PLOS ONE

RESEARCH ARTICLE

Integrative QTL mapping and selection signatures in Groningen White Headed cattle inferred from whole-genome sequences

Rayner Gonzalez-Prendes[1]*, Catarina Ginja[2], Juha Kantanen[3], Nasser Ghanem[4], Donald R. Kugonza[5], Mahlako L. Makgahlela[6,7], Martien A. M. Groenen[1], Richard P. M. A. Crooijmans[1]

1 Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands, **2** BIOPOLIS/CIBIO/ InBIO, Research Center in Biodiversity and Genetic Resources, University of Porto, Vairão, Portugal, **3** Natural Resources Institute Finland, Jokioinen, Finland, **4** Animal Production Department, Faculty of Agriculture, Cairo University, Giza, Egypt, **5** Department of Agricultural Production, College of Agricultural and Environmental Sciences, Makerere University, Kampala, Uganda, **6** Agricultural Research Council-Animal Production Institute, Irene, South Africa, **7** Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa

* rayner.gonzalezprendes@wur.nl

Citation: Gonzalez-Prendes R, Ginja C, Kantanen J, Ghanem N, Kugonza DR, Makgahlela ML, et al. (2022) Integrative QTL mapping and selection signatures in Groningen White Headed cattle inferred from whole-genome sequences. PLoS ONE 17(10): e0276309. https://doi.org/10.1371/journal.pone.0276309

Editor: Arnar Palsson, University of Iceland, ICELAND

Received: August 4, 2021

Accepted: October 4, 2022

Published: October 26, 2022

Data Availability Statement: The VCF files with variant from all breeds will be available at https://zenodo.org/deposit/6616286. Raw reads will be accessed on https://www.ebi.ac.uk with the Accession number PRJEB56301 before publication.

Funding: The research presented in this publication was funded by the Long-term EU-Africa research and innovation Partnership on food and nutrition security and sustainable Agriculture (LEAP-Agri) as

Abstract

Here, we aimed to identify and characterize genomic regions that differ between Groningen White Headed (GWH) breed and other cattle, and in particular to identify candidate genes associated with coat color and/or eye-protective phenotypes. Firstly, whole genome sequences of 170 animals from eight breeds were used to evaluate the genetic structure of the GWH in relation to other cattle breeds by carrying out principal components and model-based clustering analyses. Secondly, the candidate genomic regions were identified by integrating the findings from: a) a genome-wide association study using GWH, other white headed breeds (Hereford and Simmental), and breeds with a non-white headed phenotype (Dutch Friesian, Deep Red, Meuse-Rhine-Yssel, Dutch Belted, and Holstein Friesian); b) scans for specific signatures of selection in GWH cattle by comparison with four other Dutch traditional breeds (Dutch Friesian, Deep Red, Meuse-Rhine-Yssel and Dutch Belted) and the commercial Holstein Friesian; and c) detection of candidate genes identified via these approaches. The alignment of the filtered reads to the reference genome (ARS-UCD1.2) resulted in a mean depth of coverage of 8.7X. After variant calling, the lowest number of breed-specific variants was detected in Holstein Friesian (148,213), and the largest in Deep Red (558,909). By integrating the results, we identified five genomic regions under selection on BTA4 (70.2–71.3 Mb), BTA5 (10.0–19.7 Mb), BTA20 (10.0–19.9 and 20.0–22.7 Mb), and BTA25 (0.5–9.2 Mb). These regions contain positional and functional candidate genes associated with retinal degeneration (e.g., *CWC27* and *CLUAP1*), ultraviole*t* protection (e.g., *ERCC8*), and pigmentation (e.g. *PDE4D*) which are probably associated with the GWH specific pigmentation and/or eye-protective phenotypes, e.g. Ambilateral Circumocular Pigmentation (ACOP). Our results will assist in characterizing the molecular basis of GWH phenotypes and the biological implications of its adaptation.

part of the OPTIBOV project (LEAP-Agri-326) and co-founded by the European Union's Horizon 2020 research and innovation program under grant agreement No 727715. The funding bodies had no role in the design of the study, the collection, analysis, interpretation of data, or the writing of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Traditional native breeds are an important source of genetic variability adapted to local environments. They might harbor genetic variants unique to the breed due to ecosystem adaptation and, e.g. provide resistance to local diseases and/or extreme climatic conditions. Detailed analyses of the genomic structure of those native breeds can contribute to improving the knowledge about breed formation, and identify genes and variants with a significant impact on the adaptation processes that shaped animal phenotypes [1–4]. This information can be used to set up optimum breeding programs for the management of livestock genomic resources.

The skin and coat color variation in livestock breeds are important traits that impact the adaptation of breeds to the environment [5–8]. In the past years, numerous research projects, such as genome-wide association studies (GWAS) [9–11] and whole-genome selective sweeps identification [3, 12] have been performed to pinpoint candidate genomic regions with significant effects on skin and coat color variation [6, 9, 10, 13–15]. The combination of several sources of information can improve the power of candidate gene identification by reducing the number of QTLs and their intervals, as well as providing additional insights into the studied biological processes [16, 17].

The Groningen White Headed (GWH) breed, originated from the Groningen province of the Netherlands, is a dual-purpose cattle known for its longevity, minimal veterinary costs, and high fertility rate [18]. The first GWH animal was registered in the herd book in 1875, and in 1999, the breed was considered to be endangered with approximately 830 purebreed animals [19]. Recent interest in functional traits such as fertility or resistance may open up new opportunities for the expansion of this breed [18]. GWH animals are easily distinguished by their phenotype, that is, solid black or red coat color, white face, and colored areas around the eyes [18, 19].

In cattle, Ambilateral Circumocular Pigmentation (ACOP) can be distinguished by a white face and colored areas around the eyes in breeds such as the GWH [19] and Fleckvieh [9]. The presence of this phenotype can reduce the susceptibility to eye lesions [20]. It is well-known that non-pigmented animals have a higher incidence of eye lesions than animals with eye margin pigmentation [9, 21]. A plausible explanation for this is that cattle with a non-pigmented eye margin are exposed to more ultraviole*t* (UV) radiation in this region [9], which would be more intense and harmful in the tropical areas [22].

The molecular genetic background of GWH breed has not been extensively studied [23]. Therefore, the goal of this study was to gain further knowledge on the genomic basis of the GWH breed by analyzing whole-genome resequencing data to identify and characterize genomic regions that differ between GWH and other cattle breeds, and in particular to identify candidate genes associated with coat color and/or eye protective phenotypes. We studied the population structure of five Dutch traditional breeds, to evaluate the genetic distinction of the GWH, using two approaches, which are, a model-based clustering admixture analysis and a principal component study (PCA). Additionally, we implemented an integrative approach, to reduce the number of false positive candidate genomic regions, taking into account the findings from: a) a genome-wide association study using GWH with ACOP, breeds without the white head phenotype (Holstein Friesian, Dutch Friesian, Deep Red, Meuse-Rhine-Yssel and Dutch Belted) and other white headed breeds (Simmental and Hereford); b) scans for candidate selective sweeps in GWH cattle compared to those of four other traditional Dutch breeds (Dutch Friesian, Deep Red, Meuse-Rhine-Yssel, Dutch Belted), and the transboundary Holstein Friesian; c) identification of runs of homozygosity (ROH) in the GWH breed to reduce the number of false positive candidate selective sweeps, and d) identification of functional

candidate genes in the genomic regions commonly detected by GWAS, selective sweeps and ROH hostpots.

Materials and methods

Ethics statement

This study was conducted following the animal experimentation policy of Wageningen University & Research. The cattle blood samples were collected by a veterinarian during yearly routine health inspections with written informed consent by the owners. Therefore, no Ethics Committee approval for animal care was needed for this research.

Animals. We used 170 animals from eight breeds. We first sampled 120 unrelated animals as part of the LEAP-Agri project OPTIBOV (https://www.optibov.com/) and in collaboration with the respective breed associations, including 5 Holstein Friesian; 21 GWH; 23 Meuse-Rhine-Yssel; 23 Dutch Belted; 24 Dutch Friesian; and 24 Deep Red. In total, 92 cows and 28 bulls were included in this study (for more details see **S1 Table**). Secondly, white headed animals with no ACOP were retrieved from two more breeds (25 Simmental and 25 Hereford) included in the 1000 Bull Genomes Project (Run9 version) [24, 25]. These 50 animals with completely white heads (lacking ACOP) were used only for the genome-wide association analysis to contrast against the GWH breed, which exhibits ACOP.

DNA sample preparation and sequencing. The GENTRA Blood kit (Qiagen N.V.) was used for the isolation of genomic DNA from EDTA blood samples. The quantification and quality of the obtained DNA were assessed using the Qubit fluorometer (Qiagen N.V.). DNA was paired-end sequenced (read length of 150 base pair) as single-indexed genomic libraries using the Illumina Novaseq6000 (Illumina Inc., USA). Finally, raw reads were preprocessed by trimming the adapter sequences and removing the reads with 50% of low-quality nucleotides and fewer than 36 base pairs in length with fastp v0.23.1 [26].

