

# Structural Variations Associated with Adaptation and Coat Color in Qinghai-Tibetan Plateau Cattle

Xiaoting Xia, Fuwen Wang, Xiaoyu Luo, Shuang Li, Yang Lyu, Yining Zheng, Zhijie Ma, Kaixing Qu, Rende Song, Jianyong Liu, Jicai Zhang, Basang Wangdui, Basang Zhuzha, Suolang Quji, Li Zhao, Silang Wangmu, Ciren Luobu, Nima Cangjue, Danzeng Luosang, Suolang Sizhu, Haijian Cheng, Ruizhe Li, Zhipeng Wu, Ruihua Dang, Yongzhen Huang, Xianyong Lan, Luohao Xu, Haifei Hu, WaiYee Low, Zhuqing Zheng, Yu Wang, Yuanpeng Gao, Lu Deng, Johannes A. Lenstra, Jianlin Han, Xueyi Yang, Wenfa Lyu,\* Bizhi Huang,\* Chuzhao Lei,\* and Ningbo Chen\*

Structural variations (SVs) play crucial roles in the evolutionary adaptation of domesticated animals to natural and human-controlled environments, but SVs have not been explored in Tibetan cattle, which recently migrated and rapidly adapted to the high altitudes of the Qinghai-Tibetan Plateau (QTP). In this study, a *de novo* chromosome-level genome assembly for Tibetan cattle is constructed. It is found that using a lineage-specific reference genome significantly increased variant detection accuracy and completeness. Analysis of long-read sequencing data from 36 high-altitude QTP and 48 low-altitude cattle identified 222 528 SVs and 259 SV hotspot regions. Positively selected SVs in high-altitude cattle are related to energy metabolism erythropoiesis and angiogenesis, and peroxisomal metabolism. A 102-bp intronic deletion in *GNPAT* likely upregulated its expression. It is distinguished 7293 SVs that may be introgressed from yak, including variants upstream of the hypoxia-inducing gene *EGLN1*. Finally, a  $\approx$ 2-Mb heterozygous inversion and two translocations on chromosome 6 are likely associated with the cattle gray coat via regulatory effects on the *KIT* gene. The results confirm the importance of SVs in evolutionary adaptation and the contribution yak-introgressed SVs to the rapid acclimatization of QTP cattle.

high-intensity UV radiation, and large fluctuations in temperature. The geographical range of the QTP includes the high mountain and plateau region in southwestern China (Xizang Autonomous Region, most of Qinghai, parts of Sichuan, Gansu, and Yunnan Province) and extends to neighboring countries. Several studies have been carried out to understand the adaptations of yak<sup>[1]</sup> dog,<sup>[2,3]</sup> horse,<sup>[4]</sup> pig,<sup>[5]</sup> and chicken<sup>[6]</sup> to high-altitude environments. Interestingly, selection signatures suggest diverse evolution pathways underlying their adaptive selection. For example, *EPAS1*, a key gene involved in hypoxia adaptation, has undergone convergent selection in Tibetan human, horse, dog, and gray wolf.<sup>[2,4,7-9]</sup> However, in yak and chicken, adaptation to high altitudes is achieved through the selection of other genes (e.g., *ADAM17* and *RYR2*).<sup>[1,6]</sup>

The domestication of *Bos taurus* (taurine cattle) and *B. indicus* (indicine cattle) dates back more than 10 000 and 8000 years, respectively.<sup>[10,11]</sup> During the history and

development of human society, cattle have spread and adapted to a vast variety of environmental conditions and have also experienced admixture with other bovine species.<sup>[12-16]</sup> QTP cattle include mainly Tibetan taurine and indicine cattle. Ancient DNA evidence indicates that taurine cattle arrived on the QTP at least 3900 years ago and represent the earliest domesticated cattle

## 1. Introduction

The Qinghai-Tibetan Plateau (QTP) is one of the most extreme environments on Earth, with an average altitude of 4268 m above sea level. Species that inhabit the QTP must adapt to harsh conditions, including low oxygen concentrations,

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.202503258>

© 2025 The Author(s). Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202503258

X. Xia, F. Wang, X. Luo, S. Li, Y. Zheng, R. Dang, Y. Huang, X. Lan, Y. Wang, L. Deng, J. Han, C. Lei, N. Chen  
Key Laboratory of Animal Genetics  
Breeding and Reproduction of Shaanxi Province  
College of Animal Science and Technology  
Northwest A&F University  
Yangling 712100, China  
E-mail: [leichuzhao1118@nwfau.edu.cn](mailto:leichuzhao1118@nwfau.edu.cn); [ningbochen@nwfau.edu.cn](mailto:ningbochen@nwfau.edu.cn)

entering East Asia.<sup>[17]</sup> Tibetan cattle live at altitudes of 3000–4600 m and retain distinct morphological and other characteristics,<sup>[13]</sup> such as small body size, thick coat, low metabolic rate, and strong adaptability to high altitudes. High-altitude indicine cattle include mainly the Shigatse Humped cattle breed in Xizang Autonomous Region ( $\approx 4000$  m), whereas molecular data show a clear indicine influence on the humpless Diqing cattle in Yunnan Province ( $\approx 3460$  m).<sup>[18,19]</sup>

Structural variations (SVs) are believed to have a broader influence on gene regulation and protein functionality than single nucleotide polymorphisms (SNPs) and shorter insertions or deletions (InDels) and are therefore also more important for understanding certain aspects of the adaptation process.<sup>[20–24]</sup> Previous studies have used SNPs to identify selection signals or detect adaptative introgression from yak into Tibetan taurine cattle: this revealed several candidate genes related to the hypoxia-inducible transcription factor pathway, including

*EGLN1*.<sup>[13,14,19]</sup> A recent study utilizing large-scale genomic data revealed selected genes related to body size (*HMGA2* and *NCAPG*) and energy metabolism (*DUOX2*) in Tibetan cattle.<sup>[19]</sup> However, research on the adaptation and phenotypic associations of QTP cattle through SVs has not yet been conducted.

Here, we generated a high-quality *de novo* genome assembly of Tibetan cattle based on the Pacific Biosciences high-fidelity (HiFi) long-read and high-throughput chromosome conformation capture (Hi-C) data. We detected the SVs associated with high-altitude adaptation by collecting long-read sequencing (LRS) data from 84 cattle across various altitudes, including 36 cattle on the QTP at altitudes above 3000 m, specifically from the Xizang Autonomous Region, Qinghai, and Yunnan Province in China (Figure 1a). To expand the sample size, these SVs were further genotyped in a larger panel of short-read sequencing (SRS) data of 281 cattle and 12 individuals from other bovine species. Combined with epigenomic analysis, we highlight the unique contributions of SVs to high-altitude adaptation and gray coat color in QTP cattle.

## 2. Results

### 2.1. Phylogenomic Relationships and *De Novo* Genome Assembly of QTP Cattle

To describe the population genetic structure of the QTP cattle populations, we generated an SNP dataset using 84 LRS and

F. Wang, Z. Ma  
State Key Laboratory of Plateau Ecology and Agriculture  
Qinghai University  
Xining 810016, China

Y. Lyu, W. Lyu  
Key Laboratory of Animal Production  
Product Quality  
and Security  
Ministry of Education  
College of Animal Science and Technology  
Jilin Agricultural University  
Changchun 130118, China  
E-mail: lwfenfa@jlau.edu.cn

Z. Ma, R. Li, Z. Wu  
Qinghai Academy of Animal Science and Veterinary Medicine  
Qinghai University  
Xining 810016, China

K. Qu  
Academy of Science and Technology  
Chuxiong Normal University  
Chuxiong 675099, China

R. Song  
Animal Disease Prevention and Control Centre of Yushu Tibetan Autonomous Prefecture  
Yushu 810099, China

J. Liu, J. Zhang, B. Huang  
Yunnan Academy of Grassland and Animal Science  
Kunming 650500, China  
E-mail: hbz@ynbp.cn

B. Wangdui, B. Zhuzha, S. Quji, L. Zhao, S. Wangmu, C. Luobu, N. Cangjue, D. Luosang  
Institute of Animal Husbandry and Veterinary Science  
Tibet Academy of Agricultural and Animal Husbandry Sciences  
Lhasa 850030, China

S. Sizhu  
Animal Science College  
Tibet Agriculture and Animal Husbandry University  
Nyingchi 860000, China

H. Cheng  
Institute of Animal Science and Veterinary Medicine  
Shandong Academy of Agricultural Sciences  
Shandong Key Lab of Animal Disease Control and Breeding  
Jinan 271018, China

L. Xu  
Key Laboratory of Aquatic Science of Chongqing  
College of Life Sciences  
Southwest University  
Chongqing 400715, China

H. Hu  
Rice Research Institute  
Guangdong Key Laboratory of New Technology in Rice Breeding and  
Guangdong Rice Engineering Laboratory  
Guangdong Academy of Agricultural Sciences  
Guangzhou 510640, China

W. Low  
The Davies Research Centre  
School of Animal and Veterinary Sciences  
University of Adelaide  
Roseworthy, SA 5371, Australia

W. Low  
College of Computing and Information Technology  
University of Doha for Science and Technology  
Doha 00974, Qatar

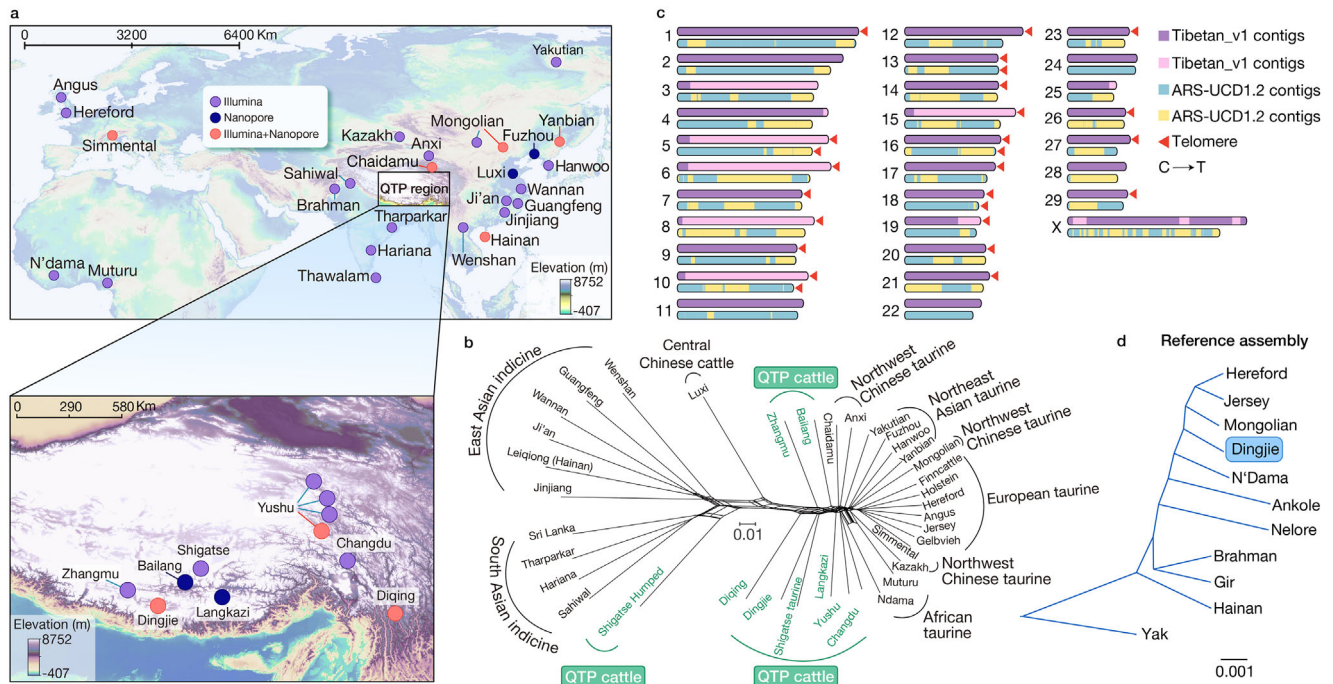
Z. Zheng  
College of Food and Biology  
Jingchu University of Technology  
Jingmen, Hubei 448000, China

Y. Gao  
College of Veterinary Medicine  
Northwest A&F University  
Yangling 712100, China

J. A. Lenstra  
Faculty of Veterinary Medicine  
Utrecht University  
Utrecht, NL 7758, The Netherlands

J. Han, W. Lyu  
Yazhouwan National Laboratory  
Sanya 572024, China

X. Yang  
Life Science College  
Luoyang Normal University  
Luoyang 471934, China



**Figure 1.** Geographic distribution, phylogenetic relationships, and genome assembly of Qinghai-Tibetan Plateau cattle. a) Geographic locations of the cattle breeds used in this study and their sequencing platforms. b) NeighborNet graph of 37 cattle populations. An allele frequency-dependent distance metric (Reynolds) was used to construct NeighborNet. c) Chromosome diagrams for the ARS-UCD1.2 (bottom) and Tibetan\_v1 (top) assemblies for autosomes from the centromere (left) to the telomere (right); the X chromosome has two telomeres. The positions of the telomeres are marked with orange triangles. d) A phylogenetic tree of MASH distances derived from 11 bovine assemblies, including the taurine cattle reference genome. The yak assembly was used as the outgroup.

