rs137472016	14	494,621	Upstream	MODIFIER	TONSL	✓	✓	✓	✓	✓	6.28E–53|8.39E–63|5.94E–29|9.1E–242|7.87E–08
rs445616049	14	64,454,721	Missense	MODERATE	VPS13B	✓			✓		4.18E–08|1.75E–32
rs41946451	20	37,085,370	Missense	MODERATE	CPLANE1					✓	1.14E–07
rs41255693	26	21,272,422	Missense&splice	MODERATE	SCD		✓		✓		4.11E–07|1.85E–08

Table 3. Upstream and missense candidate causal variants for heat tolerance (slope) traits at genomic sites that are highly conserved (conservation score > 0.9) across 100 vertebrate species*. *Conservation scores (PhastCon score) of variants at conserved genomic sites were computed by[34] based on the LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver) human sites calculated across 100 vertebrate species; single-trait GWAS for heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slope traits. [a]Meta-analysis combining single-trait GWAS results for slope traits. [b]Meta-analysis combining single-trait conditional GWAS results for slope traits. [c]For each analysis that identified this variant as significant.
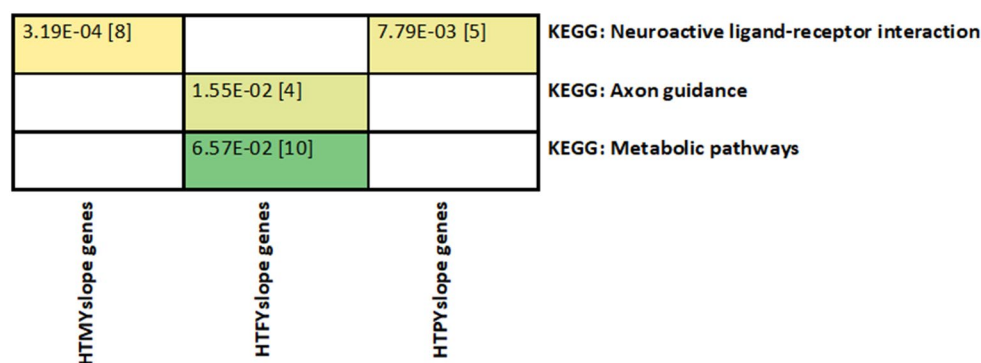

Figure 6. Enriched Kyoto Encyclopedia of Gene and Genomes (KEGG) pathways obtained from our gene-list for single-trait GWAS analysis of slope traits. HTMYslope (heat tolerance milk yield slope); HTFYslope (heat tolerance fat yield slope); and HTPYslope (heat tolerance protein yield slope). Cells are color-coded according to the strength of the significance for each pathway. Values in brackets are the number of genes within each pathway.

Discussion

In this study, we performed a GWAS using a large sample size of Australian dairy cows (N = 29,107) with milk production records and imputed sequence data (~ 15 million SNPs) to identify candidate causal variants and functional genes and pathways associated with heat tolerance. Australia's dairy cattle are uniquely placed for studying heat tolerance in mammals for two main reasons: (1) they are subjected to a wide range of seasonal climatic variations across diverse dairying regions spanning one of the geographically largest countries in the world, and (2) Australia's dairying is predominantly pasture-based with limited heat stress mitigation measures in contrast with those, for example, in North America, where extensive managerial strategies are used more to reduce thermal stress. Overall, we have identified novel candidate causal variants in the neuronal pathways that contribute significantly to heat tolerance in animals.

We leveraged two statistical approaches to identify genetic loci and pathways for heat tolerance: single-trait GWAS linear models and multi-trait meta-analysis. Single-trait GWAS is based on regressing phenotypes on each SNP one at a time. On the other hand, a meta-analysis that combines results of the single-trait GWAS allowed
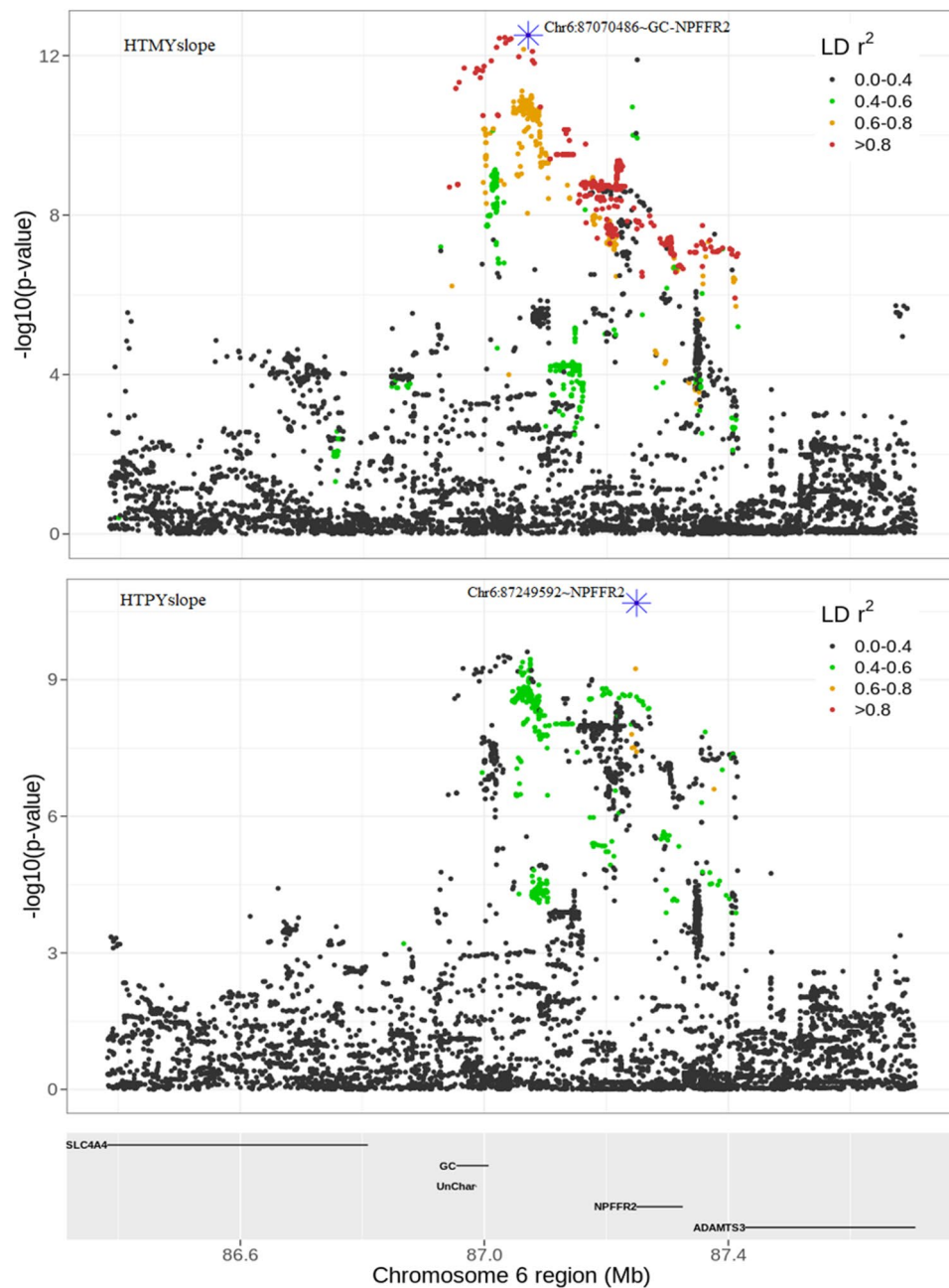
Figure 7. QTL discovery for heat tolerance milk (HTMYslope) and protein (HTPYslope) yield slope traits around the *NPFFR2* gene in bovine chromosome 6.