Short read alignment, mapping, variant detection, and filtering. The mem option from BWA v0.7.17-r1188 [27] was used to map the cleaned reads to the bovine reference genome (assembly version ARS-UCD1.2) [28]. Aligned reads from each animal were stored in binary BAM files using SAM tools v0.1.19 [29]. Freebayes software [30] was used for population-based variant calling with default parameters except for: -min-alternate-count = 3, -haplotype-length = 0, -ploidy = 2, -min-alternate-fraction = 0.2, and -min-base-quality = 30. Variants with a phred-scaled probability < 20 and a depth of coverage by sample <5 were removed using the Bcftools v1.9 [31] software.

Population structure assessment with principal component analysis and individual ancestry estimation. We used PC analysis to assess the population structure of the Dutch cattle breeds. This analysis was conducted using the variance-standardized relationship matrix [32] with PLINK v1.9 [32]. We considered only autosomal and biallelic variants with an $r^2 < 0.5$ between variants within a window of 50 SNPs and with a genotyping rate > 0.95. The results from the PCA were visualized using the R package ggplot2 v3.3.5 [33].

Individual ancestry was evaluated by a model-based clustering method with the ADMIXTURE software v1.23 [34]. This method used the allele frequencies and the proportions of the ancestral populations in each sample to model the probability of the observed genotypes [34]. In the model, the K-value (optimal number of clusters) was estimated as the one with the lowest cross-validation error (CV) [34]. The ADMIXTURE algorithm was performed using values of K ranging between 2 and 6. The analysis was performed with a total of 120 unrelated animals from Dutch breeds and included 1,354,139 autosomal variants with a $r^2 < 0.5$ within windows of 50 variants over the genome and a minor allele frequency (MAF) > 0.05.

Genome-wide association study. A genome-wide association study was used to identify and characterize genome regions that differ between GWH and other breeds to find out candidate genes funtionally related with pigmentation and/or the eye protective phenotypes, e.g. ACOP. We used a mixed-model approach developed by Zhou and Stephens [35] in the Genome-wide Efficient Mixed-Model Association v0.98.1 [35] program. The mixed-model approach accounted for the population structure by including in the random effect the covariance structure from the genomic kinship matrix. In a first step, the association analysis was performed between GWH and non-white headed Dutch breeds (5 Holstein Friesian; 24 Dutch Friesian; 23 Meuse-Rhine-Yssel; 23 Dutch Belted; and 24 Deep Red). A total of 14,285,317 autosomal variants with a MAF > 0.05 were used to evaluate the relationship between each variant and the GWH breed phenotypes:

$$\boldsymbol{y} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{x}\delta + \mathbf{u} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ was the binary phenotype, one for the GWH individuals with ACOP and zero for Dutch Belted, Deep Red, Meuse-Rhine-Yssel, Dutch Friesian, and Holstein Friesian; $\mathbf{W}$ the matrix of incidence for the fixed effects; $\boldsymbol{\alpha}$ the intercept vector of ones; $\mathbf{x}$ contains the vector with SNP genotypes by sample; $\boldsymbol{\delta}$ represents the marker effect size; $\mathbf{u}$ contains a vector with the random genetic effects that follow a n-dimensional multivariate normal distribution with $\mathbf{u} \sim \text{MVN}_n\,(\mathbf{0}, \lambda\,\tau^{-1}\,\mathbf{K})$ for n individuals and being $\boldsymbol{\lambda}$ the ratio from two components of variance, τ^{-1} is the variance of the residual error, and $\mathbf{K}$ the kinship matrix derived from the genotypes from each sample; $\boldsymbol{\varepsilon} \sim \text{MVN}_n\,(\mathbf{0}, \tau^{-1}\,\text{I}_n)$ the vector containing the errors, with I representing the identity matrix. The nominal *p-values* from the association study were corrected using the false discovery rate (FDR) approach implemented in the R function p.adjust [36] and Benjamini & Hochberg [37] method. We considered those variants with a *q-value* (from the FDR test) lower than 0.001 as significantly associated. Here, a QTL and the co-localization between QTLs and significant selective sweeps were defined following the method reported by Gonzalez-Prendes et al. [38]. In brief, we considered only genomic regions with more than two significantly associated variants as candidate QTL. The co-localization between QTLs or between QTLs and selective sweeps was considered if the genomic regions overlapped by at least one base pair.

In a second step, variants from two additional breeds (25 animals from the Simmental breed and 25 from Hereford) with white heads and no ACOP were retrieved from the 1000 Bull Genomes Project (Run9 version) [24, 25] to perform the GWAS between these and GWH. We decided to keep the analysis with those two transboundary breeds separated from the remaining five Dutch breeds because we used different approaches to detect variants from whole genome resequencing data and we did not want to lose informative variants segregating in the populations at low frequency for subsequent analyses. The Simmental and Hereford sequence data, with a mean depth of coverage of 11.68 X (between 1.84 and 44.17) [24, 25], were merged with the data obtained from the 120 animals in our study, including 21 GWH individuals using PLINK v1.9 [32] with default parameters. The association study was performed with a total of 9,655,666 variants with a genotype call rate above 0.9, a MAF higher than 0.05 and using the model described above.

Identification and annotation of selective sweeps

The variants identified in each sample were used to explore the presence of genomic regions under selection in each breed with two complementary methods. First, Sweep Detector (SweeD) v3.0 [39] software, was applied using a composite likelihood ratio test to find candidate selective sweeps across the genome based on Site Frequency Spectrum patterns of

variations [40]. We defined a window size of 5 kb across the genome to calculate the Site Frequency Spectrum patterns, and the outlier regions falling within the top 1% of the composite likelihood-ratio test distribution were selected as significant regions. Second, a complementary approach based on linkage disequilibrium implemented in OmegaPlus v3.0.3 [41] was applied. Here, the *ω-statistic* is calculated based on patterns of linkage disequilibrium close to a recently fixed mutation. A high value of *ω-statistic* at a specific genomic location can indicate a genomic region under selection. In this method, we used the same window size of 5 kb bins across the genome and outlier regions with the highest values (top 1%) of *ω-statistic* were considered significant. Finally, only candidate selective sweeps within the 1% of the highest scores obtained by both methods were annotated using Ensembl 101 [42] database and used for subsequent analyses.

Runs of homozygosity identification in the GWH breed. The detection of ROH in the GWH breed was implemented with detectRUNS v0.9.6 [43] program. This analysis was used as a complementary method to confirm and reduce the number of candidate genomic regions that co-localize between the GWAS signals and selective sweeps. Genomic regions with ROH hotspots were selected to control the number of false positive candidate selective sweeps and GWAS signals by selecting only genomic regions that co-localize between them. A sliding window-based method was applied to detect regions with at least 15 variants in a run with 250 kb as the minimum length and a maximum distance between consecutive variants of one Mb. Additionally, we considered one variant per 10 kb as the lower density limit and only one missing or heterozygous variant per run. Potential ROH hotspots were identified by selecting only genomic regions containing the most frequent (top 1%) variants in a run in the GWH population [44–46].

Results and discussion

After the mapping of the Dutch cattle breeds and Holstein Friesian short read sequences to the bovine reference genome (assembly ARS-UCD1.2), the depth of coverage across samples, in average, was 8.7X ranging from 7X to 13X (**S1 Table**). The number of variants per breed, biallelic variants and variants that are specific to each breed are shown in **Table 1**. The overall number of annotated variants was 21,313,663, and the number of SNPs per animal (between 6 and 7 million, **S1 Table**) and per breed (between 13 and 17 million, **Table 1**) are within the range of that obtained in other studies on *B. taurus* [47–53]. The breed with the highest number of breed-specific variants was Deep Red (558,909), whereas the Holstein Friesian showed the lowest number (148,213). The low number of specific variants detected in Holstein Friesian compared with the remaining breeds in this study is most likely because of the small effective population size associated with a strong artificial selection pressure [54]. However, as the number of samples (n = 5) for Holstein Friesian is low, specific variants with low frequency may be

Table 1. Number of variants by breeds and breed-specific variants detected in 6 cattle populations.

Breeds	Mean genome coverage	Number of variants	Number of biallelic variants	Number of biallelic breed-specific variants
Holstein Friesian	9.20	13,218,695	12,376,751	148,213
Meuse-Rhine-Yssel	9.78	16,642,547	15,751,146	304,064
Groningen White Headed	8.19	15,554,352	14,675,130	368,389
Dutch Frisian	8.46	16,025,075	15,139,002	374,333
Dutch Belted	8.30	16,463,618	15,574,124	475,279
Deep Red	8.75	17,804,474	16,906,802	558,909
Mean	8.78	15,951,460	15,070,493	371,531

https://doi.org/10.1371/journal.pone.0276309.t001

underestimated and the results must be taken with caution. Functional annotation analysis revealed that the detected variants mapped to intronic (46.18%) or intergenic (42.61%) regions. Only, 1.1% (389,472 variants) mapped to exonic regions, of which 146,057 were missense and 212,473 were synonymous variants (S2 Table).