281 SRS samples, which includes data from 148 QTP cattle (36 LRS and 112 SRS data) encompassing Changdu, Dingjie, Langkazi, Diqing, Yushu, Zhangmu, and Shigatse populations living at different altitudes (Figure 1a). A phylogenetic network (Figure 1b) revealed that QTP taurine cattle are closely related to Northeast Asian taurine and Northwest Chinese taurine cattle, whereas QTP indicine cattle (Shigatse Humped) are close to South Asian indicine cattle. The Diqing cattle exhibited mixed taurine-indicine ancestry. These results were confirmed by phylogenetic tree of individuals (Figure S1a, Supporting Information) and principal component analysis (PCA, Figure S1b, Supporting Information).

We selected a male Tibetan cattle (Dingjie DJ12) living at an altitude of 4610 m to generate a Tibetan cattle reference genome sequence (Tibetan\_v1). Assembly of 102.0 Gb of HiFi data by the Hifiasm program<sup>[25]</sup> (Table S1 and Figure S2, Supporting Information) resulted in 709 contigs with an N50 of 84.6 Mb. After the incorporation of 188.5 Gb Hi-C data, we obtained a genome of 2.92 Gbp, consisting of 31 chromosomes in 60 scaffolds, with a scaffold N50 of 110.9 Mb. The chromosomes of Tibetan\_v1 are up to 24.50 Mb longer than those of the taurine cattle reference genome (ARS-UCD1.2) (Figure 1c and Table S2, Supporting Information), which, for a large part (49%) was due to additional repetitive regions. We detected telomere sequences (TTAGGG)<sub>n</sub> on 21 autosomes of Tibetan\_v1 (ARS-UCD1.2: 5 autosomes) (Figure 1c). A phylogenetic tree based on the Mash distances<sup>[26]</sup> between genomes positioned Dingjie (Tibetan\_v1)

between the Mongolian (Eurasian taurine) and N'Dama (African taurine) genomes<sup>[13]</sup> (Figure 1d).

An alignment of the Tibetan\_v1 scaffolds against ARS-UCD1.2 revealed high collinearity with the Tibetan\_v1 genome (Figure S3, Supporting Information). SVs are nonrandomly distributed throughout the genome but are highly prevalent in the terminal regions of chromosomes (Figure S4, Supporting Information). Benchmarking Universal Single-Copy Orthologs (BUSCO) evaluation revealed that 93.4% of the complete single-copy genes in Tibetan\_v1 were assembled out of the 9226 orthologous single-copy genes (Table S3, Supporting Information). A K-mer evaluation of the quality by the Merqury program<sup>[27]</sup> gave a consensus quality score (QV) of Tibetan\_v1 of 52.67, which was clearly higher than that of the other references (31.1 to 35.9), confirming that the Tibetan\_v1 assembly is more complete than other cattle reference genomes (Table 1). Interspersed repeats accounted for 44.25% of Tibetan\_v1, among which long interspersed nuclear elements (LINEs) constituted the greatest proportion (26.15%), whereas short interspersed nuclear elements (SINEs) accounted for 10.88% (Table S4, Supporting Information).

## 2.2. Application of Lineage-specific Assembly in Population Genetic Analysis

We aligned the SRS data of Tibetan\_v1 to each of the four assemblies (ARS-UCD1.2, Tibetan\_v1, Hainan\_v1, and Mongolian\_v1).

**Table 1.** Comparison of the Tibetan\_v1 *de novo* assembly with representative cattle assemblies.

Statistics	Tibetan_v1	Hainan_v1	Mongolian_v1	ARS-UCD1.2
Genome size (bp)	2919463950	2651892370	2638491090	2715853794
Contig N50 (bp)	84618144	44352259	47790958	25896116
Scaffold N50 (bp)	110861251	104224000	104053164	103308737
Number of scaffolds	60	301	400	2211
Max scaffold (bp)	163509520	157081543	146429245	158534110
Min scaffold (bp)	21316	749	1000	1034
Scaffold mean (bp)	48657732	8810273	6596227	1228337
Consensus quality score (QV)	52.7	35.9	39.9	31.1
K-mer based error rate	$5.4 \times 10^{-6}$	$2.6 \times 10^{-4}$	$1.0 \times 10^{-4}$	$7.8 \times 10^{-4}$
K-mer completeness (%)	94.4	91.7	93.9	96.7

As expected, Tibetan\_v1 had the highest mapping rate (Table S5, Supporting Information). We then identified small variants by aligning the SRS data of 27 Dingjie cattle to ARS-UCD1.2 and Tibetan\_v1 reference genomes. Alignments to Tibetan\_v1 resulted in more singleton and fixed variants than alignment to ARS-UCD1.2 (Figures S5a,b, Supporting Information), indicating that lineage-specific assembly facilitates the discovery of rare variants. The use of a lineage-specific reference could convert “multiallelic” into “biallelic” loci in a target lineage.<sup>[28]</sup> We found that from the multiallelic variants identified in the 27 Dingjie cattle against the ARS-UCD1.2, 5.58–8.41% (708–1417 per genome) of them were converted into biallelic variants by alignment to Tibetan\_v1 (Table S6, Supporting Information), resulting in 11928 biallelic variants linked with 5056 genes. Among them, 265 SNPs were found in exons of 181 genes. These genes were significantly enriched in pathways related to immunity (corrected *P* value < 0.05), which might be important for the adaptation of Tibetan cattle to high altitudes<sup>[13]</sup> (Table S7 and Figure S5c, Supporting Information). This example indicates that calling for variants based on non-lineage-specific references may miss important variants. However, a PCA of 119 East Asian taurine cattle on the basis of SNP genotypes (Figure S5d, Supporting Information) was not sensitive to the use of ARS-UCD1.2 or Tibetan\_v1 as references. This was also found for the calculation of effective population size (*N<sub>e</sub>*) via a multiple sequentially Markovian coalescent (MSMC),<sup>[29]</sup> which for the Tibetan cattle population has remained relatively stable over the past 10 000 years (Figure S5e, Supporting Information). Therefore, this study confirms that for East Asian taurine cattle, the use of lineage-specific references can improve variant calling but has a minimal effect on optimizing the evaluation of population genetic structure and effective population size.

### 2.3. SV Detection in QTP Cattle Based on LRS Data

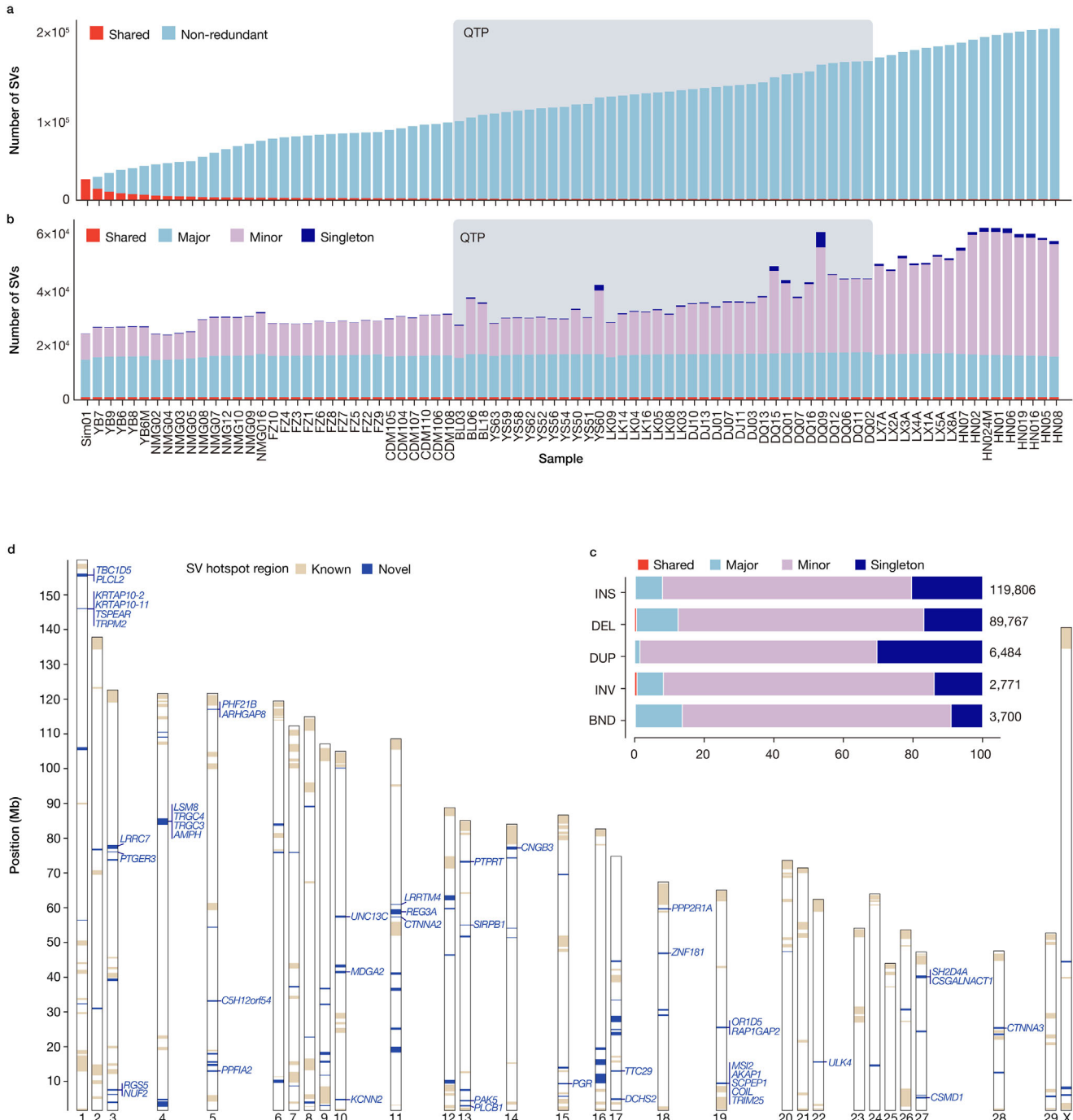
We performed LRS of 36 QTP cattle sampled from six locations of Yushu (*n* = 10), Langkazi (*n* = 7), Diqing (*n* = 10), Dingjie (*n* = 6), and Bailang (*n* = 3) (Figure 1a; Table S8, Supporting Information). We also generated 28 new LRS genomes and retrieved 20 published LRS genomes of other cattle breeds of Chaidamu (*n* = 6), Mongolian (*n* = 10), Yanbian (*n* = 5), Simmental (*n* = 1), Hainan (*n* = 9), Fuzhou (*n* = 10), and Luxi (*n* = 7). To

keep the coordinates of the genome consistent with previous reports, ARS-UCD1.2 was used as the reference genome in this study. SVs > 50 bp relative to ARS-UCD1.2 were detected and retained if they were supported by at least two of the following tools: CuteSV,<sup>[30]</sup> SVIM,<sup>[31]</sup> and Sniffles<sup>[32]</sup> (Figure S6, Supporting Information). On average, 36300 SVs per sample were identified and merged into 222528 nonredundant SVs (Figure 2a,b), including 89767 deletions (DELs), 6484 duplications (DUPs), 119 806 insertions (INSs), 2771 inversions (INVs), and 3700 complex SVs (break-ends, BNDs) (Figure 2c). Annotation using the ANNOVAR program<sup>[33]</sup> revealed that most of these SVs were located in intergenic and intron regions, with exonic SVs accounting for only 2.2%. Among the 222 528 SVs, 46 028 (20.68%) SVs exhibited homozygous reference genotypes (0/0) across all examined low-altitude cattle (Yanbian, Simmental, Luxi, Fuzhou, and Hainan), including 16 229 DELs, 27 112 INSs, 1679 DUPs, 458 INVs, and 550 BNDs.

We classified these SVs into four categories: shared (present in all samples), major (present in ≥ 50% of the samples), minor (present in more than one but < 50% of the samples), and singleton (present in only one sample). Among the QTP cattle, Diqing had the most singletons (4.93%), which probably reflects an admixture of East Asian indicine cattle (Figure 1b) or other bovine species (e.g., yak).<sup>[19]</sup> The number of nonredundant SVs initially increased steeply but began to plateau as more samples were added (Figure 2a), suggesting that the 36 cattle captured a substantial proportion of common SVs within the QTP population. The number of nonredundant SVs increases with the addition of taurindicine (Diqing and Luxi) and indicine (Hainan) cattle, which suggests that these are indicine-specific variants. Although the rate of novel SV discovery gradually slowed with increasing sample size, the continued contribution of non-redundant SVs—particularly from genetically distinct groups—implies that SV diversity in cattle may not yet be fully saturated. We detected 795 SVs in all samples (Figure 2b), which probably corresponded to singletons or errors in the cattle reference genome.