us to discern putative pleiotropic genetic variants for heat tolerance. Consequently, we identified multiple novel loci for heat tolerance, including 61 potential functional variants at genomic sites highly conserved across 100 vertebrates (Table 3 and Table S4), which could be valuable for fine-mapping and genomic prediction. Studies in humans[38] and cattle[34] have demonstrated that the conserved genomic sites have strong enrichment of trait heritability. Moreover, the results revealed specific candidate causal variants and genes related to neuronal functions for heat tolerance in animals, which we now discuss in more detail.

Heat stress responses are complex adaptations in animals involving many biological pathways, including the nervous system, which connects the internal and external environment to maintain stable core body temperature[39]. Among the candidate gene-list that contribute significantly to heat tolerance in the study animals (Holstein cows), the neuroactive ligand–receptor interaction and glutamatergic synapse pathways (Fig. 5), as components of the nervous system, were highly enriched ($p < 1E-03$) biological features.

At least two candidate variants in the intronic region of *ITPR2* (Chr5:83,330,185; $p = 1.3E-05$) and *GRIA4* (Chr15:2,461,074; $p = 5.8E-05$) genes in the glutamatergic synapse pathway could be potential targets for resilience to environmental stress in animals. *ITPR2* gene was associated with heat stress in the US Holsteins[14] or

sweating rate in humans and mice[40], while the *GRIA4* gene has been linked to thermoregulation in the Siberian cattle[41]. Another candidate variant (Chr22:21,783,956) detected for heat tolerance milk ($p = 3.87$E–05) and protein (7.15E–05) yield slope traits mapped to the intronic region of *ITPR1*—a gene associated with environmental adaptation in the domestic yak[42]. These three lead SNPs for slope traits overlapped with those for intercept traits, with opposing effect direction, suggesting that selecting for these variants may negatively impact milk production.

Previous studies show that the neuroactive ligand–receptor interaction is involved in maintaining energy homeostasis during heat stress in ducks[43]. As protein production is the most valuable output from dairy farms, the focus of breeding programs has been traits associated with yield, with the average milk volume per cow/year almost doubling within the past three decades in Australia[44]. The environmental heat stress, coupled with the elevated metabolic-induced thermogenesis, means that the genetic and cellular reprogramming of pathways such as the nervous system may be necessary to regulate a cascade of hormonal processes such as growth factors, insulin, serotonin, thyroid, prolactin, and mineralocorticoids associated with milk synthesis[45]. We identified 15 genes (FDR < 0.10) associated with the neuroactive ligand–receptor interaction, which could be relevant for metabolic homeostasis in cattle during thermal stress, of which three candidate genes (*GHR, NPFFR2,* and *CALCR*) showed the strongest evidence ($p < 1$E–05).

Here we discuss the evidence for each of these three candidate genes:

(1) Zhang et al.[24] demonstrated that the *NPFFR2* (neuropeptide FF receptor-2) gene, which is mainly expressed by neurons in the brain, plays a crucial role in regulating diet-induced thermogenesis and bone homeostasis in mice. In this study, two lead SNPs (Chr6:87,070,486 and Chr6:87,249,592), detected from single-trait GWAS for HTMYslope and HTPYslope ($p < 1$E–05) mapped to the intergenic and intronic regions of *NPFFR2* gene in BTA 6, respectively. Physiological studies suggest that NPFF family genes regulate feeding behaviour and energy expenditure in mammals reviewed in[46]. During heat events, dairy cattle typically reduce their dry matter intake by up to 30%, perhaps as part of an adaptive mechanism to depress metabolic heat production[47]. Other studies, e.g.,[48] show that inhibition of NPFF receptors induces hypothermia in mice. A recent review by Nguyen et al.[49] indicates that NPFF and its receptors have many promising therapeutic applications including pain, cardiovascular, and feeding regulations in mammals. By examining the genomic region around the *NPFFR2* gene (Fig. 7), it is more likely that the two lead SNPs within this QTL represent separate candidate causal mutations since they are not in strong LD. Interestingly, although the lead SNP (Chr6:87,070,486) for slope trait overlapped with the lead SNP detected for the milk yield intercept (MYint), we observed stronger evidence for the slope (HTMYslope; $p = 3.05$E–13) than the intercept (MYint; $p = 4.19$E–10), suggesting that this SNP is a good candidate for heat tolerance. Besides, this lead SNP (Chr6:87,070,486) remained significant ($p = 6.36$E–06) following single-trait conditional GWAS analysis for HTMYslope trait (conditioning slopes on the intercept traits) as well as in the meta-analysis of single-trait conditional GWAS results for slope traits ($p = 3.74$E–06).

(2) Calcitonin receptors regulate daily body temperature rhythm in mammals and insects and are essential for maintaining homeostasis[50]. In this study, the lead SNP (Chr4:10,815,768) was intronic in the *CALCR* (calcitonin receptor) gene, perhaps indicating that it could be relevant for animals experiencing recurrent or chronic stress, such as in Australian seasonal summers. The strong GWAS signal around this QTL (Supplementary Fig. S8) suggests that the *CALCR* gene likely harbours causal mutations affecting heat tolerance. Dairy cattle employ various adaptive behavioural strategies during heat stress such as reduced feed intake, increased volume, and frequency of water intake, increased standing time, shade seeking, and grazing at cooler day time. We think that *CALCR* is likely involved with some of these heat-stress adaptive behaviours in dairy cattle. Future studies are needed to confirm this, particularly by combining production traits with other relevant behavioural phenotypes such as panting scores from high-throughput recording devices, e.g., activity-based collars.

(3) The expression of the *GHR* (growth hormone receptor) gene is down-regulated during heat stress in livestock, including dairy cows[51] and avian species[43]. The adaptive physiological significance of this down-regulation is not well understood, and it is partly independent of the nutritional level of the animal[51]. In this study, the lead SNP (Chr20:32,103,408; $p = 2.01$E–08) identified only in one slope trait (HTMYslope) based on significant cut-off of $p < 1$E–05 mapped to intronic region of *GHR* gene (Supplementary Fig. S9). However, we found a stronger signal after combining the GWAS results for all the slope traits in a meta-analysis with the lead SNP (Chr20:32,201,287; $p = 1.7$E–47) mapping to the intergenic (~ 22 kb) region of the *GHR* gene, which confirms the pleiotropic effect of this QTL[22]. Also, we observed no significant SNP ($p < 1$E–05) around this QTL following single-trait conditional analyses, but a somewhat strong signal emerged when we combined single-trait conditional GWAS results in the meta-analysis, for which the lead SNP (Chr20:32,226,298; $p = 5.35$E–07) mapped to the intergenic region (~ 47 kb) of GHR. This further supports a possible second QTL that is independent of the level of milk production and shows pleiotropy for the heat tolerance traits. Other published GWAS have also reported an association of the GHR gene with milk production in heat-stressed cows[14] and respiratory rates in pigs during heat stress[19]. Several studies have also implicated the *GHR* polymorphisms to milk production in cattle, e.g., Chr20:31,888,449 phenylalanine-to-tyrosine missense mutation[29]. This mutation was not in strong LD ($r^2 > 0.8$) with the lead SNP detected for slope traits in our study. Taken together, polymorphisms around the *GHR* gene could be candidate targets for improving thermotolerance in livestock, although with possible antagonistic effect on milk production considering, for example, the opposing effect direction observed for the lead SNP (Chr20:32,103,408) within this QTL on the slope (HTMYslope) and intercept (MYint) traits.

There is general agreement that heat stress decreases milk yield (milk, proteins, fat, etc.) in dairy cattle. However, the genetic and biological basis for this reduction is still unclear. Evidence suggests that the reduced feed intake in heat-stressed dairy cows partially explains (35–50%) reduced milk yield and composition[47]. The molecular control and pathways for individual milk traits during heat stress are scarce and inconclusive. In this study, the QTLs detected for the heat tolerance traits varied across the three milk traits (HTMYslope, HTFYslope, HTPYslope), suggesting that they are, in part, regulated by different genes in heat-stressed cows. The greater overlap of candidate genes observed for HTMYslope and HTPYslope traits was expected due to their relatively high correlation (0.90) compared to HTMYslope and HTFYslope (0.56) or HTPYslope and HTFYslope (0.62). These correlations appear to mirror the proportions of SNPs with the same or inconsistent effect direction observed for significant SNPs between slope traits. Considering that heat stress alters carbohydrate, lipid, and amino acid metabolism[52], the large proportion of SNPs with inconsistent effect direction, particularly between HTPYslope and HTFYslope, suggest that these traits are somewhat differently regulated in heat-stressed dairy cows.

Several pair-fed studies suggest that pathways related to the mammary gland protein synthesis govern protein production under heat stress in dairy cows, in part, via reduced amino acid supply to the mammary gland, e.g.,[53,54]. We found that the candidate genes for HTMYslope and HTPYslope traits were overrepresented ($p < 0.005$) in the neuroactive ligand–receptor interaction pathway. This agrees with Pegolo et al.[55] that genes associated with milk proteins are involved in neuronal signaling pathways in dairy cattle. However, it remains unclear how this pathway is regulated during heat stress conditions in dairy cows to impact protein production.