Genetic differentiation of the GWH breed

The genetic relationships between samples were evaluated using a PCA approach. As shown in Fig 1A, the distribution of the samples is in concordance with the breed histories and in line with previous results obtained for traditional Dutch populations [23]. While, Holstein Friesian occupied the central position, PC1 separated the dual-purpose breeds Meuse-Rhine-Yssel and Deep Red, which are genetically closely related [55], from all others. This is in agreement with the history of these two breeds where Deep Red originated from the Meuse-Rhine-Yssel in the east of the Noord-Brabant province following multiple generations of selection for coat color [55]. The PC2 separated the GWH from other breeds, providing further support for the genetic differentiation of this population. The model-based clustering analysis supported the PCA results. We used the information obtained from the PCA, which showed six different clusters, to run the model-based clustering analysis from $K = 2$ to $K = 6$, and the smallest CV error to estimate the best number of K ancestral populations. The results (Fig 2) supported the high differentiation of the GWH breed at $K = 3$ in an independent genetic cluster. The separation of Dutch Friesian and Dutch Belted breeds occurred at $K = 4$, and finally the Meuse-Rhine-Yssel and Deep Red formed two distinct clusters at $K = 5$, which had the smallest CV error (0.54), reflecting their close genetic relationship [55]. In this analysis, we included the Holstein Friesian breed, however, determining the extent of admixture in this breed requires further studies of a larger sample size [56]. In the admixture analysis, populations with a low number of samples are less likely to be assigned to their own ancestral cluster and as a consequence, they are depicted across multiple drifted groups [56].

With the separation of the GWH population from the non-white-headed breeds, we decided to investigate if this breed, with ACOP, is also isolated from white headed breeds

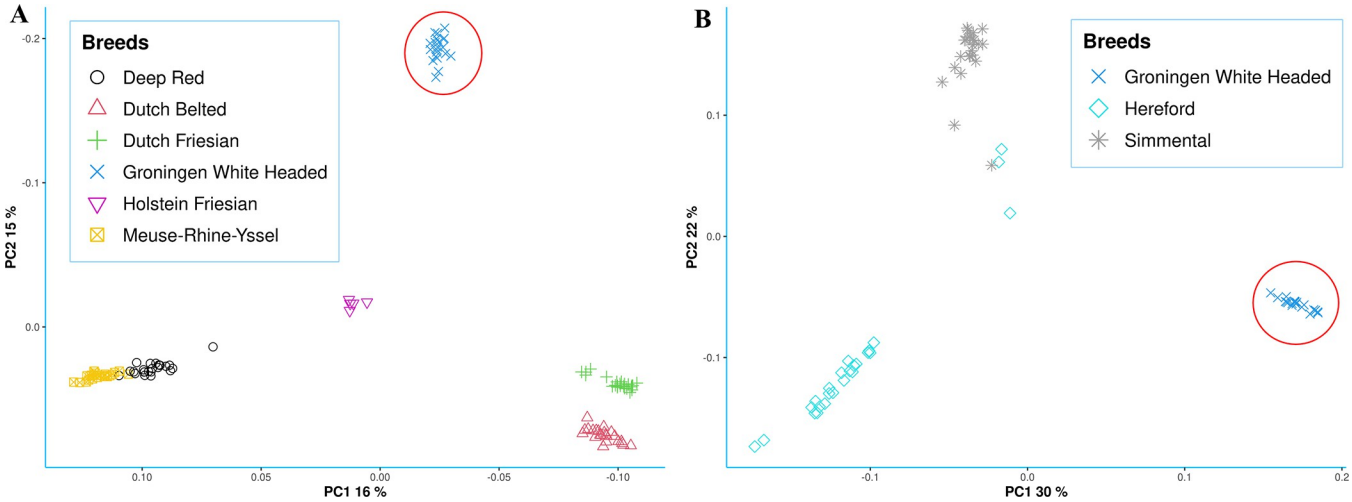

Fig 1. Principal component analysis (A) 120 samples from six breeds (Holstein Friesian, Dutch Friesian, Dutch Belted, Deep Red, Meuse-Rhine-Yssel and GWH); (B) The 70 animals from the three populations, GWH breed plus two white headed breeds (Hereford and Simmental) used for the GWAS in the second step. Individuals from the GWH breed (red circle) are distantly positioned from all other breeds in both plots. The % symbol indicates the percentage of the explained variance for the first and second components calculated from the eigenvalues.

https://doi.org/10.1371/journal.pone.0276309.g001

Fig 2. Population structure plot determined by the model-based clustering analysis of ADMIXTURE. Samples are represented by stacked columns of the 2 to 5 K-proportions and the number of clusters with the lowest cross validation error (CV = 0.54) was obtained for $K = 5$.

https://doi.org/10.1371/journal.pone.0276309.g002

without pigmentation around the eyes, that are, Hereford and Simmental (without ACOP). The PCA separated the breeds into three clusters based on their genomic information (Fig 1B). Animals represented in Fig 1B were used for the GWAS in the second step. The PC1, which explains 30% of the observed variation, divided the animals with and without ACOP and confirms the genetic differentiation of the GWH breed. The PC2, divided the Hereford and Simmental breeds into two clear clusters indicating two separate populations in accordance with previous reports [57]. This pattern, which confirms the GWH differentiation was also obtained when the five Dutch breeds and the three commercial populations (Holstein, Simmental, and Hereford) were combined (S1 Fig).

Genomic regions showing significant association with the GWH breed

The GWA study was used to identify and characterize genome regions that differ between GWH and other breeds to find candidate genes possibly associated with pigmentation and/or eye protective phenotypes e.g. ACOP, which is typical of GWH breed. Animals with ACOP (GWH) were classified as cases and animals of the Dutch Belted, Deep Red, Meuse-Rhine-Yssel, Dutch Friesian, and Holstein Friesian breeds were considered as controls. At the genome-wide level ($q\text{-}val <= 0.001$), 137 genome hits (S3 Table and Fig 3) with more than one significantly associated variant were detected. The associated regions were distributed across the 29 chromosomes (Fig 3) and the regions with the most significant associations ($p\text{-}value$ <-4.9E-14) and with the highest number of associated variants (>100 significant

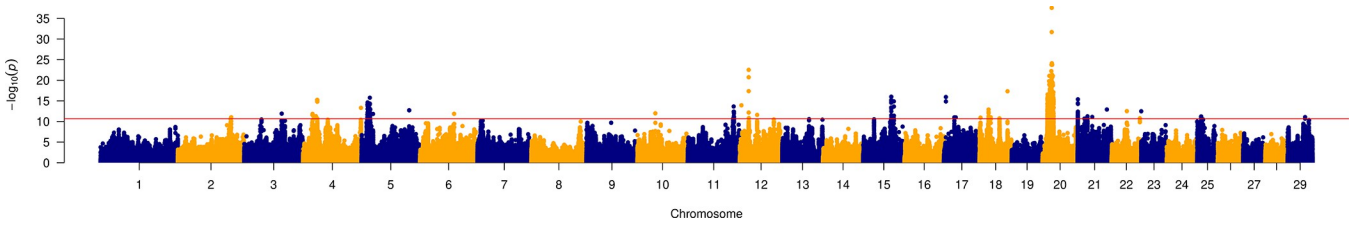

Fig 3. Manhattan plots showing the GWAS results from contrasting GWH animals with the ACOP phenotype and those of the Dutch Belted, Deep Red, Meuse-Rhine-Yssel, Dutch Friesian, and Holstein Friesian breeds without white head and non-ACOP phenotype. The y-axis of the plot represents the -$\log_{10}$ (*P*-values) from the GWAS and the x-axis shows the genomic location of each variant. The horizontal red line indicate the significant association (*q-value* $\leq$0.001) at the genome-wide level.

https://doi.org/10.1371/journal.pone.0276309.g003

associations) mapped to BTA4 (20.0–29.9 Mb and 116.8–118.8 Mb), BTA5 (10.1–19.7 Mb), BTA12 (12.0–18.5 Mb), BTA15 (50.6–59.8 Mb and 60.3–67.8 Mb), BTA20 (10.2–19.9 Mb and 20.2–29.5 Mb) and BTA21 (0.4–8.8 Mb). A total of four genomic regions co-localized with those detected by Pausch et al. [9] in Fleckvieh breed, which are, two regions located on BTA5 (10–19.7 Mb; 57.5–58.9 Mb), one on BTA13 (50.1–59.9 Mb) and one on BTA22 (30.4–32.5 Mb). The low coincidence between the studies may indicate that most associations are breed-specific suggesting that this phenotype may have a different genetic background in these breeds. However, multiple methodological and biological factors can influence these differences. Pausch et al. [9] used genomic information from a combination of SNP arrays (version 1 and 2 of Illumina BovineSNP 50K Bead chip®, and Illumina BovineHD Bead chip® 777k), whereas we used whole-genome sequence variants. Additionally, Pausch et al. [9] used a quantitative trait (a proportion of progeny with ACOP) in the GWAS study while in the current study we used the ACOP traits as a binary phenotype. Finally, while large sample sizes are needed for GWAS of complex traits, the sample size can be dramatically reduced for a case and control analysis in binary phenotypes [58, 59].