### 2.4. Characterization of Genomic SVs

Transposable elements (TEs) are important sources of SVs in cattle.<sup>[34–36]</sup> A total of 21147 TEs were identified across INS, DEL, and DUP, accounting for 9.50% of all the SVs. The TEs in-



**Figure 2.** Discovery of structural variations (SVs) in 84 cattle genomes. a) SVs from each sample were merged via a nonredundant strategy starting with Simmental (Sim01) and iteratively adding unique calls from additional samples. The growth rate of the nonredundant set decreases as the number of samples increases. Shared SVs are shown in red color. b) Number of SVs per sample that are classified into shared (identified in all samples), major (in  $\geq 42$  samples), minor (in  $> 1$  and  $< 42$  samples), and singleton (in only one sample) SVs. c) Proportions of SVs in the four categories defined in (b). d) Chromosomal distribution of SV hotspot regions and 47 SV-lined genes in newly identified hotspot regions.

cluded 20904 retrotransposons (17126 LINES, 2257 SINEs, and 1521 LTRs) and 243 DNA transposons (Table S9, Supporting Information). The enrichment of LINES has been reported previously for sheep<sup>[37]</sup> and pigs.<sup>[38]</sup> The SVs displayed sharp peaks at  $\approx 300$ , 1300, and 8000 bp corresponding to BOV-A2 (SINE),

BTLTR1/BTLTR1B (LTR/ERV), and L1\_BT/LIMA9 (LINE/L1), respectively (Figure S7, Supporting Information).

We identified 259 SV hotspots (see Materials and Methods) with a total size of  $\approx 281$  Mb, 157 of which have previously been identified<sup>[34,39,40]</sup> (Figure 2d). Random sampling (1000) and a Z

test revealed that the SV hotspots are significantly enriched in the TE-derived SVs, with a major contribution of DEL ( $Z$  score = 19.67,  $P = 1.81 \times 10^{-86}$ ) (Table S10, Supporting Information). The same analysis revealed that hotspots are not enriched in conserved element regions found by the genomic evolutionary rate profiling (GERP) ( $Z$  score =  $-5.76$ ,  $P = 4.14 \times 10^{-9}$ ). These findings were consistent with previous reports, which showed that SV hotspots are present mainly in genomic regions rich in genes and high in diversity within populations.<sup>[34,41]</sup> We observed that many newly identified SV hotspot regions, such as those on *B. taurus* autosome 4 (BTA4, *TRGC3* and *TRGC4*), BTA13 (*SIRPB1*), and BTA19 (*TRIM25* and *RAP1GAP2*), are related to immunity (Figure 2d). Notably, the genes *TRGC3* and *TRGC4* on BTA4 belong to the TRG family, which is one of the gene T-cell receptor subgroups.<sup>[42]</sup>

## 2.5. Identification of High-Altitude-Adapted SVs

To identify candidate SVs that may play a role in high-altitude adaptation, we analyzed SVs that have extremely different allele frequencies in high-altitude QTP cattle relative to low-altitude cattle using the  $d_i$  statistic.<sup>[43,44]</sup> We defined outlier SVs as  $d_i$ -SVs based on the  $Z$  test ( $P < 0.001$ ) (Figure 3a). Thus, we identified 1692  $d_i$ -SVs, including 906 INSS, 763 DELs, 16 DUPs, and 10 INV, involving 601 genes (Table S11, Supporting Information). Functional enrichment analyses of the 601 genes via Kyoto Encyclopedia of Genes and Genomes (KEGG) revealed significant enrichment in pathways (corrected  $P < 0.05$ ) related to cardiac health, metabolic regulation, and neuronal function, including arrhythmogenic right ventricular cardiomyopathy, metabolic pathways, amino sugar and nucleotide sugar metabolism, and the calcium signaling pathway (Figure S8 and Table S12, Supporting Information).

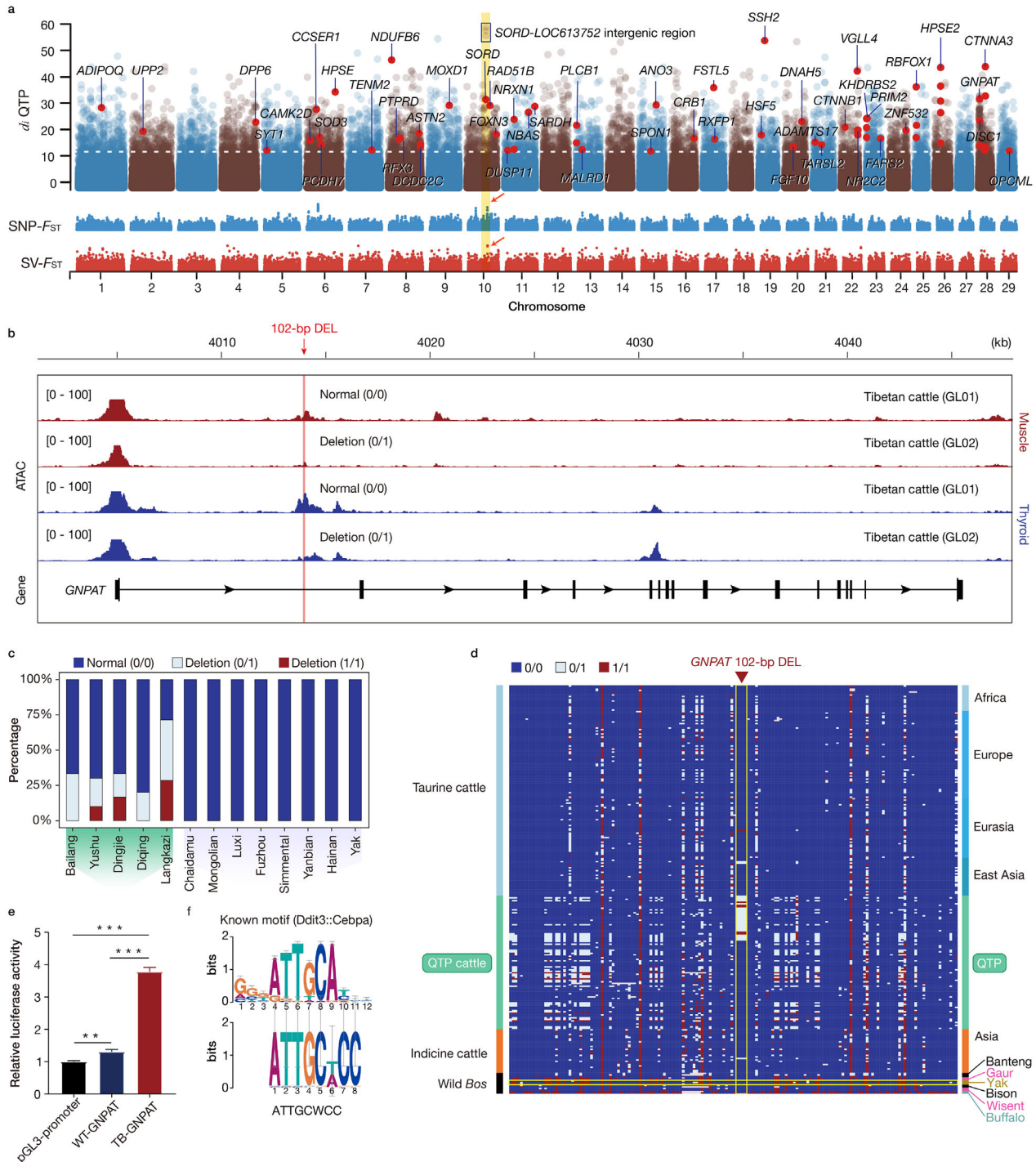
Among the 601 genes, 45 genes have been previously reported with known functions in hypoxic responses and/or as gene candidates associated with physiological traits,<sup>[45]</sup> encompassing 78 SVs (Figure 3a; Table S13, Supporting Information). Several of these SVs are located in genes involved in erythropoiesis and angiogenesis, such as *SSH2*,<sup>[46]</sup> *VGLL4*,<sup>[47]</sup> *PLCB1*,<sup>[48]</sup> *HPSE2*,<sup>[49]</sup> and *HPSE*<sup>[50]</sup> (Figure 3a). Notably, the 67-bp INS (BTA19:20937384) in *SSH2* exhibited the highest  $d_i$  signal among these genes. As a member of the slingshot phosphatase family, *SSH2* may play an important role in cofilin-mediated vascular smooth muscle cell migration and neointima formation.<sup>[46]</sup> *HPSE* and *PLCB1* are associated with the adaptation of Tibetan<sup>[51]</sup> and alpine Merino sheep<sup>[52]</sup> to high-altitude hypoxia, respectively. In addition, we detected an 8493-bp INS (BTA22:13818542) in *CTNNB1*, with an allelic difference of 0.211 (SV- $F_{ST}$ ) between QTP and low-altitude cattle. *CTNNB1* knock-down inhibited hypoxia-regulated gene expression in vitro and reduced the protein level of hypoxia-inducible factor-1 $\alpha$  (HIF-1 $\alpha$ ).<sup>[53]</sup>

We further detected several important candidate SV-related genes associated with lipid metabolism and mitochondrial metabolism, which are crucial for oxygen homeostasis.<sup>[54]</sup> The strongest signals were observed on BTA10, including four consecutive SVs of 74-bp DEL (BTA10:65366980-65367054), 71-bp INS (BTA10:65418831), 601-bp INS (BTA10:65422143), and

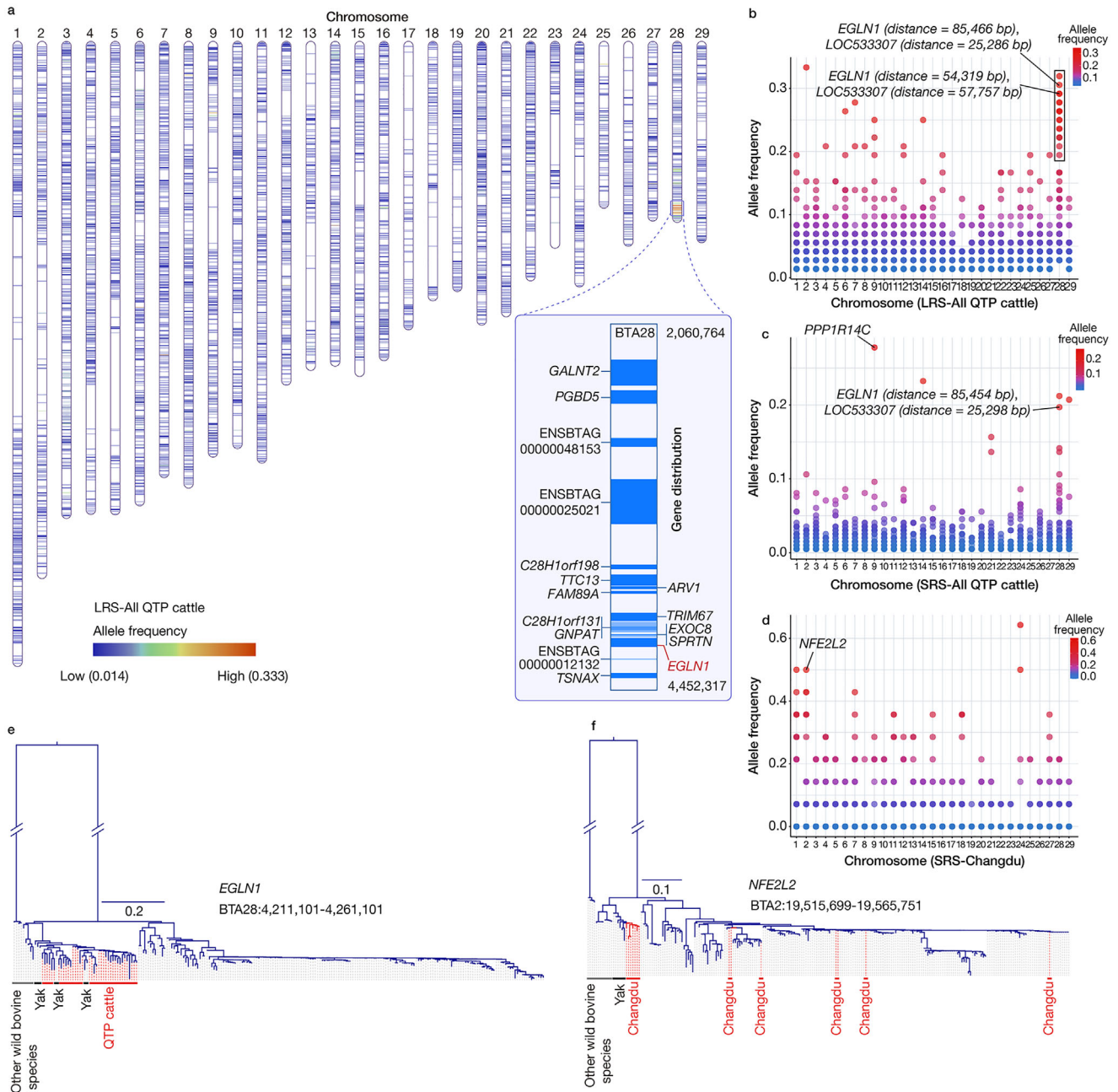
1056-bp DEL (BTA10:65424402-65425458) in the intergenic and intronic regions of the sorbitol dehydrogenase (*SORD*) gene (Figure 3a; Figure S9, Supporting Information). The allele frequencies of these SVs were as high as 35.7 to 83.3% in QTP cattle (Figure S9, Supporting Information). These SVs shared the same selection signatures of their linked SNPs (Figure 3a). *SORD* is the second enzyme of the polyol pathway. Low expression of *SORD* is expected to attenuate the increase in the cytosolic NADH/NAD<sup>+</sup> ratio and increase glycolysis.<sup>[55]</sup> Another highly  $d_i$ -SV is an 86-bp DEL in the intron of *NDUFB6* (BTA8:11510584-11510670). This is a crucial component of respiratory complex I, which, together with complex III generates most of the mitochondrial ROS.<sup>[56,57]</sup> A third top signal is a 52-bp DEL (BTA11:104554345-104554397) within the intron of *SARDH*. This gene encodes an enzyme localized to the mitochondrial matrix and plays an important role in promoting energy balance in the context of nutrient stress through the CREB/CRTC pathway.<sup>[58]</sup> We also detected a 65-bp INS (BTA1:80427709) located upstream of *ADIPOQ*, which is involved in adipose metabolism, body homeostasis, and cardiovascular protection.<sup>[59]</sup>