On the other hand, the molecular pathways for fat production under heat stress conditions have not been widely studied. Some studies e.g.,[56] suggest that the reduced activation of PPAR (peroxisome proliferator-activated receptor) signaling pathways leads to decreased expression of genes associated with fat metabolism. Candidate genes for HTFYslope identified in this study are associated with the KEGG term "metabolic pathways" (Fig. 6). Five candidate genes (DMGDH, PDHA2, UGP2, MDH1, PRDX6, NDUFA13) within this pathway may be involved with alleviating oxidative stress in heat-stressed cows. In line with these findings, we found that the candidate genes for heat tolerance (Fig. 5) are overrepresented in the citrate cycle/TCA pathway, which is central to mitochondrion energetics, and might serve to reduce substrate oxidation and reactive oxygen species (ROS) production, thus preventing cellular damage during heat stress.

Notably, our pathway results are perhaps not directly comparable to most previous work in which the study cows were subjected to short-term acute heat stress under experimental conditions e.g.,[56] whereas the current work mimics recurrent or chronic stress that dairy cows experience during summer seasons in Australia. The effects of heat stress in livestock depend on its duration and severity, with the most recent work in Arabian camels somatic cells showing that acute heat stress elevates the expression of heat shock proteins and DNA repair enzymes while chronic heat leads to changes in cell integrity and reduction of total protein levels, metabolic enzymes, and cytoskeletal proteins[57]. Our candidate QTLs are particularly important since it provides novel insights into the molecular aspect of chronic stress considering that the study animals are predominantly reared under outdoor conditions with limited heat stress mitigations. Future studies are required to confirm if these QTLs are involved with recurrent chronic stress in other animal species.

We could not replicate most of the candidate genes with published GWAS results for heat tolerance in cattle, likely for several reasons. First, all comparable earlier studies were much smaller (< 5,000 animals) and therefore were under-powered, and the marker density used was typically 50 k or 600 k SNP array e.g.,[13,14,58]. As expected, we observed that our sequence variants showed markedly higher significance levels than the 50 k SNP array and increased the number of significant peaks across the genome (Supplementary Fig. S1). Second, the trait used to define heat tolerance in this study (i.e., the rate of milk yield decline under heat stress) differs from many other studies e.g.,[12], which used measures of core body temperatures in their GWAS. Given that heat tolerance is a complex trait involving a wide array of adaptive responses (behavioural, physiological, cellular, etc.), different QTLs may be captured by different traits used in GWAS. Notably, although heat tolerance traits (slopes) used in earlier studies in Australia[13,58] were comparable to those used in our study, we could not confirm most candidate genes (except HSF1 gene). This is likely due to the reduced power in earlier GWAS studies (they used smaller sample size and 50 k or 600 k SNP data). Third, differences in the patterns of LD among study populations used and imputation quality may have implications on GWAS, particularly in the detection of putative causal mutations[59]. Here we explored QTLs for heat tolerance in purebred Holstein cows, while some other studies, e.g.,[15] have used crossbred cattle. Collectively, these factors likely impacted the replication of previous GWAS candidate genes for heat tolerance.

Although we detected multiple candidate causal variants for heat tolerance in this study, it appears that larger sample size (we used N = 29,107) would be beneficial considering the polygenic architecture of this trait. Larger sample size is required to detect causal variants with very small effects and the effects of rare causal variants[16]. For example, many of the lead SNPs (most significant) for heat tolerance were tagged by none or very few significant SNPs (Table S1), which may be false-positive variants passing the GWAS cut-off ($p < 1E−05$).

With the increasing availability of high-throughput data from automatic sensor devices such as activity-based collars or tags, it is now feasible to obtain large-scale data for thousands of animals; if genotyped, it would allow a comprehensive genetic evaluation of heat tolerance from a wide array of phenotypes e.g., mid-infrared (MIR) predicted traits from milk recording data[60]. Furthermore, we used conditional analyses of slope traits in a bid to separate production and heat tolerance genes. It may be useful to consider alternative heat tolerance traits in the future GWAS (besides milk decays) that are independent of production, such as those derived from milk yield based on the principal components (PC)[61] or eigen-functions[62]. Overall our results support the highly polygenic nature of heat tolerance characterised by multiple small-effect variants, suggesting that this trait is more amenable to genomic selection tools such as those currently implemented in the Australian dairy industry[3, 63] rather than approaches that exploit few QTLs with large effects. The significant variants detected in this study will be tested in a follow-up study to assess their benefits in the genomic prediction of heat tolerance in dairy cattle.

In conclusion, we performed GWAS for heat tolerance using large sample size and genotype dataset for dairy cattle. The increased sample size and high-resolution SNP data in our study compared to previous reports allowed us unprecedented power and precision of the GWAS to pinpoint multiple putative causal mutations, including 61 potential functional variants at genomic sites highly conserved across 100 vertebrate species. Also, results indicate that different genes and pathways, in part, regulate different milk production traits (milk, fat, and proteins) in heat-stressed dairy cows with a substantial overlap of genes for heat tolerance milk and protein yield. Overall, the results revealed the importance of variation in genes related to the neuronal functions for heat tolerance in mammals, which is of interest for future research towards understanding and managing heat stress for warm climates and particularly in view of the anticipated rising global temperatures.

Materials and methods

Animals and phenotypes.
No live animals were used in this study. Phenotypes used for GWAS were part of our previous study[64] obtained from DataGene (DataGene Ltd., Melbourne, Australia; https://datagene.com.au/)—the organisation responsible for genetic evaluation of dairy animals in Australia. The phenotypes were test-day milk, fat, and protein yield for Holstein dairy cows collected from dairy herds that were matched with climate data (daily temperature and humidity) obtained from weather stations across Australia's dairying regions. The distribution of dairy herds and weather stations; and the calculation of environmental covariate (i.e., temperature–humidity index (THI)) used here were described in our earlier studies[3,64].