QTL detection in white headed cattle with and without ACOP. As there were no GWH animals with a completely white head and without ACOP, 50 animals from Hereford and Simmental breeds were selected from the 1000 Bull project [24]. These data were merged with variants from our GWH to carry out a GWAS analysis using a total of 15,751,624 variants to: 1) detect GWAS signals associated with the phenotype variation of GWH breed to find candidate genes related with pigmentation and/or eye protection phenotypes, e.g. ACOP, by contrasting breeds with ACOP (GWH) and without ACOP (Hereford and Simmental) and completely unpigmented area around the eyes; and 2) to reduce the number of candidate genomic regions by retrieving the QTLs overlapping with the GWAS (breeds without white head vs GWH). A total of 187 genomic significant hits with at least two significant SNPs were detected (**S4 Table**), and 100 (53.4%) co-localized with the QTLs identified when the six breeds were included in the analysis (**S4 Table**). This result may suggest that those regions specifically affect the GWH breed and may be associated with its color phenotype. Interestingly, the QTL on BTA5 (region, 10.1–19.7 Mb), was also identified by contrasting GWH vs breeds without white head (BTA5, region 10.0–13.7 Mb). Pausch et al. [9] reported the same QTL earlier at BTA5 (15.6–20.6 Mb, remapped to ARS-UCD1.2 assembly) which explained around 7.9% of the total phenotypic variation of ACOP in the Fleckvieh breed [9].

Our GWAS analyses were limited by the fact that significantly associated genomic regions can be observed due to the different genetic backgrounds between the breeds. This confounding effect should either be eliminated through a better study design (e.g. F2 crosses with another white face breed that does not show ACOP) [60–62] or by reducing the number of false positives using a combined approach in a downstream analysis [17, 63]. For example, the

application of complementary methods to investigate whether loci significantly associated were recently selected in the population [16], the description of functions of the genes in candidate regions, and finally the experimental validation. As we did not have animals from the GWH breed without ACOP we decided to investigate if our significantly associated genomic regions were recently selected in our GWH population to detect positional candidate genes functionally associated with pigmentation, eye disease, and/or UV protection.

Detection of a breed-specific selective sweeps in GWH

We used the whole genome resequencing data from six cattle breeds (GWH, Dutch Belted, Deep Red, Meuse-Rhine-Yssel, Dutch Friesian, and Holstein Friesian) to find out breed-specific selective sweeps (BSSS) in the GWH breed with two complementary methods: SweeD, which detects selective sweeps based on the variant frequencies using a composite maximum likelihood approach [39]; and OmegaPlus, that identifies patterns of linkage disequilibrium using the ω statistic [41]. Only significant genome regions (top 1% of the empirical distribution) in both algorithms (SweeD [39] and OmegaPlus [41]) were selected for furher analysis. With this approach, 257 breed-specific putative genomic regions under selection were detected (**Fig 4**, **S5 Table**). The candidate regions were distributed across the 29 autosomes (**Fig 4**) with sizes that ranged from 3.4 kb to 140.4 kb and a mean of 17.8 kb. The breed with the lowest number of candidate regions was GWH (31), followed by Meuse-Rhine-Yssel (40), Dutch belted (41), Dutch Friesian (46), Holstein Friesian (48), and Deep Red (51). The highly significant BSSS migth indicate "divergence signals" between breeds [3]. Thus, the BSSS might be an indicator of genomic regions affecting unique phenotypic characteristics for which the selection signal was detected [3] and therefore can be used to validate the GWAS signals for the phenotypic variation of the GWH breed. The regions with the most significant associations obtained by both methods were found on BTA5 (12 Mb) and BTA20 (14–20 Mb) in GWH; BTA3 (115–118 Mb) on Dutch Belted and BTA3 (12–13 Mb), BTA11 (93–94 Mb) and BTA22 (45–48 Mb) on Holstein Friesian (**Fig 4**). When we evaluated the co-localization between the BSSS (±500 kb up-and downstream) in GWH and QTLs detected by GWAS, eight genomic regions were also mapped with all methods (**S3**–**S5 Tables**) as follows: one on BTA4 (70.2–71.3 Mb), one on BTA5 (10.0–19.7 Mb), one on BTA10 (26.9–29.4 Mb), one on BTA13 (60.0–61.4 Mb), one on BTA15 (55.5–59.8 Mb), two on BTA20 (10.0–19.9, 20.0–22.7 Mb) and one on BTA25 (0.5–9.2 Mb).

We also evaluated if the candidate selective sweep co-localized with known bovine QTLs deposited in the AnimalQTLdb [64] database. A total of 4,558 different QTLs affecting 260 traits were found within 257 candidate BSSS (**S6 Table**). Several of the candidate selective sweeps highlighted loci which were mainly associated with milk quality, milk production, feed efficiency, body weight, and several meat-related phenotypes. To be noted, these results are in line with the economic objective established for the studied breeds; Dutch local cattle (Meuse-Rhine-Yssel, Deep Red, and GWH) have been selected for dual-purpose characteristics including milk production. Although our candidate selective sweeps were selected as unique in each breed, we still can find BSSS affecting the same trait. This can be explained by the fact that in livestock populations, including traditional cattle breeds, the selection for economically important traits, e.g. complex traits, might happen across many loci with small effects [2, 65]. The successful identification and characterization of those BSSS that are associated with economically relevant traits can be used to: 1) improve the knowledge about the processes influencing the genetic diversity of each breed; and 2) identify candidate genes and/or causal variants affecting phenotypes under selection. Thus, further studies are encouraged to explore the relationship between our candidate BSSS and the impact that they may have on economically relevant traits in detail as this was not an objective of the current study.

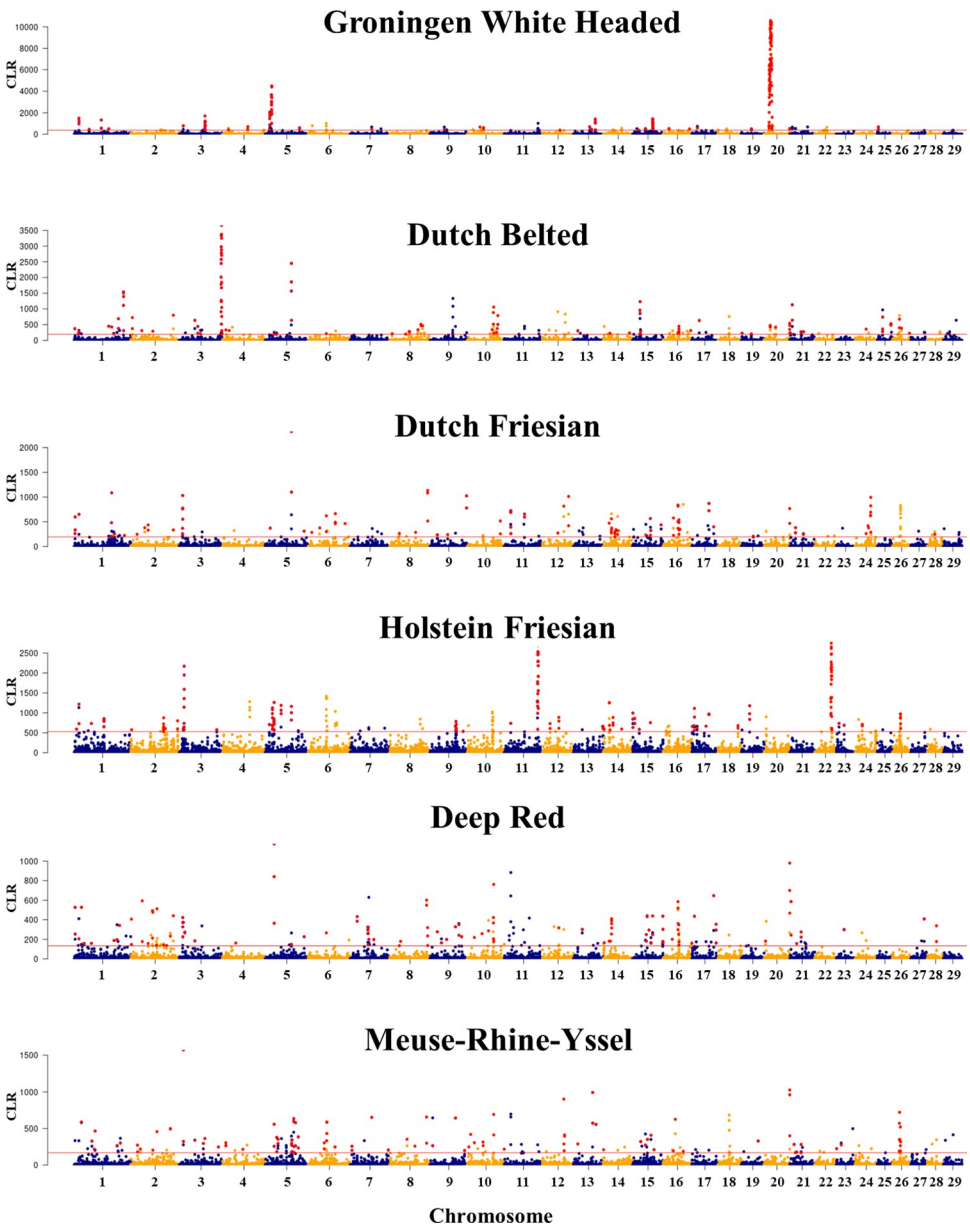

Fig 4. Genome-wide selective sweep scans using SweeD in each breed. Manhattan plots representing the composite likelihood ratio values (y-axis) from SweeD for each marker across the genome (x-axis). The threshold of the significant association (top 1% of the highest composite likelihood ratio values) for declaring candidate selective sweeps is indicated by the red line. Red points indicate candidate genomic regions detected by both the SweeD and OmegaPlus methods.