To explore the potential regulatory roles of the 1692  $d_i$ -SVs under selection, we integrated ATAC-seq data from six key tissues (adipose, liver, heart, lung, muscle, and thyroid) from high- and low-altitude cattle to identify chromatin accessible regions (CARs) (Table S14, Supporting Information). By intersecting with CARs, we identified 59  $d_i$ -SVs (3.5%) and 75 SV-associated genes (Table S15, Supporting Information). Notably, three genes (*GNPAT*, *CAT*, and *ACOT2*) were significantly enriched in the GO pathway of the peroxisomal matrix. *GNPAT* encodes glyceronephosphate O-acyltransferase, which is located in the peroxisomal membrane and catalyzes the first step of ether lipid synthesis involved in the decomposition of hydrogen peroxide and the maintenance of reactive oxygen species homeostasis.<sup>[60,61]</sup> We identified a 102-bp DEL in the first intron of *GNPAT* (BTA28:4013916-4014018) (Figure 3a,b), which had the highest signal among these three SVs. This 102-bp DEL was found in 13 out of the 36 LRS-based (36%, Figure 3c) and 36 out of the 112 SRS-based genomes (32%, Figure 3d; Figure S10, Supporting Information) of QTP cattle, with a very low detection rate in other cattle breeds. A previous study indicated that peroxisomal genes play crucial roles under hypoxic conditions.<sup>[54]</sup> ATAC-seq data from different tissues of QTP cattle revealed that the 102-bp DEL falls within an open chromatin region. Chromatin accessibility at this locus was reduced in the heterozygous GL02 individual (genotype 0/1) compared to the wild-type GL01 (genotype 0/0) (Figure 3b; Figure S11, Supporting Information).

Compared with the normal taurine haplotype, the QTP cattle-specific haplotype carrying the 102-bp DEL exhibited greater activity in a luciferase assay ( $P < 0.001$ ) (Figure 3e). This finding indicates that the DEL may disrupts the structure of the open chromatin and subsequently affect *GNPAT* expression. Within the 102-bp sequence, we identified a putative CCAAT/enhancer-binding protein (C/EBP)-binding motif (Ddit3::Cebpa), which controls lineage-specific gene expression and acts as a master regulator of the terminal differentiation of several cell types<sup>[62]</sup> (Figure 3f). Notably, the *GNPAT* gene region was found to be introgressed from yak to QTP cattle.<sup>[19]</sup> However, yak carried this 102-bp fragment (Figure 3d), suggesting that this 102-bp DEL was not derived from yak introgression.



**Figure 3.** Differentiated structural variations (SVs) associated with the high-altitude adaptation of Qinghai-Tibetan Plateau cattle. a) Genome-wide distribution of  $d_i$  for each SV across all autosomes between high- and low-altitude cattle (top panel). The blue block in the Manhattan plot for SNP- $F_{ST}$  and SV- $F_{ST}$  (bottom panel) indicates the consistency of *SORD* region selection for SNPs and SVs. b) The red and blue tracks represent the ATAC-seq data from muscle and thyroid tissues, respectively, of Tibetan cattle (GL01 and GL02). The pink line shows the 102-bp DEL (BTA28:4013916-4014018) in *GNPAT*, located within chromatin accessibility signals. c) Genotype frequency of the 102-bp DEL in *GNPAT* of the 96 LRS-based genomes, including data from 84 cattle and 12 yak. d) SNP and SV genotypes in the *GNPAT* region. e) Comparison of the relative luciferase activity between the wild-type (WT) and Tibetan (TB) cattle haplotypes of *GNPAT*. The bars display the means  $\pm$  SD ( $n = 3$  technical replicates). A two-sided t-test was used, where \*\* indicates  $P < 0.01$  and \*\*\* indicates  $P < 0.001$ . f) The top panel shows the known binding motif of the transcription factor Ddit3::Cebpa, while the bottom panel represents the sequence motif identified within the 102-bp DEL, which matches the binding site of Ddit3::Cebpa. This alignment indicates that the 102-bp DEL contains the Ddit3::Cebpa binding motif, suggesting that this DEL may affect the function of the regulatory element.



**Figure 4.** Genome-wide introgression from yak into Qinghai-Tibetan Plateau (QTP) cattle. a) The distribution of yak introgressed fragments in the genome based on LRS-SV results, as well as the SV hotspot region on BTA28. b–d), Allele frequencies of the SVs resulting from yak introgression b) in the LRS-based genomes of all QTP cattle (LRS-All QTP cattle); c) in the SRS-based genomes of all QTP cattle (SRS-All QTP cattle); and d) in the SRS-based genomes of Changdu cattle (SRS-Changdu). e) Phylogenetic analysis of the SNPs in the introgressed *EGN1* gene. f) Phylogenetic analysis of the SNPs in the introgressed *NFE2L2* gene.

## 2.6. Yak Introgression

Previous studies have reported that historic yak introgression events have contributed to the rapid adaptation of QTP cattle.<sup>[13,14,63]</sup> We used LRS-based genomes of 12 yak to identify the SVs that may have resulted from the introgression of yak into QTP cattle (Table S8, Supporting Information). We then searched for SVs that were shared by yak genomes but not present in other

cattle genomes. Using LRS-based genomes, 7293 autosomal SVs possibly derived from yak introgression were identified in QTP cattle, totaling 3.38 Mb (Figure 4a). To further validate these regions, we constructed a neighbour-joining (NJ) tree using SNPs from 50-kb flanking regions of the SVs in QTP cattle and yak, tracing the haplotype origins. In this way, we identified 5642 SVs in QTP cattle as probable yak introgressions (Table S17, Supporting Information). We found an introgressed SV hotspot region

on BTA28, including two SVs (1324-bp DEL in BTA28:4236113-4237437 and 570-bp INS in BTA28:4204966) located upstream of the well-known hypoxia-inducing gene *EGLN1*.<sup>[64]</sup> The two SVs had high frequencies in QTP cattle (Figure 4a-c,e). Another high-signal SV was a 57-bp DEL (BTA9:87336841-87336898) located in the first intron of *PPP1R14C*, a gene associated with angiogenesis.<sup>[65]</sup>

To expand the population size, the SVs were further genotyped using 293 SRS-based genomes, from 112 QTP cattle, 169 cattle from various global populations, and 12 individuals from other bovine species with high-coverage genome sequences (Table S16, Supporting Information). Applying the same identification criteria used for the LRS data, we detected 2334 SVs possibly derived from yak introgression (Table S18, Supporting Information), of which 971 showed overlap with the LRS SVs. Consistent with a recent study using SNP data,<sup>[19]</sup> our results indicate that not all QTP cattle populations have the same yak introgression signals. For example, the top candidate regions in the Changdu cattle population differ from those in other QTP cattle populations (Figure 4d). Interestingly, an introgressed region on BTA2, including a 52-bp DEL in the intron of *NFE2L2* (BTA2:19540699-19540751), has a high frequency in the Changdu population and is absent in other cattle (Figure 4f). *NFE2L2* (NFE2-like bZIP transcription factor 2, also known as *NRF2*) plays a key role in the response to oxidative stress and induces the upregulation of glucose transporter 1 (*GLUT1*), which allows increased glucose uptake and HIF-1-driven augmentation of glycolysis.<sup>[66]</sup>

## 2.7. A 2-Mb Heterozygous Inversion Related to the Gray Coat Phenotype of QTP Cattle

We further examined QTP-specific SVs and found a 2-Mb heterozygous INV and two translocations (TRAs) related to the gray coat phenotype in cattle. The gray phenotype of mammals occurs when at birth the hairs lose their pigment and turn white. This is observed in gray horses,<sup>[67]</sup> but also with a low frequency in QTP cattle. To confirm the genomic candidate region of the gray coat phenotype, SRS data from 23 gray cattle and 258 non-gray cattle were used for SNP-based GWAS (SNP-GWAS) and SNP-based  $F_{ST}$  (SNP- $F_{ST}$ ) analyses (Figure 5a). Both SNP-GWAS, which uses the genome-wide efficient mixed model association (GEMMA) program with a mixed linear model, and SNP- $F_{ST}$  confirmed a strong signal on BTA6. The GWAS signal spans 2.99 Mb (BTA6:69765842-72752916;  $P < 1 \times 10^{-8}$ ). This region harbors 23 non-LOC genes and 13 LOC genes, including *KIT* (Figure S12, Supporting Information). The signals of the SNP- $F_{ST}$  contrasting 22 gray and 29 non-gray QTP cattle (BTA6:69750001-72725001,  $P < 0.001$ ) overlapped with the SNP-GWAS signals (Figure 5a). However, no functional SNP or InDel was found. We then performed a whole-genome SV-based GWAS (SV-GWAS) (Figure 5b) and SV-based  $F_{ST}$  (SV- $F_{ST}$ ) analysis (Figure S13, Supporting Information), and confirmed the highest signal of the 2-Mb heterozygous INV (BTA6:69772181-71772202), together with two TRAs in BTA17 (47058249) and BTA6 (70471991) (Figure 5b), overlapping with the region detected by SNPs for the gray coat color phenotype in cattle.

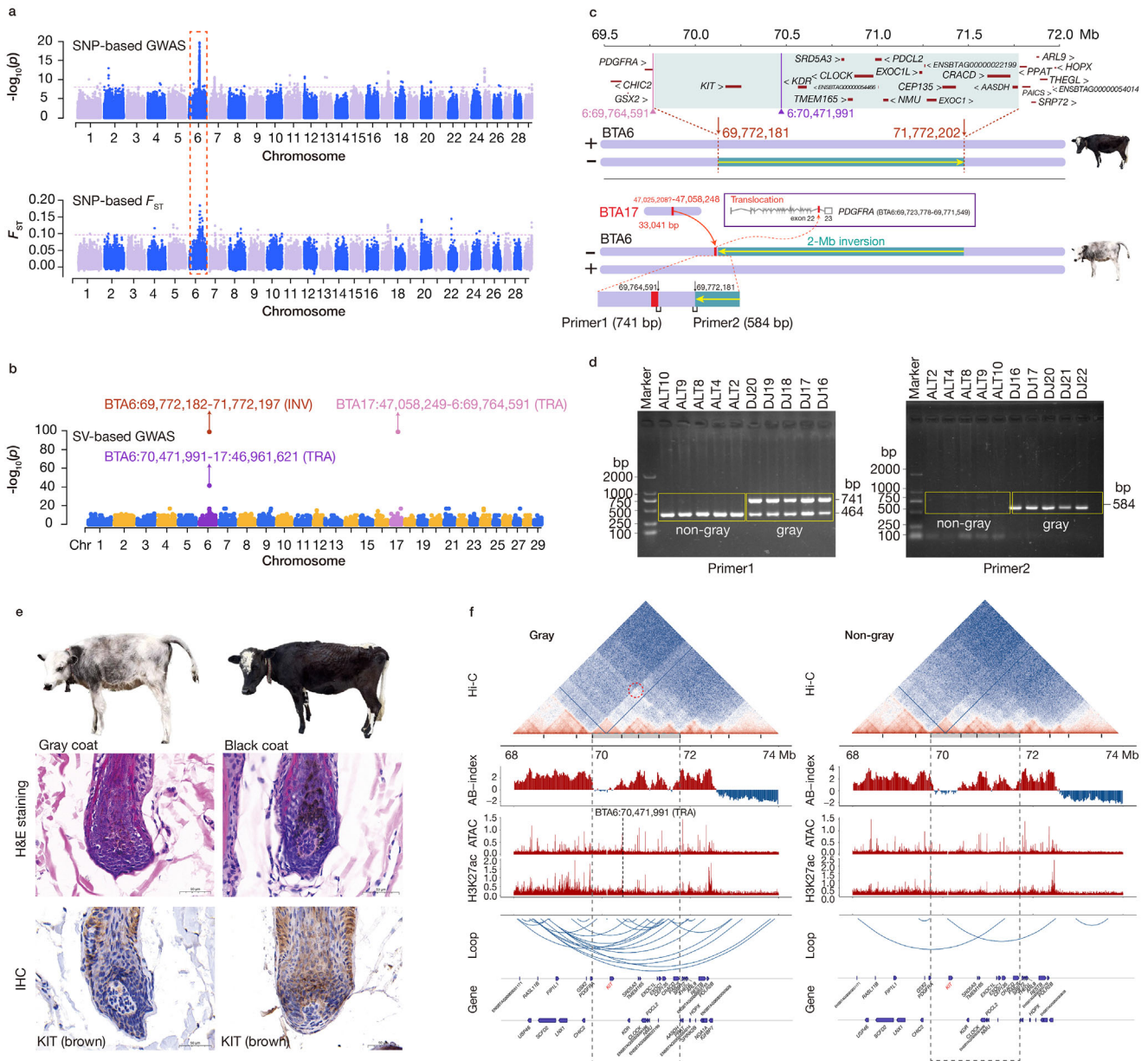
The INV was confirmed by analysis of discordant and split reads at the junctions in gray cattle compared with those at the

junctions in non-gray cattle (Figure S14, Supporting Information). We further confirmed the presence of this 2-Mb heterozygous INV via the program npInv,<sup>[68]</sup> an accurate tool for detecting and genotyping inversions using multiple alignment long reads. This INV was probably caused by error-prone nonhomologous end joining (NHEJ) events and contains 20 coding genes, including *KIT* (Figure 5c). There was no disrupted gene present at the INV boundary; however, a translocation fragment of  $\approx 33$  kb from BTA17 was found 7590 bp upstream of the INV proximal breakpoint (Figure 5b,c, Supporting Information). This fragment was inserted into the 22nd intron of *PDGFRA*. We confirmed the INV and TRAs by PCR in 23 other gray cattle (Figure 5d; Figure S15, Supporting Information).