Calculation of heat tolerance phenotypes for cows.
The dataset used to calculate heat tolerance phenotypes for cows was similar to that used by[64], comprising a total of 424,846 test-day milk records for first, second and third lactations from 312 herds and 15,906 herd-test days (HTD) collected over 15 years (2003–2017). A summary of the final dataset is given in Table 1. The rate of decline (slope) in milk, fat, and protein yield due to heat stress events was estimated using a reaction norm models[64]. In these models, data on milk, fat, or protein yield were adjusted for the fixed effects, including herd test day, year season of calving, parity, Legendre polynomials (order 3) on the cow age on the day of test, and the Legendre polynomials (order 8) on the interaction between parity and DIM. Random effects fitted in the model included a random regression on a linear orthogonal polynomial of THI, where the intercept represents the level of mean milk yield and the linear component represents the change in milk yield (slope) due to heat stress for each cow and a residual term. In the model, the threshold of THI was set to 60 following[65]. The analyses to derive trait deviation (TD) which represents a phenotype adjusted for all fixed effects (i.e., the mean/intercept and slope for each cow) were conducted using ASReml v4.2[66]. To facilitate convergence, milk, fat and protein yield traits were scaled by a factor of 10, 100, and 100, respectively. The description of heat tolerance traits (i.e., slopes) used in this study are comparable to those used in previous GWAS in Australia[13,58].

We refer to milk intercept traits as [MYint (i.e., milk yield intercept), FYint (i.e., fat yield intercept), and PYint (i.e., protein yield intercept)] and the slopes traits as [HTMYslope (i.e., heat tolerance milk yield slope), HTFYslope (i.e., heat tolerance fat yield slope), and HTPYslope (i.e., heat tolerance protein yield slope)], respectively.

Genotypes.
Two genotype datasets were analysed for 29,107 Holstein cows with the above phenotypes: 50 k SNP chip and 15,098,486 imputed whole-genome sequence variants (WGS). Most of the cows were originally genotyped with a custom low-density 10 k SNP panel or a standard medium density 50 k SNP array (BovineSNP50k BeadChip: Illumina Inc). The low-density genotypes were imputed to the 50 k array using a reference set of approximately 14,000 animals with real 50 k genotypes, with approximately 7,000 SNPs of the low-density SNP panel overlapping the 50 k SNP array. The 50 k genotypes were then imputed to the high-density Bovine SNP array (HD: BovineHD BeadChip, Illumina Inc) using a reference set of 2,700 animals with real HD genotypes. All SNP BeadChip genotypes were first converted to the ARS-UDC1.2 reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1/)[67] positions from reference genome UMD3.1 and imputed using Fimpute3[68]. The WGS was imputed from the HD genotypes using a reference set of 3,090 *Bos taurus* sequences in the Run7 of the 1000 Bull Genome Project (http://1000bullgenomes.com/)[20] aligned to the ARS-UCD1.2 reference genome. Only bi-allelic sequence variants with a minor allele count (≥ 4) and GATK[69] quality tranche 99.0 or better were retained for imputation. Pre-imputation, we also removed sequence variants from the imputation reference that had a higher than expected proportion of heterozygous calls (> 0.5) if these variants fell in a 500 kb window enriched for variants showing excessive heterozygosity (as a proxy to indicate regions where WGS mapping/alignment may be poor). A total of 31,994,954 sequence variants remained for imputation. Minimac3[70] was used for WGS imputation, having first pre-phased both the HD genotypes and the WGS reference using Eagle v2[71]. For the analysis, we retained only the variants with Minimac3 imputation accuracy, $R^2 > 0.4$ and MAF > 0.005 (N = 15,098,486 sequence variants).

Single-trait GWAS and multi-trait meta-analysis.
A genome-wide association analysis (GWAS) using a mixed linear model was used to test associations between individual SNP and cows' slope [HTMYslope, HTPYslope and HTFYslope] and intercept [MYint, FYint, PYint] traits using GCTA software[72]. Because phenotypes were TD already adjusted for nongenetic effects, for each autosomal SNP i with minor allele frequency (MAF) > 0.005, the fitted model per trait was,

$$y = \text{mean} + x\beta + g + \varepsilon,$$

where y was the vector of TD (intercept or slope traits) for cows (n = 29,107), β was the allele substitution effect of SNP i, x was the vector of genotype dosages (0, 1, or 2) for SNP i, g was the vector of polygenic effect

12

with $g \sim N(0, GRM\sigma_g^2)$ and ε was a vector of the residual effect with $e \sim N(0, I\sigma_e^2)$, where I was an $n \times n$ identity matrix. The variance of y was $var(y) = GRM\sigma_g^2 + I\sigma_e^2$ where GRM is the genomic relationship matrix between cows, and σ_g^2 and σ_e^2 were the genetic and residual variances. For animal j and k relationship was calculated using GCTA[72] as follows:

$$A_{jk} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)}$$

where A_{jk} are the off-diagonal elements of GRM for animal j and k; N = total number of SNPs from 50 k SNP array data (MAF > 0.005; 45,504 SNPs); x_{ij} and x_{ik} are the number of copies for reference allele for the ith SNP; and p_i is the allele frequency for ith SNP.

Genomic heritability was calculated for each trait using variance component estimates from –reml option of GCTA for 50 k SNP array (45,504 SNPs) data of cows (N = 29,107): $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$.