https://doi.org/10.1371/journal.pone.0276309.g004

Several ROH hotspots map to QTLs and putative selective sweeps

The identification of the genomic regions in ROH in the GWH breed was implemented as a complementary method to confirm and reduce the number of candidate genomic regions that co-localize between the GWAS signals and BSSS. We found 4,911 ROH regions that cover on average a total of 207.4 Mb of the genome. Of these ROH regions, around 73% (3,615) can be classified as small (0.5–1 Mb) regions, indicating more ancient consanguinity or population founder effects [66]. This result is common in cattle populations, where longer ROH regions

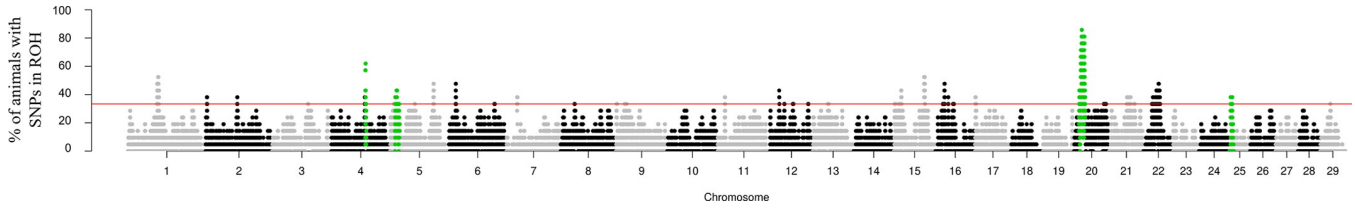

Fig 5. Genome-wide ROH hotspots disribution in GWH breed. The y-axis represents the percentage (%) of animals with SNPs in ROH regions and the x-axis the genomic coordinate of each variant. The significance threshold indicating the genomic regions (ROH hotspots) containing the variants present in more than 99% of a ROH region across the samples is indicated by the red line. Green dots represent genomic regions on BTA4 (70.2–71.3 Mb), BTA5 (10.0–19.7 Mb), BTA20 (10.0–19.9 Mb and 20.0–22.7 Mb), and BTA25 (0.5–9.2 Mb) that co-localize with significant SNPs commonly detected by the GWAS, selective sweeps and ROH hotspots.

https://doi.org/10.1371/journal.pone.0276309.g005

have been found less frequently than shorter ones [67]. To reduce the number of identified genomic regions in ROH, the ROH hotspots were defined by identifying genomic regions containing the variants with the highest frequency (top 1%) in a ROH across the GWH population (**Fig 5**, **S7** and **S8 Tables**). With this approach, 57 genomic regions were detected as ROH hotspots. With their genomic coordinates, we were able to reveal genomic regions that co-localize with the previously detected BSSS and GWAS signals. Five genomic regions that mapped to BTA4 (70.2–71.3 Mb), BTA5 (10.0–19.7 Mb), BTA20 (10.0–19.9 Mb and 20.0–22.7 Mb), and BTA25 (0.5–9.2 Mb) were overlapped between the three methods, and thus genes on those regions are probably under selection in the GWH breed [68, 69].

Positional and functional candidate genes associated with pigmentation and retinal diseases

We also investigated whether the function of the positional candidate genes are specifically associated with pigmentation and/or metabolism of melanocytes. First, we focused on genes that mapped to regions that overlapped between ROH hotspots, BSSS, and the GWAS signals (GWH vs other Dutch breeds, and GWH vs Hereford and Simmental breeds) (**Fig 5**). These regions included 141 genes (**S9 Table**), of which some are functional candidate genes. For example, on BTA 5 (12–17 Mb), the transmembrane o-mannosyltransferase targeting cadherins 2 (*TMTC2*) located at 12.2 Mb, is associated with calcium ion homeostasis [70]. Calcium homeostasis is of major importance in melanocytes and is suggested to be regulated by melanosomes [71]. The KIT Ligand (*KITLG*) locus (BTA 5, 18.2–18.3 Mb), which encodes a ligand for the receptor-type-tyrosine kinase KIT and contributes to coat color in various species, including cattle [72, 73]. On BTA20 (10.9–20 Mb), the region with the most significant SNPs contains the *DEPDC1B* (DEP domain-containing protein 1B) gene at position 18.5–18.6 Mb, which is associated with the hyperproliferation of abnormal melanocyte cells [74]. This gene is overexpressed in melanoma and encodes DEPDC1B protein that contains a DEP domain [75, 76], which plays an active role in controlling cell functions, including specific signal of retinal photoreceptor and cell polarity [76, 77].

Interestingly, there are two genes (**S7 Table**) in our list related with retinal diseases, for example *CWC27* (CWC27 Spliceosome Associated Cyclophilin) associated with Retinitis Pigmentosa [78]; on BTA25 (1.1–1.2 Mb) the function of the Clusterin Associated Protein 1 (*CLUAP1*) in the vertebrate eye is important for ciliogenesis and photoreceptor maintenance [79]. Although only few cases of eye degenerative diseases with a genetic background have been reported in cattle [80–82], recently Michot et al. [83] evidenced a group of mutations related with eye diseases that are segregating in European cattle breeds with direct impact on animal health e.g., the recessive frameshift mutation on *RP1* gene that causes loss of vision in cattle populations.

The most significant SNPs on BTA20 mapped to genes related with UV protection and melanocyte differentiation

The analysis of the whole genome resequencing data allowed to identify variants within candidate genomic regions that can help to clarify the cause of the phenotypic differences that exist between GWH and the remaining breeds. We investigated the genomic regions on BTA20 (10.0–19.9 Mb and 20.0–22.7 Mb) because those regions contained the most significant associations at three levels (GWAS, Fig 3; BSSS, Fig 4; and ROH, Fig 5). We studied the top ten significant SNPs in these regions to identify putatively associated genes. Nine of these SNPs mapped to four genes (*RAB3C*, *NDUFAF2*, *ZSWIM6*, and *PDE4D*; Table 2), and 11 of them to intergenic regions (Table 2). The linkage desequilibrium between those SNPs was high, ranging from $r^2 = 0.91$ to one (Table 2), and one of these SNPs (rs381052637, *p-value* = 8.64E-22) mapped to the 3′UTR of the *PDE4D* gene. SNPs located in 3′-UTR sequences may abolish or create a microRNA target and consequently may lead to different activities of the gene thereby contributing to interindividual variability [84, 85].

Four of the most significant SNPs (Table 2) mapped to the Phosphodiesterase 4D (*PDE4D*) gene. *PDE4D* is involved in the degradation of the Cyclic AMP. In humans, the skin pigment production and its protection against the UV radiation improved with the up-regulation of cAMP in melanocytes [86]. However, the function of PDE isoforms in pigmentation and melanocyte biology has not been extensively studied. Khaled et al. [87] reported that the up-regulation of *PDE4D* loci mediated by the MC1R-cAMP-MITF pathway led to a reduced melanocyte pigmentation in mice [88–90]. Interestingly, genes in the MITF pathway have been linked in many cattle breeds with coat color phenotypes [11, 91, 92], and also in other species [93]. As far as we know, there is no evident relationship between the Ubiquinone Oxidoreductase Complex Assembly Factor 2 (*NDUFAF2*) or Related Protein Rab-3C (*RAB3C*) genes with coat color or melanogenesis. However, the *RAB3C* gene is part of the Rab GTPases proteins, which were involved in cell membrane trafficking and associated with melanosomes [94]. Finally, another interesting candidate gene that maps to 40 kb downstream of the rs381810091 SNP (*p-value* = 1.74E-24) is the ERCC Excision Repair 8 (*ERCC8*) gene, involved in protein ubiquitination and UV response. In humans, the *ERCC8* gene is associated with Ultraviolet-sensitive syndrome [95] a genetic disorder characterized by cutaneous photosensitivity that causes differentiated skin pigmentation and greater freckling, without an increased risk of skin tumors [95, 96].

Table 2. Genomic localization of the most significant SNPs on BTA20.