To further investigate the possible origin of this INV in QTP cattle, we conducted a comprehensive haplotype phylogenetic analysis using sequences from the INV region (BTA6:69722181-71822202). The results indicate that the INV-carrying haplotypes show strong phylogenetic clustering with South Asian indicine cattle (*Bos indicus*), while the alternative haplotype primarily groups with East Asian taurine cattle (*Bos taurus*) (Figure S16, Supporting Information). This suggests that the inversion haplotype segment was likely originated from South Asian indicine cattle, consistent with the presence of South Asian indicine ancestry components in QTP cattle populations as previously reported.<sup>[19]</sup>

Due to the well-established role of *KIT* in determining white-spotting phenotypes in several species, this gene is an obvious candidate for the gray coat phenotype in QTP cattle. *KIT* is expressed in several tissues, particularly in mast cells and melanocytes.<sup>[69]</sup> Using H&E staining and immunohistochemistry of skin samples, we observed a more intense presence of melanin and more pronounced *KIT* signals around the hair follicles of black cattle than in those of gray cattle (Figure 5e). Hi-C data from gray cattle and non-gray cattle indicated that the breakpoints of this INV were located at topologically associated domains (TAD) boundaries (Figure 5f; Table S19, Supporting Information). The presence of the INV led to stronger interactions across the TAD boundaries and stronger H3K27ac signals upstream of the inversion in gray cattle. Evolutionary conservation was confirmed through cross-species alignment (LiftOver; BTA6:68-70 Mb to GRCm39), showing mice *KIT* enhancer sequences are conserved in cattle (Figure S17a, Supporting Information), with these enhancers known to reside within *KIT* TAD in mice melanocytes.<sup>[69]</sup> Furthermore, the TRA at BTA6:70471991 in gray cattle precisely localizes to the CTCF-defined TAD border between *KIT* and *KDR* (Figure 5f; Figure S17b, Supporting Information), while deletion of the *Kit-Kdr* TAD boundary in mice results in a lighter coat color phenotype.<sup>[69]</sup> Together, these findings suggest that the INV and associated TRA events may alter *KIT* expression through TAD boundary disruption and consequent enhancer dysregulation, contributing to the gray coat phenotype in QTP cattle. We then examined the linkage disequilibrium (LD) level between the SNPs positioned within or adjacent to the 2-Mb heterozygous INV (Figures S18 and S19, Supporting Information). The LD levels between distant SNPs are generally higher in gray cattle than in non-gray cattle, indicating a suppression of recombination by the large inversion.

Previous studies have established the cell-specific expression pattern of the *Kit* gene in mice, leading us to hypothesize



**Figure 5.** A 2-Mb heterozygous inversion related to the gray coat phenotype of Qinghai-Tibetan Plateau (QTP) cattle. a) Genome-wide Manhattan plots generated via GWAS and  $F_{ST}$  based on SNPs for the gray coat phenotype. b) Genome-wide Manhattan plot using the GWAS based on SVs for gray coats. c) Rearrangement of the 2-Mb INV in non-gray (top) and gray (bottom) QTP cattle. This region displays the genes located on BTA6 between positions 69.50-72.00 Mb in the *B. taurus* reference genome (ARS-UCD1.2) from the Ensembl database, with the blue shaded area indicating the specific location of the inversion at BTA6:69772181-71772202. In addition to the 2-Mb inversion on BTA6, we also identified two translocations (pink and purple lines), one of which involves a fragment from BTA17 that is translocated into intron 22 of *PDGFRA*. d) PCR validation of the 2-Mb INV. We designed two pairs of PCR primers at the insertion site of the translocation (Primer1, BTA6:69764591) and the starting position of the inversion (Primer2, BTA6:69772181) and validated them in gray and non-gray cattle. e) Skin sections of gray and black cattle stained with hematoxylin and eosin (H&E) and subjected to immunohistochemistry (IHC), respectively. For the IHC analysis, a rabbit antibody against KIT was used, with the brown color indicating positive staining and the blue color representing the cell nuclei. Scale bars, 50  $\mu$ m. f) Chromatin architecture and regulatory landscape in the gray and non-gray cattle genomic regions flanking the 2-Mb INV region on BTA6. The top section represents a Hi-C contact map, which visualizes chromatin interactions, with red and blue colors indicating high interaction densities. The grey shaded area encompasses four consecutive TADs in the cattle genome, with a highlighted region (red dashed circle) indicating strong chromatin interactions in gray cattle. The AB index plot, where red bars represent active (A) compartments and blue bars denote inactive (B) compartments. The ATAC and H3K27ac tracks show regulatory activity in this region. Loops derived from Hi-C data highlight chromatin interactions across this region. The bottom panel shows an annotation map of genes located on BTA6 between positions 68–74 Mb in the ARS-UCD1.2 genome from the Ensembl database.

that the effects of INV might exhibit cell-type specificity (e.g., melanocytes). However, current bulk-level Hi-C interaction data lack resolution to discern such cellular heterogeneity. Future investigations employing single-cell Hi-C or other high-resolution chromatin conformation capture methods could help verify these possibilities and elucidate cell-type-specific regulatory mechanisms.

### 3. Discussion

In this study, we generated a high-quality assembly of Tibetan taurine cattle (Tibetan\_v1) using a combination of PacBio HiFi and Hi-C technologies. Tibetan\_v1 has quality metrics comparable to or even exceeding those of other currently available cattle reference genomes. It may serve as a reference genome serve as an important data resource for further analysis of Asian cattle breeds, particularly for studies of the bovine pangenome.<sup>[70]</sup> Bovine autosomes are acrocentric,<sup>[71]</sup> and a complete bovine assembly is close to “centromere-to-telomere”.<sup>[72]</sup> We identified 20 autosomes with telomeric repeats in the Tibetan\_v1 assembly.<sup>[72]</sup> In the application of population genetics, the use of a population-specific reference genome (Tibetan\_v1) instead of ARS-UCD1.2 more accurately distinguishes East Asian cattle, detects rare low-frequency variants, and allows the conversion of multiallelic into biallelic variants. The multiallelic variants covered by the European bovine reference genome in QTP cattle are mainly restricted to highly polymorphic immune-related genes (*BoLA*, *BoLA-NC1*, etc.).<sup>[73,74]</sup> Our results suggest that sequences with high nucleotide diversity will benefit from population-specific reference genomes. This observation aligns with findings in other species, such as Tibetan goat, where population-specific reference genomes have been shown to improve the detection of novel selective regions.<sup>[75,76]</sup>

Whole-genome scanning of nucleotide variations has been widely used to identify candidate genes for hypoxia adaptation in high-altitude species.<sup>[19]</sup> In this study, we focused on the role of SVs in high-altitude adaptation, which involves both polygenic inheritance and gene introgression. We generated the largest dataset of QTP cattle SVs to date and identified SVs that are highly differentiated between high- and low-altitude cattle populations, potentially linked to high-altitude adaptation. Increasing oxygen delivery and reducing oxygen consumption are the two primary responses to hypoxia; the former relies on the improvement of blood and vascular conditions, and the latter mainly refers to glucose and lipid metabolism.<sup>[77]</sup> We identified SVs associated with high-altitude adaptation in QTP cattle, primarily overlapping with genes related to erythropoiesis and angiogenesis (*SSH2*, *VGLL4*, *PLCB1*, *HPSE2*, and *HPSE*) and energy metabolism (*SORD*, *NDUFB6*, *SARDH*, and *ADIPOQ*). Similar adaptive pathways (e.g., hypoxia response and energy metabolism) were previously observed in yak,<sup>[1,63,78]</sup> though involving distinct genes, suggesting convergent mechanisms for high-altitude survival between QTP cattle and native yak.

The selection signal around the *SORD* gene was highly consistent in both the SNPs and the SVs between the high- and low-altitude cattle populations, indicating its potential crucial role in high-altitude adaptation. A previous study on ischemia/reperfusion-induced injury of the rat heart revealed that *SORD* (also called *SDH*) affects the induction of HIF-1 $\alpha$  by

regulating the level of NAD<sup>+</sup>.<sup>[55]</sup> HIF-1 $\alpha$  is a subunit of a transcription factor called hypoxia-inducible factor 1.<sup>[79]</sup> The physiological function of HIFs is to promote the adaptation of cells to low oxygen by inducing neovascularization and glycolysis.<sup>[80]</sup> Other high-signal SV-associated genes, such as *CTNNB1*<sup>[53]</sup> and *HPSE*,<sup>[81]</sup> may also help the high-altitude adaptation of QTP cattle through the HIF pathway. Notably, some strongly selected SVs (e.g., *SSH2*) were located outside ATAC-seq predicted open chromatin regions. This observation suggests two non-exclusive possibilities: 1) The regulatory effects of adaptive SVs may operate through mechanisms independent of canonical chromatin accessibility; 2) Tissue-specific regulatory landscapes-while our ATAC-seq covered six major tissues (heart, liver, lung, muscle, adipose and thyroid), key regulatory elements for certain genes might be active in other cell types. This highlights the importance. This highlights the necessity for future studies to integrate broader tissue sampling, single-cell multi-omics profiling, and functional validation to comprehensively resolve the regulatory mechanisms underlying high-altitude adaptation. Additionally, we identified *d*<sub>i</sub>-SVs in three peroxisomal genes (*GNPAT*, *CAT*, and *ACOT2*). We detected a 102-bp DEL in the first intron of *GNPAT*, which is located within a predicted chromatin open region. The results from the luciferase reporter assay showed that the fragment containing the 102-bp DEL significantly increased luciferase expression in mouse 293T cells, suggesting that the DEL sequence may block the function of a regulatory element.<sup>[82]</sup> In Tibetans, *GNPAT* has also been identified as a key gene involved in the response to UV treatment through the promotion of melanin synthesis.<sup>[83]</sup> *CAT* encodes a catalase that plays a key role in protecting cells and dismutates H<sub>2</sub>O<sub>2</sub> to water and oxygen in response to oxidative stress. Combined with previous study,<sup>[54]</sup> it suggests that peroxisomal genes may play crucial roles under hypoxic conditions in QTP cattle.

BTA28 contains relatively many yak-introgressed regions, which is consistent with previous SNP-based analyses.<sup>[19]</sup> Notably, this region includes hypoxia genes such as *EGLN1*, *EXOC8*, *GNPAT*, *SPRTN*, *TSNAX*, *FAM89A*, *TRIM67*, and *ARV1*.<sup>[54,83,84]</sup> Interestingly, the same genomic region has also been reported in high-altitude species such as the zoker (*Eospalax baileyi*)<sup>[85]</sup> and Tibetan humans.<sup>[86]</sup> SNPs revealed that the proportions of yak ancestry ranged from 0.64% to 3.26% in the QTP cattle genomes, among which Nangqian, Changdu, and Yushu were the highest, followed by Diqing, Zhangmu, Dingjie, and Shigatse taurine cattle.<sup>[19]</sup> In this study, we identified distinct hotspots of SV introgression across different QTP populations, indicating various patterns of yak introgression. For example, in the Changdu population, the highest introgression region signal was found on BTA2, which includes the *NFE2L2* gene, which is involved in the oxidative stress response.<sup>[87]</sup> Changdu cattle inhabit an agro-pastoral zone where interbreeding with yak is common. Our results indicate that not all QTP cattle populations have the same yak introgression signals or similar adaptive mechanisms, which may reflect the size of the QTP with geographical barriers. This has also influenced human variation. The top candidate genes that carry adaptive genetic variation previously found in Tibetan humans, such as *EPAS1* and *EGLN1*, do not have strong positive selection signals in Deng people, who also live on the Plateau and have selection signals in *KLHL12* and other genes.<sup>[88]</sup>

Chromosomal INVs may affect recombination and chromosome structure. However, INVs can be challenging to detect, so the prevalence and hence the significance of INVs segregating within a species are often overlooked. LRS data for gray cattle revealed a 2-Mb heterozygous INV that distinguished phenotypically gray cattle from non-gray cattle. These INVs may very well be favored by natural selection because they suppress recombination between coadapted genes,<sup>[89]</sup> where the gray coat phenotype contributes to adaptation to high-altitude environments. Previous studies in humans and other mammals have shown that large INVs modify gene expression via two major mechanisms: i) directly disrupting regulatory elements adjacent to breakpoints and ii) rearranging regulatory elements positioned within or near the inverted region.<sup>[89]</sup> At the same time, we observed that the 2-Mb INV was accompanied by two TARs, a phenomenon that appears to be common in cytogenetically detected INVs.<sup>[90]</sup> At the INV breakpoints, complexity and imbalances are often observed, which may result from error-prone repair of processed double-strand breaks.<sup>[91]</sup> Inhibitory recombination of inverted heterozygotes has been reported in several species, such as flat-fruit peach<sup>[92]</sup> and deer mice.<sup>[93]</sup> A light coat color may contribute to adaptation to high-latitude or tropical/subtropical environments by reflecting a relatively large proportion of incident solar radiation.<sup>[94]</sup> We propose that the gray coat phenotype in cattle arises through structural rearrangement of regulatory elements by this INV, or alternatively, through potential disruption of the *KIT-KDR* TAD by the TRA, leading to dysregulated *KIT* expression.