To increase the power of GWAS and identify pleiotropic variants, we next combined single-trait GWAS results obtained above in a multi-trait meta-analysis following[21]. The multi-trait chi-squared (χ^2) statistics for ith SNP was calculated separately for intercept [MYint, FYint, PYint] and slope [HTMYslope, HTFYslope, and HTPYslope] traits as follows:

$$\chi^2 = t_i' V^{-1} t_i$$

where t_i is the vector of 3×1 vector of signed t-values (i.e., b/se) of ith SNP for either intercept or slope traits; and V^{-1} is the inverse of 3×3 correlation matrix of the signed t-values calculated based on all pairs for the intercept or slope traits. The significance of χ^2 value for ith SNP was calculated based on chi-squared distribution with 3 degrees of freedom—that is number of traits for either intercept or slope traits.

Conditional GWAS analysis. Next, we performed two conditional GWAS strategies of slope traits using GCTA software[72] to test somewhat different hypotheses:

Conditional analysis of slope traits on lead SNP (i.e., most significant SNP within a chromosome from first-round GWAS)—aimed at identifying additional or secondary putative causal variants beside those detected from first-round GWAS. We performed a conditional analysis strategy on two chromosomes (BTA 14 and BTA 20), which showed the strongest GWAS signal for slope traits in the first-round GWAS (Supplementary Figs. S1 and S2) and are known to harbour QTLs with major effects on milk production (i.e., BTA14 ~ *DGAT1* and BTA20 ~ *GHR* gene).

Conditional analysis of slope traits on intercept traits—aimed at identifying QTLs for heat tolerance that are independent, or not also strongly associated with the level of milk production. We fitted the intercept traits of MYint, FYint, and PYint, as a covariate in the linear model when analysing the HTMYslope, HTFYslope, or HTPYslope, respectively. To increase the power of GWAS, we then combined conditional GWAS results for the three slope traits [HTMYslope, HTFYslope, and HTPYslope] in a multi-trait meta-analysis following[21] as described earlier.

Identifying candidate causal variants. We used the following criteria to select candidate variants ($p < 1.0E-05$) from the three analytical approaches (single-trait GWAS, meta-analysis, conditional analysis).

1. For each trait, select all SNPs with $p < 1E-05$ (FDR < 0.10).
2. Split each chromosome (N = 1…29) into 5 Mb non-overlapping windows from the start to the distal end of the chromosome.
3. Within the ith 5 Mb window, select the most significant SNP (i.e., the SNP with the smallest p value below the threshold of $p < 1E-05$) defined as the 'lead SNP'. We chose this arbitrary 5 Mb window size to obtain a small set of significant lead SNPs representing independent QTL (that is, not in linkage disequilibrium) for further detailed examination.
4. Calculate the LD between each lead SNP and all the other SNPs within 500 kb up and downstream of the lead SNP using Plink v1.9[73].
5. For each lead SNP, extract all the significant SNPs ($p < 1E-05$) in strong LD ($r^2 > 0.80$) with the lead SNP within 500 kb up or down downstream – to account for the fact that the lead SNP (most significant) is not necessarily the causal variant.

Annotation of sequence variants and enrichment analysis. Annotation of all variants (~ 15 million SNPs) was performed using SnpEff[33] tool. Using the annotation, we grouped the candidate causal variants for heat tolerance (slopes) into 9 classes (intergenic, intronic, missense, upstream, downstream, 3_prime_UTR, synonymous, 5_prime_UTR, and Other) and performed enrichment analysis using phyper in R v3.61[74]. The class "Other" comprised variants including 5_prime_UTR_premature/_start_codon_gain, frameshift, missense&splice, splice&intron, stop_gained, etc. Supplementary Table S1 provides the number of candidate causal variants for heat tolerance within the 9 classes.

Candidate variants at conserved genomic sites. We identified candidate causal variants for heat tolerance at highly conserved genomic sites using data from[34]. Briefly, these authors documented over 300 k sequence

variants at conserved sites in cattle based on the LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver) human sites with conservation scores (PhastCon score) > 0.9 calculated across 100 vertebrate species (see https://www.pnas.org/content/pnas/suppl/2019/09/07/1904159116.DCSupplemental/pnas.1904159116.sapp.pdf for more details).

Pathway enrichment analysis. We generated candidate gene-list mapping near or underlying lead SNPs (most significant SNPs within 5 Mb QTL windows) identified at FDR < 0.10 cut-off threshold from both single-trait and multi-trait analyses of intercept or slope traits. For intergenic lead SNPs, we selected the closest gene on either side of the SNP. We chose this cut-off (FDR < 0.10) instead of a more stringent $p < 1E-05$ to include genes associated with smaller effects while guarding against false positives. We then performed the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis using DAVID[75].

We also performed enrichment test separately for the gene-list associated with potential major effects on heat tolerance identified across all analyses (i.e., gene-list with the strongest ($p < 1E-05$) evidence of association defined as the candidate causal variants (i.e., lead SNP + other significant SNPs in strong LD ($r^2 > 0.80$) with the lead SNP within 500 kb up or downstream passing the cut-off p value of $1 < 1E-05$). For all the analyses, we considered functional pathways with Fisher's $p < 0.05$ as significantly enriched.

Data availability
Positions and annotations for all the lead SNPs (most significant SNPs) with $p < 1E-5$ are in Tables S1–S3. Data-Gene (DataGene Ltd., Melbourne, Australia; https://datagene.com.au/) are the custodians of the raw phenotype and genotype data of Australian farm animals. Research related requests for access to the data may be accommodated on a case-by-case basis.