GWAS results					Candidate Genes			
BTA	**Position (bp)**	**SNP ID**	**Localization**	***p-value***	**BTA**	**Gene Start**	**Gene End**	**Gene Symbol**
20	17,974,182	rs382263925	intronic	5.89E-23	20	17,842,584	18,052,859	*ZSWIM6*
	18,330,187	rs381810091	intronic	1.74E-24		18,210,359	18,370,943	*NDUFAF2*
	20,044,595	20:20044595	intronic	8.64E-22		20,014,955	20,315,593	*PDE4D*
	20,044,910	rs380360322	intronic	8.64E-22				
	20,278,747	20:20278747	intronic	8.64E-22				
	20,314,517	rs381052637	3 prime UTR	8.64E-22				
	20,524,538	20:20524538	intronic	8.64E-22		20,440,181	20,735,999	*RAB3C*
	20,542,032	20:20542032	intronic	8.64E-22				
	20,598,559	20:20598559	intronic	8.64E-22				

[1]BTA: *Bos taurus* chromosomes, Position (bp): position in base pair of SNP, SNP ID: Variant displaying the significant association with GWH breed, *p-value*: nominal *p-value*.

*Linkage disequilibrium based on the squared correlation (r^2) from genotypic allele counts was higher than 0.91 between presented SNPs.

https://doi.org/10.1371/journal.pone.0276309.t002

Conclusion

The used integrative approach based on the combined use of GWAS, selective sweep and ROH analyses identified several regions of the cattle genome (BTA4,70.2–71.3 Mb; BTA5,10.0–19.7 Mb; BTA20,10.0–19.9 Mb, and 20.0–22.7 Mb; and BTA25,0.5–9.2 Mb) as candidates to explain phenotype variation in the GWH breed. Importantly, those regions contained breed-specific genetic markers and candidate genes that are functionally related with pigmentation (e.g. *PDE4D*), UV protection (e.g. *ERCC8*), or retinal degeneration (e.g. *CWC27*, and *CLUAP1*). This finding contributes to characterizing the genetic background of the GWH breed and provides insights to further investigate the biological pathways and causative mutations influencing skin pigmentation and/or eye protective phenotypes e.g. Ambilateral Circumocular Pigmentation, and the biological implications of skin pigmentation for animal adaptation.

Supporting information

S1 Fig. Principal component analysis of the 170 cattle samples from five local Dutch Breeds (Dutch Belted, Dutch Friesian, Meuse-Rhine-Yssel, Deep Red, and GWH) and three commercial breeds Holstein Friesian, Hereford and Simmental. Individuals from the GWH breed (red circle) were distantly positioned from all other breeds.
(TIF)

S1 Table. Read coverage and number of variants by animals and breeds from five local Dutch Breeds (Dutch Belted, Dutch Friesian, Meuse-Rhine-Yssel, Deep Red, and GWH) and the commercial breed Holstein Friesian.
(XLSX)

S2 Table. Variant effect predicted from whole-genome variant from Dutch Belted, Dutch Friesian, Meuse-Rhine-Yssel, Deep Red, Holstein Friesian, and GWH breeds.
(XLSX)

S3 Table. Genome-wide significant QTL for GWH and five cattle breeds without white head (Dutch Belted, Deep Red, Meuse-Rhine-Yssel, Dutch Friesian and Holstein Friesian).
(XLSX)

S4 Table. Genome-wide significant QTL for phenotypic variation of GWH from contrasting Simmental, Hereford white headed breeds.
(XLSX)

S5 Table. Breed-specific selective sweep regions detected in Dutch Belted, Dutch Friesian, Meuse-Rhine-Yssel, Deep Red, GWH, and Holstein Friesian breeds.
(XLSX)

S6 Table. Quantitative trait locus, from the Animal QTL database, in breed-specific selective sweeps detected in Dutch Belted, Dutch Friesian, Meuse-Rhine-Yssel, Deep Red, Holstein Friesian and GWH breeds.
(XLSX)

S7 Table. Genomic distribution of SNPs in RHO hotspots in GWH breed.
(XLSX)

S8 Table. Genomic coordinates of RHO hotspots in GWH breed.
(XLSX)

S9 Table. Genes that mapped to genome regions on BTA4 (70.2–71.3 Mb), BTA5 (10.0–19.7 Mb), BTA20 (10.0–19.9 Mb, and 20.0–22.7 Mb) and BTA25 (0.5–9.2 Mb).
(XLSX)

Acknowledgments

The authors are indebted to The Groningen White Headed association for its collaboration and for providing the animal material. We gratefully acknowledge Bert Dibbits and Kimberley Laport from Animal Breeding and Genomics (WUR) for technical laboratory support. This work is part of the OPTIBOV project financed within the Long term European African research and innovation Partnership on food and nutrition security and sustainable Agriculture (LEAP-Agri) program (https://leap-agri.com). The authors gratefully acknowledge the Fundação Nacional para a Ciência e a Tecnologia (FCT), Portugal, contract grant 2020.02754. CEECIND (C.G.) and the National Research Foundation (NRF) of South Africa, Leap Agri-326, Grant number 115577.

Author Contributions

Conceptualization: Richard P. M. A. Crooijmans.

Data curation: Rayner Gonzalez-Prendes.

Formal analysis: Rayner Gonzalez-Prendes.

Funding acquisition: Richard P. M. A. Crooijmans.

Investigation: Rayner Gonzalez-Prendes.

Methodology: Rayner Gonzalez-Prendes, Catarina Ginja, Juha Kantanen.

Resources: Juha Kantanen, Richard P. M. A. Crooijmans.

Supervision: Catarina Ginja, Martien A. M. Groenen, Richard P. M. A. Crooijmans.

Writing – original draft: Rayner Gonzalez-Prendes, Richard P. M. A. Crooijmans.

Writing – review & editing: Rayner Gonzalez-Prendes, Catarina Ginja, Juha Kantanen, Nasser Ghanem, Donald R. Kugonza, Mahlako L. Makgahlela, Martien A. M. Groenen, Richard P. M. A. Crooijmans.

References

1. Ramey HR, Decker JE, McKay SD, Rolf MM, Schnabel RD, Taylor JF. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics. 2013; 14: 382. https://doi.org/10.1186/1471-2164-14-382 PMID: 23758707
2. Ghoreishifar SM, Eriksson S, Johansson AM, Khansefid M, Moghaddaszadeh-Ahrabi S, Parna N, et al. Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. Genet Sel Evol. 2020; 52. https://doi.org/10.1186/s12711-020-00571-5 PMID: 32887549
3. Gutiérrez-Gil B, Arranz JJ, Wiener P. An interpretive review of selective sweep studies in Bos taurus cattle populations: Identification of unique and shared selection signals across breeds. Frontiers in Genetics. Frontiers Media S.A.; 2015. p. 167. https://doi.org/10.3389/fgene.2015.00167 PMID: 26029239
4. Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, et al. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics. 2011; 12: 318. https://doi.org/10.1186/1471-2164-12-318 PMID: 21679429
5. Jablonski NG, Chaplin G. The evolution of human skin coloration. J Hum Evol. 2000; 39: 57–106. https://doi.org/10.1006/jhev.2000.0403 PMID: 10896812