GWAS based on SNPs constitute a powerful experimental strategy for identifying genetic variations underlying animal traits.<sup>[95,96]</sup> However, causative genomic variations often involve long stretches of DNA, such as the 2-Mb heterozygous INV identified in this study, and are not detected by genome-wide SNP genotyping. Our study suggests that LRS analysis can identify important SVs that SRS data cannot detect, providing high-confidence candidate genes and variants for subsequent functional studies.

## 4. Conclusion

We generated a high-quality genome assembly for Tibetan cattle and presented a comprehensive catalog of SV variations. We evaluated the important role of the population-specific reference genome in detecting rare and low-frequency variations. Leveraging the advantages of population-scale long-read sequencing, we identified genes and related SVs associated with high-altitude adaptation and coat phenotypes, offering new insights into the plateau adaptability of cattle to the QTP.

## 5. Experimental Section

**Ethics Statement:** This study was approved by the Institutional Animal Care and Use Committee of Northwest A&F University (FAPWCNWFU, Protocol no. NWFAC 1008).

**Samples Collected for Tibetan Cattle Genome Assembly:** Whole blood from the Dingjie bull DJ12 was collected to construct the Tibetan cattle genome assembly. The bull was located in Dingjie County, Shigatse City, Xizang Autonomous Region, China. Genomic DNA was extracted from the tissues of the animals using the phenol/chloroform method. The result-

ing DNA was sequenced using the Pacific Biosciences (PacBio) Sequel II platform at Novogene, yielding a total of 102 Gbp, corresponding to a genomic coverage of  $\approx 39\times$ . In addition to long reads, the same animal was resequenced using Illumina HiSeq X Ten paired-end short-read sequencing, yielding 376 Gbp with an average insert size of 350 bp, corresponding to a genomic coverage of  $\approx 145\times$ .

**In Situ Hi-C Library Preparation:** In this study, one library was generated from the blood of the Dingjie bull DJ12 and six in situ Hi-C libraries from the skin tissue of gray and non-gray cattle. Hi-C libraries were constructed according to previous studies.<sup>[97]</sup> Briefly, the samples were crosslinked with 1% formaldehyde for 10 min at room temperature and quenched with 0.125 M glycine for 5 min. The crosslinked cells were subsequently lysed. Endogenous nuclease was inactivated with 0.3% SDS, and chromatin DNA was digested with 100 U of MboI (NEB), marked with biotin-14-dCTP (Invitrogen), and then ligated with 50 U of T4 DNA ligase (NEB). After the crosslinking was reversed, the ligated DNA was extracted with a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. The purified DNA was sheared into 300- to 500-bp fragments and subjected to further blunt-end repair, the addition of A-tails and adaptors, followed by purification using biotin-streptavidin-mediated pull-down and PCR amplification. Finally, the Hi-C libraries were quantified and sequenced on the MGI-seq platform (BGI, China).

**Whole-Genome Nanopore LRS Samples:** Blood or ear tissue samples were collected from 36 cattle in five high-altitude regions (above 3000 m) in the Xizang Autonomous Region, Qinghai, and Yunnan Province, China. Low-altitude cattle were collected from Qinghai, Jilin, and Shandong Province, all below 3000 m above sea level, for a total of 28 blood or ear tissue samples. The sequencing libraries were prepared using an SQK-LSK109 ligation kit (Oxford Nanopore Technologies; Oxford, UK). Sequencing was performed using PromethION flow cells (R9.4). Base calling was performed using Guppy (v.5.1.13). In addition, 20 published LRS datasets were downloaded from low-altitude cattle for joint analysis. LRS data were downloaded from six domestic yak and six wild yak from the NCBI to identify SVs. All the samples used in this study are listed in Table S8 (Supporting Information).

**Samples Collected for Whole-genome Illumina SRS:** The Illumina SRS data of 281 cattle from six geographic regions (Africa, Europe, Eurasia, Northeast Asia, South China, and South Asia) as well as six wild bovine species ( $n = 12$ ) were generated in our study or retrieved from the sequence read archive of the National Center for Biotechnology Information (NCBI, Table S16, Supporting Information).

**Genome Assembly and Quality Assessment:** Hifiasm (v.0.16.1-r375) was used to generate the assembly from HiFi CCS reads with default parameters.<sup>[25]</sup> Hifiasm yields one primary contig assembly and a pair of partially phased contig assemblies. The primary contigs were scaffolded to chromosomes using clean Hi-C reads with Juicer (v.1.5)<sup>[98]</sup> and 3D-DNA (v.201008).<sup>[99]</sup> The assembly was manually reviewed and adjusted using Juicebox Assembly Tools.<sup>[100]</sup> Genome completeness was assessed using BUSCO (v.4.0)<sup>[101]</sup> with the Mammalia\_odb10 database. Merqury (v.1.0)<sup>[27]</sup> was used to estimate the quality value (QV) by employing high-depth Illumina SRS data from the same individual whose genome was assembled. Mash<sup>[26]</sup> was used to estimate the distance between the assembled genome of the Tibetan cattle genome and other genomes.

**Genome Annotation:** Repetitive elements in Tibetan\_v1 were identified by their matches to RepeatMasker (v.20140131) using RepeatMasker (v.4.0.5) (<http://www.repeatmasker.org>). To identify protein-coding genes in Tibetan\_v1, Liftoff (v.1.5.2)<sup>[102]</sup> and GffRead (v.0.12.1)<sup>[103]</sup> were used to generate gene annotations from ARS-UCD1.2. The annotation used for the transfer was NCBI GCF\_02263795.1\_ARS-UCD1.2\_genomic.gff.

**Variant Calling Evaluation:** To assess the variant calling performance of the different references, the SRS data of 27 Dingjie samples were aligned to each of the two assemblies (ARS-UCD1.2 and Tibetan\_v1) using BWA-MEM (v.0.7.13-r1126) with default parameters, and the duplicated reads were removed using Picard Tools (<http://broadinstitute.github.io/picard>). The Genome Analysis Toolkit (GATK, v.3.8-1-0-gf15c1c3ef) was used to detect SNPs and InDels. The variations were called using the "HaplotypeCaller", "GenotypeGVCFs", and "SelectVariants" of GATK. After SNP or InDel calling, "VariantFiltration" was used to discard sequencing and align-

ment artifacts from the SNPs with the parameters 'QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, and SOR > 3.0' and based on a mean sequencing depth of variants (all individuals) of "< 1/3x and > 3x." To identify the biallelic SNPs in Tibetan\_v1 that appeared to be multiallelic under ARS-UCD1.2, first the biallelic SNPs were extracted from Tibetan\_v1 and the multiallelic SNPs from ARS-UCD1.2 in 27 Dingjie cattle. Then, using the chain file created between Tibetan\_v1 and ARS-UCD1.2, LifOver tools were employed for coordinate conversion.

**Population Structure Analysis and Estimation of the Effective Population Size:** To examine the fine-scale population structure revealed by genetic variants from different assemblies (ARS-UCD1.2 and Tibetan\_v1), PCA of cattle populations was performed in the QTP region and northern China. PCA was conducted using the smartPCA program in EIGENSOFT (v.5.0).<sup>[104]</sup> A neighbor-joining (NJ) tree was constructed using the matrix of pairwise genetic distances calculated by PLINK (v.1.9).<sup>[105]</sup> The unrooted NJ tree was then visualized via MEGA (v.5.0)<sup>[106]</sup> and FigTree (v.1.4.3) (<http://tree.bio.ed.ac.uk/software/figtree/>). A NeighborNet network was constructed using Reynold's distances between populations using SplitsTree App (v.6.3.32).<sup>[107]</sup>

For the genetic variants called from different reference assemblies (ARS-UCD1.2 and Tibetan\_v1), MSMC2 analysis was applied to estimate the effective population size ( $N_e$ ) of the Tibetan cattle, Anxi cattle, and Angus populations over time. This method was applied to all groups with two deep-coverage (>14 x) individuals per group. The samples (coverage) used in this analysis were as follows: Anxi, O718844 (32.72 x) and V347862 (33.36 x); Angus, SRR1365144 (16.27 x) and SRR1425124 (16.44 x); and Tibetan cattle, Xizang22 (26.28 x) and Xizang7 (24.56 x). For the calculation of the effective population size, the parameters of MSMC2 were set to 'msmc2 -t 10 -p 1\*2+25\*1+1\*2 -l 0, 1, 2, 3' and 'msmc2 -t 10 -p 1\*2+25\*1+1\*2 -l 4, 5, 6, 7'. For the calculation of population separation, the parameters of MSMC2 were set to 'msmc2 -t 8 -P 0, 0, 0, 1, 1, 1, 1 -s -p 1\*2+25\*1+1\*2'. For effective population size inference, two individuals (4 phased haplotypes) from each group were used. A time scale for generation time of  $g = 6$  and a mutation rate per generation of  $\mu g = 1.26 \times 10^{-8}$  were used.

**SV Discovery and Genotyping:** LRS data from 84 samples were aligned to ARS\_UCD1.2. Mapping was performed using NGMLR (v.0.2.7)<sup>[32]</sup> with default parameters. SV calling was performed using CuteSV (v.2.0.3),<sup>[30]</sup> SVIM (v.1.4.0),<sup>[31]</sup> and Sniffles (v.2.2).<sup>[32]</sup> Three minimum supporting reads were needed for each SV. The minimum length of the SVs was at least 50 bp. Specifically, CuteSV was run with the parameters '-max\_cluster\_bias\_INS 100 -diff\_ratio\_merging\_INS 0.3 -max\_cluster\_bias\_DEL 100 -diff\_ratio\_merging\_DEL 0.3 -min\_support 3 -min\_size 50 -genotype -report\_readid -sample.' The following parameters were set for SVIM: '-min\_sv\_size 50 -min\_mapq 20 -minimum\_depth 3 -insertion\_sequences -sequence\_alleles -read\_names -sample.' Sniffles was run with the parameters "-minsupport 3 -output-rnames" to remove positions marked as IMPRECISE for INFO or as UNRESOLVED for FILTER. SURVIVOR (v.1.0.7)<sup>[108]</sup> was used to merge the SVs supported by two or three calling methods, with a maximum allowed pairwise distance of 500 bp between breakpoints. Individuals were genotyped using the force calling function of Sniffles with the -genotype-vcf option. Multisample VCF files were merged and deduplicated using BCFtools (v.1.17)<sup>[109]</sup> and the "collapse" module in Truvari software.<sup>[110]</sup> After filtering out unreliable genotypes (all 0/0), a total of 222 528 SVs were obtained.

These 222 528 SVs were genotyped in 281 cattle and 12 individuals from other bovine species with available SRS data. Reads were mapped to ARS-UCD1.2 using BWA-MEM (v.0.7.13-r1126) with default parameters. Paragraph (v.2.4a)<sup>[111]</sup> was used to genotype the SVs from the SRS data. BCFtools was used to combine the results for all the genotypes. Then all the unfiltered genotypes were replaced in Paragraph software with missing genotypes (./.) and excluded SVs without any remaining nonreference genotypes.

Following the methodology outlined in a previous study,<sup>[41]</sup> the midpoint of each SV was used to calculate SV hotspots via a publicly available script (<https://github.com/daewoo00/primatR/blob/master/R/hotspotter.R>) with minor modifications.

**Detection of Selective Signals:** To detect SVs with highly differentiated allele frequencies between QTP cattle (Bailang, Dingjie, Diqing, Langkazi, and Yushu) and low-altitude cattle (Fuzhou, Yanbian, Luxi and Simmental), the  $d_i$  values for 222528 SVs were calculated from 84 LRS data as described in previous studies.<sup>[43,44]</sup> Chaidamu and Mongolian cattle were not included in either of the two groups because their altitudes were between 800 and 3000 m, which may have affected the results. For the  $SV-F_{ST}$ , the Weir and Cockerham estimator were employed for the  $F_{ST}$  estimates via VCFtools (v.0.1.16)<sup>[112]</sup> to identify population-stratified SVs in five groups: all QTP cattle versus low-altitude cattle; Bailang versus low-altitude cattle; Dingjie versus low-altitude cattle; Diqing versus low-altitude cattle; Langkazi versus low-altitude cattle; and Yushu versus low-altitude cattle. SVs with  $P$  values less than 0.001 ( $Z$  test) were selected as  $d_i$ -SVs.