References
1. Li, D., Yuan, J. & Kopp, R. B. Escalating global exposure to compound heat-humidity extremes with warming. *Environ. Res. Lett.* **15**, 064003 (2020).
2. Xu, C., Kohler, T. A., Lenton, T. M., Svenning, J.-C. & Scheffer, M. Future of the human climate niche. *Proc. Natl. Acad. Sci.* **117**, 11350–11355 (2020).
3. Nguyen, T. T., Bowman, P. J., Haile-Mariam, M., Pryce, J. E. & Hayes, B. J. Genomic selection for tolerance to heat stress in Australian dairy cattle. *J. Dairy Sci.* **99**, 2849–2862 (2016).
4. Polsky, L. & von Keyserlingk, M. A. Invited review: Effects of heat stress on dairy cattle welfare. *J. Dairy Sci.* **100**, 8645–8657 (2017).
5. St-Pierre, N., Cobanov, B. & Schnitkey, G. Economic losses from heat stress by US livestock industries. *J. Dairy Sci.* **86**, E52–E77 (2003).
6. Hansen, P. Physiological and cellular adaptations of zebu cattle to thermal stress. *Anim. Reprod. Sci.* **82**, 349–360 (2004).
7. Bryant, J., López-Villalobos, N., Pryce, J., Holmes, C. & Johnson, D. Quantifying the effect of thermal environment on production traits in three breeds of dairy cattle in New Zealand. *N. Z. J. Agric. Res.* **50**, 327–338 (2007).
8. Sejian, V., Bhatta, R., Gaughan, J., Dunshea, F. & Lacetera, N. Adaptation of animals to heat stress. *Animal* **12**, s431–s444 (2018).
9. Ravagnolo, O., Misztal, I. & Hoogenboom, G. Genetic component of heat stress in dairy cattle, development of heat index function. *J. Dairy Sci.* **83**, 2120–2125 (2000).
10. Dikmen, S., Cole, J., Null, D. & Hansen, P. Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. *J. Dairy Sci.* **95**, 3401–3405 (2012).
11. Littlejohn, M. D. *et al.* Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nat. Commun.* **5**, 1–8 (2014).
12. 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).
13. Wang, T. *et al.* Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. *BMC Genomics* **18**, 618 (2017).
14. Sigdel, A., Abdollahi-Arpanahi, R., Aguilar, I. & Peñagaricano, F. Whole genome mapping reveals novel genes and pathways involved in milk production under heat stress in US Holstein cows. *Front. Genet.* **10**, 928 (2019).
15. Otto, P. I. *et al.* Genome-wide association studies for heat stress response in Bos taurus× Bos indicus crossbred cattle. *J. Dairy Sci.* **102**, 8148–8158 (2019).
16. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
17. Taye, M. *et al.* Whole genome detection of signature of positive selection in African cattle reveals selection for thermotolerance. *Anim. Sci. J.* **88**, 1889–1901 (2017).
18. Freitas, P. H. *et al.* Genetic diversity and signatures of selection for thermal stress in cattle and other two bos species adapted to divergent climatic conditions. *Front. Genet.* **12**, 102 (2021).
19. Kim, K.-S. *et al.* Characterization of the acute heat stress response in gilts: III. Genome-wide association studies of thermotolerance traits in pigs. *J. Anim. Sci.* **96**, 2074–2085 (2018).
20. Hayes, B. J. & Daetwyler, H. D. 1000 bull genomes project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annu. Rev. Anim. Biosci.* **7**, 89–102 (2019).
21. Bolormaa, S. *et al.* A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet* **10**, e1004198 (2014).
22. Xiang, R., van den Berg, I., MacLeod, I. M., Daetwyler, H. D. & Goddard, M. E. Effect direction meta-analysis of GWAS identifies extreme, prevalent and shared pleiotropy in a large mammal. *Commun. Biol.* **3**, 88. https://doi.org/10.1038/s42003-020-0823-6 (2020).
23. Carvalheiro, R. *et al.* Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. *Genet. Sel. Evol.* **51**, 29 (2019).
24. Zhang, L. *et al.* Diet-induced adaptive thermogenesis requires neuropeptide FF receptor-2 signalling. *Nat. Commun.* **9**, 1–19 (2018).
25. Joutsen, J. *et al.* Heat shock factor 2 protects against proteotoxicity by maintaining cell-cell adhesion. *Cell Rep.* **30**, 583–597. e586 (2020).

26. Grisart, B. *et al.* Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proc. Natl. Acad. Sci.* **101**, 2398–2403 (2004).

27. Winter, A. *et al.* Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA: Diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc. Natl. Acad. Sci.* **99**, 9300–9305 (2002).

28. Littlejohn, M. D. *et al.* Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. *Sci. Rep.* **6**, 25376 (2016).

29. Blott, S. *et al.* Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* **163**, 253–266 (2003).

30. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**, 9440–9445 (2003).

31. Menegazzi, P., Yoshii, T. & Helfrich-Förster, C. Laboratory versus nature: The two sides of the Drosophila circadian clock. *J. Biol. Rhythms* **27**, 433–442 (2012).

32. Sveinbjornsson, G. *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* **48**, 314 (2016).

33. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

34. Xiang, R. *et al.* Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc. Natl. Acad. Sci.* **116**, 19398–19408 (2019).

35. Dado-Senn, B. *et al.* RNA-Seq reveals novel genes and pathways involved in bovine mammary involution during the dry period and under environmental heat stress. *Sci. Rep.* **8**, 1–11 (2018).

36. Deng, L., Feng, J. & Broaddus, R. The novel estrogen-induced gene EIG121 regulates autophagy and promotes cell survival under stress. *Cell Death Dis.* **1**, e32–e32 (2010).

37. Fang, N. N., Zhu, M., Rose, A., Wu, K.-P. & Mayor, T. Deubiquitinase activity is required for the proteasomal degradation of misfolded cytosolic proteins upon heat-stress. *Nat. Commun.* **7**, 1–16 (2016).

38. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228 (2015).

39. Nakamura, K. & Morrison, S. F. A thermosensory pathway that controls body temperature. *Nat. Neurosci.* **11**, 62–71 (2008).

40. Klar, J. *et al.* Abolished InsP 3 R2 function inhibits sweat secretion in both humans and mice. *J. Clin. Investig.* **124**, 4773–4780 (2014).