6. Finch VA, Western D. Cattle colors in pastoral herds: natural selection or social preference? Ecology. 1977; 58: 1384–1392. https://doi.org/10.2307/1935090
7. Flori L, Moazami-Goudarzi K, Alary V, Araba A, Boujenane I, Boushaba N, et al. A genomic map of climate adaptation in Mediterranean cattle breeds. Mol Ecol. 2019; 28: 1009–1029. https://doi.org/10.1111/mec.15004 PMID: 30593690
8. Cieslak M, Reissmann M, Hofreiter M, Ludwig A. Colours of domestication. Biol Rev Camb Philos Soc; 2011. pp. 885–899. https://doi.org/10.1111/j.1469-185X.2011.00177.x PMID: 21443614
9. Pausch H, Wang X, Jung S, Krogmeier D, Edel C, Emmerling R, et al. Identification of QTL for UV-protective eye area pigmentation in cattle by progeny phenotyping and genome-wide association analysis. Moore S, editor. PLoS One. 2012; 7: e36346. https://doi.org/10.1371/journal.pone.0036346 PMID: 22567150
10. Senczuk G, Guerra L, Mastrangelo S, Campobasso C, Zoubeyda K, Imane M, et al. Fifteen shades of grey: Combined analysis of genome-wide snp data in steppe and mediterranean grey cattle sheds new light on the molecular basis of coat color. Genes (Basel). 2020; 11: 1–16. https://doi.org/10.3390/genes11080932 PMID: 32823527
11. Jivanji S, Worth G, Lopdell TJ, Yeates A, Couldrey C, Reynolds E, et al. Genome-wide association analysis reveals QTL and candidate mutations involved in white spotting in cattle. Genet Sel Evol. 2019; 51: 1–18. https://doi.org/10.1186/s12711-019-0506-2 PMID: 31703548
12. Mei C, Wang H, Liao Q, Wang L, Cheng G, Wang H, et al. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. Mol Biol Evol. 2018; 35: 688–699. https://doi.org/10.1093/molbev/msx322 PMID: 29294071
13. Andersson L. Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet; 2001. pp. 130–138. https://doi.org/10.1038/35052563 PMID: 11253052
14. Rees JL. Genetics of hair and skin color. Annu Rev Genet. 2003; 37: 67–90. https://doi.org/10.1146/annurev.genet.37.110801.143233 PMID: 14616056
15. Wang C, Li H, Guo Y, Huang J, Sun Y, Min J, et al. Donkey genomes provide new insights into domestication and selection for coat color. Nat Commun. 2020; 11: 1–15. https://doi.org/10.1038/s41467-020-19813-7 PMID: 33293529
16. Igoshin A V., Yurchenko AA, Belonogova NM, Petrovsky D V., Aitnazarov RB, Soloshenko VA, 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. 2019;20. https://doi.org/10.1186/s12863-019-0725-0 PMID: 30885142
17. Niu Q, Zhang T, Xu L, Wang T, Wang Z, Zhu B, et al. Integration of selection signatures and multi-trait GWAS reveals polygenic genetic architecture of carcass traits in beef cattle. Genomics. 2021; 113: 3325–3336. https://doi.org/10.1016/j.ygeno.2021.07.025 PMID: 34314829
18. CGN, 2009. Groningen White Headed breed assessment. EU GENRES 870/04 project EURECA. Centre for Genetic Resources, the Netherlands (CGN) of Wageningen UR. Available: https://edepot.wur.nl/5600
19. Blaarkopnet -. [cited 10 Dec 2020]. Available: https://zeldzamerassen.nl/blaarkopnet/
20. Tsujita H, Plummer CE. Bovine ocular squamous cell carcinoma. Vet Clin North Am Food Anim Pract; 2010. pp. 511–529. https://doi.org/10.1016/j.cvfa.2010.08.003 PMID: 21056799
21. Frisch JE. The relative incidence and effect of bovine infectious keratoconjunctivitis in Bos indicus and Bos taurus cattle. Anim Prod. 1975; 21: 265–274. https://doi.org/10.1017/S0003356100030737
22. Ward JK, Nielson MK. Pinkeye (bovine infectious keratoconjunctivitis) in beef cattle. J Anim Sci. 1979; 49: 361–366. https://doi.org/10.2527/jas1979.492361x PMID: 574506
23. Hulsegge I, Schoon M, Windig J, Neuteboom M, Hiemstra SJ, Schurink A. Development of a genetic tool for determining breed purity of cattle. Livest Sci. 2019; 223: 60–67. https://doi.org/10.1016/J.LIVSCI.2019.03.002
24. Hayes BJ, Daetwyler HD. 1000 Bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. Annual Review of Animal Biosciences. 2019. pp. 89–102. https://doi.org/10.1146/annurev-animal-020518-115024 PMID: 30508490
25. Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014; 46: 858–865. https://doi.org/10.1038/ng.3034 PMID: 25017103
26. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34: i884–i890. https://doi.org/10.1093/bioinformatics/bty560 PMID: 30423086
27. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26: 589–595. https://doi.org/10.1093/bioinformatics/btp698 PMID: 20080505

28. Modernizing the bovine reference genome assembly | WCGALP Archive. [cited 3 May 2022]. Available: http://www.wcgalp.org/proceedings/2018/modernizing-bovine-reference-genome-assembly

29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

30. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012 [cited 11 Jan 2021]. Available: http://arxiv.org/abs/1207.3907

31. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. https://doi.org/10.1086/519795 PMID: 17701901

33. Villanueva RAM, Chen ZJ. ggplot2: Elegant graphics for data analysis (2nd ed.). Meas Interdiscip Res Perspect. 2019; 17: 160–167. https://doi.org/10.1080/15366367.2019.1565254

34. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009; 19: 1655–1664. https://doi.org/10.1101/gr.094052.109 PMID: 19648217

35. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012; 44: 821–824. https://doi.org/10.1038/ng.2310 PMID: 22706312

36. The R Foundation. R: The R Project for Statistical Computing. 2018 [cited 16 Nov 2021]. Available: https://www.r-project.org/

37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995; 57: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

38. González-Prendes R, Quintanilla R, Mármol-Sánchez E, Pena RN, Ballester M, Cardoso TF, et al. Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine gluteus medius and longissimus dorsi muscles. BMC Genomics. 2019;20. https://doi.org/10.1186/s12864-019-5557-9 PMID: 30832586

39. Pavlidis P, Živković D, Stamatakis A, Alachiotis N. SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Mol Biol Evol. 2013; 30: 2224–2234. https://doi.org/10.1093/molbev/mst112 PMID: 23777627

40. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Res. 2005; 15: 1566–1575. https://doi.org/10.1101/gr.4252305 PMID: 16251466

41. Alachiotis N, Stamatakis A, Pavlidis P. OmegaPlus: A scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics. 2012; 28: 2274–2275. https://doi.org/10.1093/bioinformatics/bts419 PMID: 22760304

42. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020; 48: D682–D688. https://doi.org/10.1093/nar/gkz966 PMID: 31691826

43. Biscarini F, Cozzi P, Gaspa G, Marra G. detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. [cited 21 Oct 2021]. Available: https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html

44. Noce A, Qanbari S, González-Prendes R, Brenmoehl J, Luigi-Sierra MG, Theerkorn M, et al. Genetic diversity of bubalus bubalis in germany and global relations of its genetic background. Front Genet. 2021; 11. https://doi.org/10.3389/fgene.2020.610353 PMID: 33552127

45. Mastrangelo S, Sardina MT, Tolone M, Di Gerlando R, Sutera AM, Fontanesi L, et al. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. Animal. 2018; 12: 2480–2488. https://doi.org/10.1017/S1751731118000629 PMID: 29576040

46. Xu L, Zhao G, Yang L, Zhu B, Chen Y, Zhang L, et al. Genomic patterns of homozygosity in chinese local cattle. Sci Reports 2019 91. 2019; 9: 1–11. https://doi.org/10.1038/s41598-019-53274-3 PMID: 31740716

47. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, et al. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science (80). 2009; 324: 528–532. https://doi.org/10.1126/science.1167936 PMID: 19390050

48. Ramírez-Ayala LC, Rocha D, Ramos-Onsins SE, Leno-Colorado J, Charles M, Bouchez O, et al. Whole-genome sequencing reveals insights into the adaptation of French Charolais cattle to Cuban tropical conditions. Genet Sel Evol. 2021; 53: 1–11. https://doi.org/10.1186/s12711-020-00597-9 PMID: 33397281

49. Xia X, Zhang S, Zhang H, Zhang Z, Chen N, Li Z, et al. Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. BMC Genomics. 2021; 22. https://doi.org/10.1186/s12864-020-07340-0 PMID: 33421990

50. Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun. 2018; 9: 1–13. https://doi.org/10.1038/s41467-018-04737-0 PMID: 29904051

51. Kõks S, Reimann E, Lilleoja R, Lättekivi F, Salumets A, Reemann P, et al. Sequencing and annotated analysis of full genome of Holstein breed bull. Mamm Genome. 2014; 25: 363–373. https://doi.org/10.1007/s00335-014-9511-5 PMID: 24770584

52. Stafuzza NB, Zerlotini A, Lobo FP, Yamagishi MEB, Chud TCS, Caetano AR, et al. Single nucleotide variants and InDels identified from whole-genome re-sequencing of Guzerat, Gyr, Girolando and Holstein cattle breeds. PLoS One. 2017; 12. https://doi.org/10.1371/journal.pone.0173954 PMID: 28323836

53. Das A, Panitz F, Gregersen VR, Bendixen C, Holm LE. Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes. BMC Genomics. 2015; 16: 1043. https://doi.org/10.1186/s12864-015-2249-y PMID: 26645365

54. Doekes HP, Veerkamp RF, Bijma P, Hiemstra SJ, Windig JJ. Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein-Friesian breeding program from 1986 to 2015. Genet Sel Evol. 2018; 50: 15. https://doi.org/10.1186/s12711-018-0385-y PMID: 29642838

55. CGN, 2009. Deep Red Cattle breed assessment. EU GENRES 870/04 project EURECA. Centre for Genetic Resources, the Netherlands (CGN) of Wageningen UR. Available: https://edepot.wur.nl/5599.