**Annotation of SVs:** Consistent sequences of SVs were extracted and compared with consistent repeat sequences of mammals using RepeatMasker (v.4.0.5) (<http://www.repeatmasker.org>) to determine whether they were repetitive sequences. Specifically, for DELs and DUPs, BEDTools (v.2.25.0)<sup>[113]</sup> were used to intersect these segments with the repetitive sequences of the ARS-UCD1.2, which requires a minimum reciprocal overlap of 80%. For INs, the sequences were extracted and annotated using RepeatMasker. INs were classified as repetitive sequences if the query sequence aligned completely, allowing for a maximum of 20 bp unaligned regions at the ends (left < 20). The functional regions of the SVs in the genomes were annotated using ANNOVAR.<sup>[33]</sup> KEGG and GO enrichment analyses were performed for SV-linked genes in population stratification, and functional categories with corrected  $P$  values less than 0.05 were considered significantly enriched.

**ATAC-seq Analysis:** Some of the ATAC-seq data used in this study were generated through our own sequencing experiments, whereas others were obtained from publicly available NCBI datasets (Table S14, Supporting Information). After the adapters were checked for and trimmed with Trim Galore (v.0.6.10) (<https://github.com/FelixKrueger/TrimGalore>), the ATAC-seq clean reads were mapped to ARS-UCD1.2 using Bowtie2 (v.2.4.5).<sup>[114]</sup> BAM files were sorted using SAMtools (v.1.3),<sup>[115]</sup> and duplicate alignments were removed with Picard (v.2.20.2) (<http://broadinstitute.github.io/picard>). The peaks for individual replicates were called separately with MACS2 (v.2.2.7.1)<sup>[116]</sup> and then merged across three duplicates within the same tissue using BEDTools (v.2.25.0).<sup>[113]</sup> The R package obtained from <https://github.com/junjunlab> is used for drawing.

**Dual-Luciferase Assay:** For the 102-bp DEL of the *GNPAT* gene, sequences containing the entire SV and the 360-bp flanking sequences were extracted from the cattle genome (ARS-UCD1.2) and inserted into the pGL3 promoter. The HEK293T cell line was purchased from the National Science & Technology Infrastructure (Shanghai, China). Transfection was performed when the cell density in the 12-well plate reached approximately 80%. Lipofectamine 2000 (Invitrogen, CA, USA) was mixed with plasmids (recombinant plasmid and pRL-TK plasmid vector) at a ratio of 1:1, incubated for 20 min, and then added to the cells. The dual-luciferase reporter assay kit was purchased from Promega (Wisconsin, USA), and the experiments were conducted according to the manufacturer's instructions. After cell lysis, the lysates were transferred to a 96-well plate, and two substrates were sequentially added. Detection was then performed using a Multiskan system (TECAN, Spark, Switzerland). Each experiment was independently performed three times. All the data are presented as the means with standard deviations, and significant differences between groups were examined via two-sided Student's  $t$  test. A  $P$  value < 0.05 was considered to indicate a significant difference.

**Discovery of the Introgressed SVs:** Yak LRS data were downloaded from the NCBI to identify SVs, including six domestic yak and 6 wild yak (Table S8, Supporting Information). Using SURVIVOR software, 12 yak and 84 cattle vcf files were merged, and the subsequent genotyping workflow was the same as that used for "SV discovery and genotyping". After filtering out unreliable genotypes (all 0/0), a total of 250156 SVs was obtained. First, using LRS data, SVs specific to QTP cattle and fixed in yak genomes but absent from low-altitude cattle populations (Fuzhou, Yanbian, Luxi, and Simmental) were investigated. A total of 7293 autosomal SVs possibly derived from yak introgression were identified in QTP cattle (Figure 4a). To further validate the SVs, haplotype trees were constructed using SNPs

in the 50-kb flanking of the SV, using a SNP dataset of 148 QTP (36 LRS and 112 SRS) samples and 16 reference SRS samples.

To expand the population size, the SVs were further genotyped using 293 SRS data, including 112 QTP cattle and 12 high-coverage individuals from other bovine species (Table S16, Supporting Information). Paragraph (v.2.4a)<sup>[111]</sup> was used to genotype the SVs from the SRS data. Cattle from low-altitude taurine breeds/populations (African taurine cattle, European taurine cattle, and East Asian taurine cattle) were used as control populations for genotyping (Table S18, Supporting Information). Then SVs that were shared with yak genomes but not present in other cattle samples were searched, and 2334 candidate introgressed tract SVs remained after SRS genotyping. The *-maf* parameter of VCFtools (v.0.1.16)<sup>[112]</sup> with the *-freq* parameter was used to calculate the allelic frequency in different breeds/populations. IQTREE (v.1.6.6)<sup>[117]</sup> was used to construct a phylogenetic tree for the introgressed yak regions. ModelFinder<sup>[118]</sup> was used to identify the best model of the phylogenetic tree.

**Hi-C Data Preprocessing and Normalization:** Hi-C read pairs were processed using the HiC-Pro pipeline<sup>[119]</sup> with default parameters. After removing duplications, valid pairs were summed across three technical replicates and used to generate an interchromosome contact matrix. The normalized observed contact matrices were generated using the Knight–Ruiz algorithm within the Juicer toolkit to eliminate intrinsic biases within the matrices. Quantile normalization was subsequently performed using the BNBC R package (v.1.0.0) to mitigate biases between samples. The correlation between normalized matrices of technical replicates was calculated using GenomeDISCO<sup>[120]</sup> (Figure S20, Supporting Information).

**A/B Compartment Determination and Analysis:** A/B compartments at 20 kb resolution were identified using PCA and the A–B index as described in a previous study.<sup>[121]</sup> In brief, observed/expected contact matrices were obtained by normalizing intrachromosomal observed contact matrices using KR normalization and quantile normalization. For loci *i* and *j*, the expected contact frequency was determined as the median observed contact frequency at the same genomic distance. The observed/expected contact matrices were then calculated by dividing the observed contact frequency by the expected contact frequency for each locus pair. The “prcomp” package was applied in R (v.4.1.2) with default parameters to these matrices to generate PC1 vectors. The gene density for each bin was defined as the number of transcriptional start sites (TSSs) within  $\pm 100$  kb of the bin. Compartments A and B were identified by calculating the Pearson’s correlation between PC1 values and gene density for each chromosome using the “cor.test” function in R (v.4.1.2). Positive PC1 values were assigned to compartment A, and negative values were assigned to compartment B if the correlation was positive. The assignment was reversed if the correlation was negative. The A–B index was subsequently calculated at a 20 kb resolution following previously described methods, representing the likelihood of a sequence interacting with either the A or B compartment.<sup>[121]</sup> Bins 20 kb in length with a positive A–B index were classified as A compartments, whereas those with a negative A–B index were classified as B compartments.

**TAD Identification and Loop Calling:** TADs were identified via a previously reported insulation score (IS)<sup>[122]</sup> method based on the KR and quantile-normalized intrachromosomal observed contact matrices at a 20-kb resolution. The IS reflects the aggregate interactions passing across each bin, and IS boundaries were identified using the public script matrix2insulation.pl (<https://github.com/dekkerlab/cworld-dekker>) with the following parameters: *-v*, *-is* 260 000, *-ids* 200 000, *-im* mean, *-nt* 0.1, and *-bmoe* 0. Fit-Hi-C software (v.2.0.8)<sup>[123]</sup> was used with default parameters to identify loops at a 5-kb resolution based on the KR and quantile-normalized intrachromosomal contact matrices, with a *q* value  $< 1 \times 10^{-8}$  used as a cutoff.

**GWAS and  $F_{ST}$  Analysis of Gray Cattle:** For both the SNP-GWAS and SNP- $F_{ST}$  analyses, read mapping and SNP calling for the SRS data were performed as described in the previous “Variant calling evaluation” section. To identify the locus associated with gray coat color, SRS data from 23 gray cattle and 258 nongray cattle were used. SNP-GWAS analysis was conducted on 62091927 SNPs using the GEMMA.<sup>[124]</sup> The genome-wide distribution of the  $F_{ST}$  values was estimated using VCFtools (v.0.1.16)<sup>[112]</sup>

with a 50-kb window size and 25-kb increment to examine pairwise genetic differentiation between 23 gray and 29 nongray cattle. For SV-GWAS analysis, five gray cattle were used and the remaining 79 non-gray cattle from the LRS dataset were analyzed with the GEMMA model. For SV- $F_{ST}$ , five gray cattle and 11 non-gray cattle were used to calculate the  $F_{ST}$  value for each SV.

**SV Validation:** Visualization of the detected SVs was performed using IGV (v.2.2).<sup>[125]</sup> To obtain more reliable detection outputs for inversions, nplnv (v.1.24)<sup>[68]</sup> was used to detect and genotype inversion variants. To further verify the authenticity and heterozygosity of the 2-Mb INV and two associated TRAs, breakpoint-spanning primers were designed and PCR validation was performed in an additional 23 gray cattle. Three primer pairs were used to independently validate each SV: Primer1 (F: 5'-GTTGACCCCTTCTTTCCA-3'; R: 5'-GACCCCTCTTCTATCCC-3') targeted TRA (BTA17:47058249-BTA6:69764591); Primer2 (F: 5'-ATCTTAGCTGTAGCATGT-3'; R: 5'-ACTTGTAGGTTACTGGTT-3') spanned the INV (BTA6:69722181-71822202) breakpoints; and Primer3 (F: 5'-TCCTTCCCTAAGATTCAGA-3'; R: 5'-GAGCCCTTGACCCACTA-3') amplified TRA (BTA6:70471991-BTA17:46961621).

For the four introgressed SVs, four primer pairs were designed: Primer4 (F: 5'-AAGACGCTTGCTCC-3'; R: 5'-CTTCTCCAAAACCACA-3') targeted DEL (BTA28:4236113-4237437); Primer5 (F: 5'-CTGGTCCATCACTTC-3'; R: 5'-TTGCCATACCTATTTT-3') spanned the INS (BTA28:4204966) breakpoint; Primer6 (F: 5'-CTTAGACTTCAGGA-3'; R: 5'-GCTCATCATAGGTTTGTGA-3') amplified DEL (BTA9:87336847-87336904); and Primer7 (F: 5'-ACCCTGTACCTCTAGCA-3'; R: 5'-GACATGACTGAGCGACTG-3') targeted DEL (BTA2:19540699-19540751) (Figure S21, Supporting Information).

PCR was performed in 25  $\mu$ L reactions with 1  $\mu$ L of genomic DNA (50 ng), 1  $\mu$ L each of forward and reverse forward primers (10  $\mu$ M), and 22  $\mu$ L of Golden Star T6 Super PCR mix (Beijing Tsingke Biotech Co., Beijing, China). The PCR products were examined using 1.5% agarose gel electrophoresis. The presence and size of the amplified fragments were determined and used to infer the genotype of the inversion.

**Immunohistochemistry and H&E Staining:** Skin samples were fixed in 4% paraformaldehyde, washed in PBS, dehydrated and embedded in paraffin. The rabbit polyclonal antibody anti-KIT (D260893, Sangon) was used for immunohistochemical staining.

**Linkage Disequilibrium:** LD was accessed across the chromosome containing the 2-Mb INV (BTA6:69770000-71780000) and the extended region (BTA6:67770000-73780000) using samples from both gray and nongray cattle. Gray cattle group: comprising 28 samples from Dingjie (*n* = 13), Langkazi (*n* = 14), and Bailang cattle (*n* = 1) in the QTP. Non-gray cattle group: consisting of Dingjie (*n* = 29), Diqing (*n* = 19), and Yushu cattle (*n* = 32) from the QTP. Randomly 8000 SNPs from the VCF file were sampled and haplotype blocks were analyzed using LDBlockShow<sup>[126]</sup> with the parameter “-SeleVar 2”.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

This work was supported by the National Key R&D Program of China (2021YFF1001000 and 2021YFD1200400), the National Natural Science Foundation of China (32372854, 32341054, 32102523 and 32260823), the Key Research and Development Program of Xizang Autonomous Region of China (XZ202301ZY0008N), the Yunnan Expert Workstations (202305AF150156), the China Agriculture Research System of MOF and MARA (CARS-37), the Postdoctoral Fellowship Program of CPSF (GZC20232149), the Postdoctoral Research Project Funding of Shaanxi Province (2023BSHEDZZ132), the Open Project of State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University (2024-KF-02), the

Program of Yunling Scholar and Yunling Cattle Special Program of Yunnan Joint Laboratory of Seeds and Seeding Industry (202205AR070001), the Construction of Yunling Cattle Technology Innovation Center and Industrialization of Achievements (2019ZG007), and Chuxiong Science and Technology Leading Talents (CXKJLJRC2023-07). The authors thank Lijing Tang and Yafei Mao for their valuable suggestions, and the High-Performance Computing (HPC) of Northwest A&F University (NWAUFU) for providing the computing resources.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

X.X., F.W., X.L., S.L. and L.Y. contributed equally to this work. N.C., C.L., W.L., and B.H. designed and supervised the project. X.X., F.W., X.L., S.L. and L.Y. performed most of the analyses. K.Q., R.S., J.L., J.Z., B.W., B.Z., S.Q., L.Z., S.W., C. Luobu, N. Cangjue, D.L., S.S., Z.M., R.L., Z.W., X.Y., and H.C. collected the samples. Y.Z. and L.D. conducted the experiments. X.X., F.W. and N.C. wrote the manuscript with input from all the authors. N.C., J.A.L., J.H., L.X., H.H., W.L., Z.Z., Y.W., Y.G., R.D., Y.H. and X.L. revised the manuscript.