41. 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).

42. Basang, W. & Zhu, Y. Whole-genome analysis identifying candidate genes of altitude adaptive ecological thresholds in yak populations. *J. Anim. breed. Genet.* **136**, 371–377 (2019).

43. Kim, J.-M. *et al.* Identification of the acclimation genes in transcriptomic responses to heat stress of White Pekin duck. *Cell Stress Chaperones* **22**, 787–797 (2017).

44. Pryce, J., Nguyen, T., Axford, M., Nieuwhof, G. & Shaffer, M. Symposium review: Building a better cow—The Australian experience and future perspectives. *J. Dairy Sci.* **101**, 3702–3713 (2018).

45. Bernabucci, U. *et al.* Metabolic and hormonal acclimation to heat stress in domesticated ruminants. *Animal Int. J. Anim. Biosci.* **4**, 1167 (2010).

46. Bechtold, D. A. & Luckman, S. M. The role of RFamide peptides in feeding. *J. Endocrinol.* **192**, 3–15 (2007).

47. Wheelock, J., Rhoads, R., VanBaale, M., Sanders, S. & Baumgard, L. Effects of heat stress on energetic metabolism in lactating Holstein cows. *J. Dairy Sci.* **93**, 644–655 (2010).

48. Wang, Y.-Q. *et al.* Neuropeptide FF receptors antagonist, RF9, attenuates opioid-evoked hypothermia in mice. *Peptides* **29**, 1183–1190 (2008).

49. Nguyen, T., Marusich, J., Li, J.-X. & Zhang, Y. Neuropeptide FF and its receptors: therapeutic applications and ligand development. *J. Med. Chem.* **63**, 12387 (2020).

50. Goda, T. *et al.* Calcitonin receptors are ancient modulators for rhythms of preferential temperature in insects and body temperature in mammals. *Genes Dev.* **32**, 140–155 (2018).

51. Rhoads, M. *et al.* Effects of heat stress and nutrition on lactating Holstein cows: II. Aspects of hepatic growth hormone responsiveness. *J. Dairy Sci.* **93**, 170–179 (2010).

52. Baumgard, L. H. & Rhoads, R. P. Jr. Effects of heat stress on postabsorptive metabolism and energetics. *Annu. Rev. Anim. Biosci.* **1**, 311–337 (2013).

53. Cowley, F., Barber, D., Houlihan, A. & Poppi, D. Immediate and residual effects of heat stress and restricted intake on milk protein and casein composition and energy metabolism. *J. Dairy Sci.* **98**, 2356–2368 (2015).

54. Gao, S. *et al.* The effects of heat stress on protein metabolism in lactating Holstein cows. *J. Dairy Sci.* **100**, 5040–5049 (2017).

55. Pegolo, S. *et al.* Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. *Sci. Rep.* **8**, 1–15 (2018).

56. Gao, S. *et al.* Heat stress negatively affects the transcriptome related to overall metabolism and milk protein synthesis in mammary tissue of midlactating dairy cows. *Physiol. Genomics* **51**, 400–409 (2019).

57. Saadeldin, I. M. *et al.* Thermotolerance and plasticity of camel somatic cells exposed to acute and chronic heat stress. *J. Adv. Res.* **22**, 105–118 (2020).

58. Hayes, B. J. *et al.* A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PloS One* **4**, e6676 (2009).

59. Pausch, H. *et al.* Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genet. Sel. Evol.* **49**, 24 (2017).

60. van den Berg, I. *et al.* The use of milk mid-infrared spectroscopy to improve genomic prediction accuracy of serum biomarkers. *J. Dairy Sci.* **104**, 2008–2017 (2021).

61. Macciotta, N. *et al.* Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. *J. Dairy Sci.* **100**, 4683–4697 (2017).

62. Carabaño, M., Bachagha, K., Ramón, M. & Díaz, C. Modeling heat stress effect on Holstein cows under hot and dry conditions: Selection tools. *J. Dairy Sci.* **97**, 7889–7904 (2014).

63. Nguyen, T. T. *et al.* Implementation of a breeding value for heat tolerance in Australian dairy cattle. *J. Dairy Sci.* **100**, 7362–7367 (2017).

64. Cheruiyot, E. K. *et al.* Genotype-by-environment (temperature–humidity) interaction of milk production traits in Australian Holstein cattle. *J. Dairy Sci.* **103**, 2460–2476. https://doi.org/10.3168/jds.2019-17609 (2020).

65. Hayes, B., Carrick, M., Bowman, P. & Goddard, M. Genotype × environment interaction for milk production of daughters of Australian dairy sires from test-day records. *J. Dairy Sci.* **86**, 3736–3744 (2003).

66. Gilmour, A., Gogel, B., Cullis, B., Welham, S. & Thompson, R. ASReml user guide release 4.2 functional specification. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK* www.vsni.co.uk *(2015)*.

67. Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience* **9**, giaa021 (2020).

68. Sargolzaei, M., Chesnais, J. P. & Schenkel, F. S. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* **15**, 478 (2014).
69. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
70. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
71. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443 (2016).
72. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Human Genet.* **88**, 76–82 (2011).
73. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* **81**, 559–575 (2007).
74. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. (2018).
75. Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44 (2009).

Author contributions

J.E.P., H.M., I.M.M., and E.K.C., conceived and designed the study; I.M.M. assisted with preparation and imputation of genotype data; E.K.C., I.M.M., M.H., J.E.P., B.G.C., R.X., contributed to the formal data analysis; E.K.C. wrote the first draft; all authors, reviewed and approved the final manuscript for publication.

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/s41598-021-95816-8.

Correspondence and requests for materials should be addressed to M.H.-M.

Reprints and permissions information is available at www.nature.com/reprints.

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