56. Lawson DJ, van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commun. 2018; 9: 1–11. https://doi.org/10.1038/s41467-018-05257-7 PMID: 30108219

57. Kelleher MM, Berry DP, Kearney JF, McParland S, Buckley F, Purfield DC. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. Animal. 2017; 11: 15–23. https://doi.org/10.1017/S1751731116001099 PMID: 27330040

58. Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, et al. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat Genet. 2008; 40: 449–454. https://doi.org/10.1038/ng.96 PMID: 18344998

59. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics. 2009. pp. 381–391. https://doi.org/10.1038/nrg2575 PMID: 19448663

60. Vilhjálmsson BJ, Nordborg M. The nature of confounding in genome-wide association studies. Nat Rev Genet. 2013; 14: 1–2. https://doi.org/10.1038/nrg3382 PMID: 23165185

61. Flint J, Eskin E. Genome-wide association studies in mice. Nat Rev Genet. 2012; 13: 807–817. https://doi.org/10.1038/nrg3335 PMID: 23044826

62. Mkize N, Maiwashe A, Dzama K, Dube B, Mapholi N. Suitability of gwas as a tool to discover snps associated with tick resistance in cattle: A review. Pathogens. 2021; 10: 1604. https://doi.org/10.3390/pathogens10121604 PMID: 34959558

63. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015; 24: 4348–4370. https://doi.org/10.1111/mec.13322 PMID: 26184487

64. Hu ZL, Fritz ER, Reecy JM. AnimalQTLdb: A livestock QTL database tool set for positional QTL information mining and beyond. Nucleic Acids Res. 2007;35. https://doi.org/10.1093/nar/gkl946 PMID: 17135205

65. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. BMC Genomics. 2014; 15. https://doi.org/10.1186/1471-2164-15-246 PMID: 24678841

66. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. BMC Genomics. 2011; 12: 1–15. https://doi.org/10.1186/1471-2164-12-460 PMID: 21943305

67. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet. 2015; 46: 110–121. https://doi.org/10.1111/age.12259 PMID: 25530322

68. Pryce JE, Haile-Mariam M, Goddard ME, Hayes BJ. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genet Sel Evol. 2014; 46. https://doi.org/10.1186/s12711-014-0071-7 PMID: 25407532

69. Kim ES, Sonstegard TS, Van Tassell CP, Wiggans G, Rothschild MF. The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. PLoS One. 2015; 10. https://doi.org/10.1371/journal.pone.0129967 PMID: 26154171

70. Sunryd JC, Cheon B, Graham JB, Giorda KM, Fissore RA, Hebert DN. TMTC1 and TMTC2 are novel endoplasmic reticulum tetratricopeptide repeat-containing adapter proteins involved in calcium homeostasis. J Biol Chem. 2014; 289: 16085–16099. https://doi.org/10.1074/jbc.M114.554071 PMID: 24764305

71. Bush WD, Simon JD. Quantification of Ca2+ binding to melanin supports the hypothesis that melanosomes serve a functional role in regulating calcium homeostasis. Pigment Cell Res. 2007; 20: 134–139. https://doi.org/10.1111/j.1600-0749.2007.00362.x PMID: 17371440

72. Weich K, Affolter V, York D, Rebhun R, Grahn R, Kallenberg A, et al. Pigment intensity in dogs is associated with a copy number variant upstream of *KITLG*. Genes (Basel). 2020; 11. https://doi.org/10.3390/genes11010075 PMID: 31936656

73. Talenti A, Bertolini F, Williams J, Moaeen-Ud-Din M, Frattini S, Coizet B, et al. Genomic analysis suggests *KITLG* is responsible for a roan pattern in two pakistani goat breeds. J Hered. 2018; 109: 315–319. https://doi.org/10.1093/jhered/esx093 PMID: 29099936

74. Xu Y, Sun W, Zheng B, Liu X, Luo Z, Kong Y, et al. *DEPDC1B* knockdown inhibits the development of malignant melanoma through suppressing cell proliferation and inducing cell apoptosis. Exp Cell Res. 2019; 379: 48–54. https://doi.org/10.1016/j.yexcr.2019.03.021 PMID: 30880030

75. Ballon DR, Flanary PL, Gladue DP, Konopka JB, Dohlman HG, Thorner J. DEP-Domain-mediated regulation of GPCR signaling responses. Cell. 2006; 126: 1079–1093. https://doi.org/10.1016/j.cell.2006.07.030 PMID: 16990133

76. Sokol S. A role for *WNTS* in morphogenesis and tissue polarity. Nature Cell Biology. 2000. pp. E124–E125. https://doi.org/10.1038/35017136 PMID: 10878822

77. Consonni S V., Gloerich M, Spanjaard E, Bos JL. cAMP regulates DEP domain-mediated binding of the guanine nucleotide exchange factor Epac1 to phosphatidic acid at the plasma membrane. Proc Natl Acad Sci U S A. 2012; 109: 3814–3819. https://doi.org/10.1073/pnas.1117599109 PMID: 22343288

78. Xu M, Xie Y (Angela), Abouzeid H, Gordon CT, Fiorentino A, Sun Z, et al. Mutations in the spliceosome component CWC27 cause retinal degeneration with or without additional developmental anomalies. Am J Hum Genet. 2017; 100: 592–604. https://doi.org/10.1016/j.ajhg.2017.02.008 PMID: 28285769

79. Lee C, Wallingford JB, Gross JM. *CLUAP1* is essential for ciliogenesis and photoreceptor maintenance in the vertebrate eye. Investig Ophthalmol Vis Sci. 2014; 55: 4585–4592. https://doi.org/10.1167/iovs.14-14888 PMID: 24970261

80. Murgiano L, Jagannathan V, Calderoni V, Joechler M, Gentile A, Drögemüller C. Looking the cow in the eye: Deletion in the *NID1* gene is associated with recessive inherited cataract in romagnola cattle. PLoS One. 2014;9. https://doi.org/10.1371/journal.pone.0110628 PMID: 25347398

81. Bradley R, Terlecki S, Clegg FG. The pathology of a retinal degeneration in Friesian cows. J Comp Pathol. 1982; 92: 69–83. https://doi.org/10.1016/0021-9975(82)90043-3 PMID: 7068955

82. Stehman SM, William DVM, Rebhun C, Ronald DVM, Riis C. Progressive retinal atrophy in related cattle. Bov Pract. 1987. https://doi.org/10.21423/bovine-vol0no22p195-197

83. Michot P, Chahory S, Marete A, Grohs C, Dagios D, Donzel E, et al. A reverse genetic approach identifies an ancestral frameshift mutation in *RP1* causing recessive progressive retinal degeneration in European cattle breeds. Genet Sel Evol. 2016; 48: 56. https://doi.org/10.1186/s12711-016-0232-y PMID: 27510606

84. Hughes TA. Regulation of gene expression by alternative untranslated regions. Trends in Genetics. Elsevier. 2006. pp. 119–122. https://doi.org/10.1016/j.tig.2006.01.001

85. Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A. 2007; 104: 3300–3305. https://doi.org/10.1073/pnas.0611347104 PMID: 17360642

86. D'Orazio JA, Nobuhisa T, Cui R, Arya M, Spry M, Wakamatsu K, et al. Topical drug rescue strategy and skin protection based on the role of Mc1r in UV-induced tanning. Nature. 2006; 443: 340–344. https://doi.org/10.1038/nature05098 PMID: 16988713

87. Khaled M, Levy C, Fisher DE. Control of melanocyte differentiation by a MITF-PDE4D3 homeostatic circuit. Genes Dev. 2010; 24: 2276–2281. https://doi.org/10.1101/gad.1937710 PMID: 20952536

88. Bang J, Zippin JH. Cyclic adenosine monophosphate (cAMP) signaling in melanocyte pigmentation and melanomagenesis. Pigm Cell Melanoma R. 2021. pp. 28–43. https://doi.org/10.1111/pcmr.12920 PMID: 32777162

89. Delyon J, Servy A, Laugier F, André J, Ortonne N, Battistella M, et al. *PDE4D* promotes FAK-mediated cell invasion in BRAF-mutated melanoma. Oncogene. 2017; 36: 3252–3262. https://doi.org/10.1038/onc.2016.469 PMID: 28092671

90. Bogunovic D, O'Neill DW, Belitskaya-Levy I, Vacic V, Yu Y Lo, Adams S, et al. Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival. Proc Natl Acad Sci U S A. 2009; 106: 20429–20434. https://doi.org/10.1073/pnas.0905139106 PMID: 19915147

91. Liu L, Harris B, Keehan M, Zhang Y. Genome scan for the degree of white spotting in dairy cattle. Anim Genet. 2009; 40: 975–977. https://doi.org/10.1111/j.1365-2052.2009.01936.x PMID: 19531114

92. Fontanesi L, Scotti E, Russo V. Haplotype variability in the bovine *MITF* gene and association with piebaldism in Holstein and Simmental cattle breeds. Anim Genet. 2012; 43: 250–256. https://doi.org/10.1111/j.1365-2052.2011.02242.x PMID: 22486495

93. Baxter LL, Hou L, Loftus SK, Pavan WJ. Spotlight on spotted mice: A review of white spotting mouse mutants and associated human pigmentation disorders. Pigment Cell Res. 2004; 17: 215–224. https://doi.org/10.1111/j.1600-0749.2004.00147.x PMID: 15140066

94. Fukuda M. Rab GTPases: Key players in melanosome biogenesis, transport, and transfer. Pigment Cell Melanoma Res. 2021; 34: 222–235. https://doi.org/10.1111/pcmr.12931 PMID: 32997883

95. Nardo T, Oneda R, Spivak G, Vaz B, Mortier L, Thomas P, et al. A UV-sensitive syndrome patient with a specific *CSA* mutation reveals separable roles for *CSA* in response to UV and oxidative DNA damage. Proc Natl Acad Sci U S A. 2009; 106: 6209–6214. https://doi.org/10.1073/pnas.0902113106 PMID: 19329487

96. Laugel V, Dalloz C, Durand M, Sauvanaud F, Kristensen U, Vincent MC, et al. Mutation update for the CSB/ERCC6 and CSA/ERCC8 genes involved in Cockayne syndrome. Hum Mutat. 2010; 31: 113–126. https://doi.org/10.1002/humu.21154 PMID: 19894250