## Data Availability Statement

The data that support the findings of this study are openly available in NCBI at <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1158668?reviewer=k781773a0e211mnj2j9j41a8l>, reference number 1158668.

## Keywords

cattle, coat color, genome assembly, high-altitude adaptation, structural variation

Received: February 20, 2025

Revised: May 7, 2025

Published online: June 5, 2025

- [1] Q. Qiu, G. Zhang, T. Ma, W. Qian, J. Wang, Z. Ye, C. Cao, Q. Hu, J. Kim, D. M. Larkin, L. Auvil, B. Capitanu, J. Ma, H. A. Lewin, X. Qian, Y. Lang, R. Zhou, L. Wang, K. Wang, J. Xia, S. Liao, S. Pan, X. Lu, H. Hou, Y. Wang, X. Zang, Y. Yin, H. Ma, J. Zhang, Z. Wang, et al., *Nat. Genet.* **2012**, *44*, 946.
- [2] G. D. Wang, R. X. Fan, W. Zhai, F. Liu, L. Wang, L. Zhong, H. Wu, H. C. Yang, S. F. Wu, C. L. Zhu, Y. Li, Y. Gao, R. L. Ge, C. I. Wu, Y. P. Zhang, *Genome Biol Evol* **2014**, *6*, 2122.
- [3] X. Gou, Z. Wang, N. Li, F. Qiu, Z. Xu, D. Yan, S. Yang, J. Jia, X. Kong, Z. Wei, S. Lu, L. Lian, C. Wu, X. Wang, G. Li, T. Ma, Q. Jiang, X. Zhao, J. Yang, B. Liu, D. Wei, H. Li, J. Yang, Y. Yan, G. Zhao, X. Dong, M. Li, W. Deng, J. Leng, C. Wei, et al., *Genome Res.* **2014**, *24*, 1308.
- [4] X. Liu, Y. Zhang, Y. Li, J. Pan, D. Wang, W. Chen, Z. Zheng, X. He, Q. Zhao, Y. Pu, W. Guan, J. Han, L. Orlando, Y. Ma, L. Jiang, *Mol. Biol. Evol.* **2019**, *36*, 2591.
- [5] M. Li, S. Tian, L. Jin, G. Zhou, Y. Li, Y. Zhang, T. Wang, C. K. Yeung, L. Chen, J. Ma, J. Zhang, A. Jiang, J. Li, C. Zhou, J. Zhang, Y. Liu, X. Sun, H. Zhao, Z. Niu, P. Lou, L. Xian, X. Shen, S. Liu, S. Zhang, M. Zhang, L. Zhu, S. Shuai, L. Bai, G. Tang, H. Liu, et al., *Nat. Genet.* **2013**, *45*, 1431.
- [6] M. S. Wang, Y. Li, M. S. Peng, L. Zhong, Z. J. Wang, Q. Y. Li, X. L. Tu, Y. Dong, C. L. Zhu, L. Wang, M. M. Yang, S. F. Wu, Y. W. Miao, J. P. Liu, D. M. Irwin, W. Wang, D. D. Wu, Y. P. Zhang, *Mol. Biol. Evol.* **2015**, *32*, 1880.
- [7] D. D. Wu, C. P. Yang, M. S. Wang, K. Z. Dong, D. W. Yan, Z. Q. Hao, S. Q. Fan, S. Z. Chu, Q. S. Shen, L. P. Jiang, Y. Li, L. Zeng, H. Q. Liu, H. B. Xie, Y. F. Ma, X. Y. Kong, S. L. Yang, X. X. Dong, A. Esmailzadeh, D. M. Irwin, X. Xiao, M. Li, Y. Dong, W. Wang, P. Shi, H. P. Li, Y. H. Ma, X. Gou, Y. B. Chen, Y. P. Zhang, *Natl. Sci. Rev.* **2020**, *7*, 952.
- [8] W. Zhang, Z. Fan, E. Han, R. Hou, L. Zhang, M. Galaverni, J. Huang, H. Liu, P. Silva, P. Li, J. P. Pollinger, L. Du, X. Zhang, B. Yue, R. K. Wayne, Z. Zhang, *PLoS Genet.* **2014**, *10*, 1004466.
- [9] T. S. Simonson, Y. Yang, C. D. Huff, H. Yun, G. Qin, D. J. Witherspoon, Z. Bai, F. R. Lorenzo, J. Xing, L. B. Jorde, J. T. Prchal, R. Ge, *Science* **2010**, *329*, 72.
- [10] R. T. Loftus, D. E. MacHugh, D. G. Bradley, P. M. Sharp, P. Cunningham, *Proc Natl Acad Sci U S A* **1991**, *88*, 2757.
- [11] J.-D. Vigne, J. Peters, D. Helmer, *The first steps of animal domestication*, Oxbow Books, Oxford **2005**.
- [12] J. E. Decker, S. D. McKay, M. M. Rolf, J. Kim, A. Molina Alcalá, T. S. Sonstegard, O. Hanotte, A. Götherström, C. M. Seabury, L. Praharani, M. E. Babar, L. Correia de Almeida Regitano, M. A. Yildiz, M. P. Heaton, W. S. Liu, C. Z. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel, J. F. Taylor, *PLoS Genet.* **2014**, *10*, 1004254.
- [13] N. Chen, Y. Cai, Q. Chen, R. Li, K. Wang, Y. Huang, S. Hu, S. Huang, H. Zhang, Z. Zheng, W. Song, Z. Ma, Y. Ma, R. Dang, Z. Zhang, L. Xu, Y. Jia, S. Liu, X. Yue, W. Deng, X. Zhang, Z. Sun, X. Lan, J. Han, H. Chen, D. G. Bradley, Y. Jiang, C. Lei, *Nat. Commun.* **2018**, *9*, 2337.
- [14] D. D. Wu, X. D. Ding, S. Wang, J. M. Wójcik, Y. Zhang, M. Tokarska, Y. Li, M. S. Wang, O. Faruque, R. Nielsen, Q. Zhang, Y. P. Zhang, *Nat Ecol Evol* **2018**, *2*, 1139.
- [15] I. Medugorac, A. Graf, C. Grohs, S. Rothhammer, Y. Zagdsuren, E. Gladyr, N. Zinovieva, J. Barbieri, D. Seichter, I. Russ, A. Eggen, G. Hellenenthal, G. Brem, H. Blum, S. Krebs, A. Capitan, *Nat. Genet.* **2017**, *49*, 470.
- [16] K. Zhang, J. A. Lenstra, S. Zhang, W. Liu, J. Liu, *Anim Genet* **2020**, *51*, 637.
- [17] X. T. Xia, A. Achilli, J. A. Lenstra, B. Tong, Y. Ma, Y. Z. Huang, J. L. Han, Z. Y. Sun, H. Chen, C. Z. Lei, S. M. Hu, N. B. Chen, *Heredity (Edinb)* **2021**, *126*, 1000.
- [18] N. Chen, X. Xia, Q. Hanif, F. Zhang, R. Dang, B. Huang, Y. Lyu, X. Luo, H. Zhang, H. Yan, S. Wang, F. Wang, J. Chen, X. Guan, Y. Liu, S. Li, L. Jin, P. Wang, L. Sun, J. Zhang, J. Liu, K. Qu, Y. Cao, J. Sun, Y. Liao, Z. Xiao, M. Cai, L. Mu, A. Z. Siddiki, M. Asif, et al., *Nat. Commun.* **2023**, *14*, 7803.
- [19] Y. Lyu, F. Wang, H. Cheng, J. Han, R. Dang, X. Xia, H. Wang, J. Zhong, J. A. Lenstra, H. Zhang, J. Han, D. E. MacHugh, I. Medugorac, M. Upadhyay, A. S. Leonard, H. Ding, X. Yang, M. S. Wang, S. Quji, B. Zhuzha, P. Quzhen, S. Wangmu, N. Cangjue, D. Wa, W. Ma, J. Liu, J. Zhang, B. Huang, X. Qi, F. Li, et al., *Sci. Bull. (Beijing)* **2024**, *69*, 3415.
- [20] Y. He, H. Lou, C. Cui, L. Deng, Y. Gao, W. Zheng, Y. Guo, X. Wang, Z. Ning, J. Li, B. Li, C. Bai, S. Liu, T. Wu, S. Xu, X. Qi, B. Su, *Natl. Sci. Rev.* **2020**, *7*, 391.
- [21] C. Chiang, A. J. Scott, J. R. Davis, E. K. Tsang, X. Li, Y. Kim, T. Hadzic, F. N. Damani, L. Ganel, S. B. Montgomery, A. Battle, D. F. Conrad, I. M. Hall, *Nat. Genet.* **2017**, *49*, 692.
- [22] C. Quan, Y. Li, X. Liu, Y. Wang, J. Ping, Y. Lu, G. Zhou, *Genome Biol.* **2021**, *22*, 159.
- [23] J. A. Lenstra, *Animal Research and One Health* **2024**, *2*, 360.
- [24] X. Li, Q. Liu, C. Fu, M. Li, C. Li, X. Li, S. Zhao, Z. Zheng, *J Genet Genomics* **2024**, *51*, 394.
- [25] H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, *Nat. Methods* **2021**, *18*, 170.

- [26] B. D. Ondov, T. J. Treangen, P. Melsted, A. B. Mallonee, N. H. Bergman, S. Koren, A. M. Phillippy, *Genome Biol.* **2016**, *17*, 132.
- [27] A. Rhie, B. P. Walenz, S. Koren, A. M. Phillippy, *Genome Biol.* **2020**, *21*, 245.
- [28] H. Lou, Y. Gao, B. Xie, Y. Wang, H. Zhang, M. Shi, S. Ma, X. Zhang, C. Liu, S. Xu, *Cell Syst* **2022**, *13*, 321.
- [29] S. Schiffels, R. Durbin, *Nat. Genet.* **2014**, *46*, 919.
- [30] T. Jiang, Y. Liu, Y. Jiang, J. Li, Y. Gao, Z. Cui, Y. Liu, B. Liu, Y. Wang, *Genome Biol.* **2020**, *21*, 189.
- [31] D. Heller, M. Vingron, *Bioinformatics* **2019**, *35*, 2907.
- [32] F. J. Sedlazeck, P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von Haeseler, M. C. Schatz, *Nat. Methods* **2018**, *15*, 461.
- [33] K. Wang, M. Li, H. Hakonarson, *Nucleic Acids Res.* **2010**, *38*, 164.
- [34] X. Dai, P. Bian, D. Hu, F. Luo, Y. Huang, S. Jiao, X. Wang, M. Gong, R. Li, Y. Cai, J. Wen, Q. Yang, W. Deng, H. A. Nanaei, Y. Wang, F. Wang, Z. Zhang, B. D. Rosen, R. Heller, Y. Jiang, *Genome Res.* **2023**, *33*, 1284.
- [35] C. J. Kelly, C. G. Chitko-McKown, E. B. Chuong, *Genome Res.* **2022**, *32*, 1474.
- [36] A. S. Leonard, X. M. Mapel, H. Pausch, *Genome Res.* **2024**, *34*, 300.
- [37] R. Li, M. Gong, X. Zhang, F. Wang, Z. Liu, L. Zhang, Q. Yang, Y. Xu, M. Xu, H. Zhang, Y. Zhang, X. Dai, Y. Gao, Z. Zhang, W. Fang, Y. Yang, W. Fu, C. Cao, P. Yang, Z. A. Ghanatsaman, N. J. Negari, H. A. Nanaei, X. Yue, Y. Song, X. Lan, W. Deng, X. Wang, C. Pan, R. Xiang, E. M. Ibeagha-Awemu, et al., *Genome Res.* **2023**, *33*, 463.
- [38] Z. Li, X. Liu, C. Wang, Z. Li, B. Jiang, R. Zhang, L. Tong, Y. Qu, S. He, H. Chen, Y. Mao, Q. Li, T. Pook, Y. Wu, Y. Zan, H. Zhang, L. Li, K. Wen, Y. Chen, *Genome Res.* **2023**, *33*, 1833.
- [39] D. Crysanto, A. S. Leonard, Z. H. Fang, H. Pausch, *Proc Natl Acad Sci U S A* **2021**, *118*, 2101056118.
- [40] A. Talenti, J. Powell, J. D. Hemmink, E. A. J. Cook, D. Wragg, S. Jayaraman, E. Paxton, C. Ezeasor, E. T. Obishakin, E. R. Agusi, A. Tijjani, K. Marshall, A. Fisch, B. R. Ferreira, A. Qasim, U. Chaudhry, P. Wiener, P. Toye, L. J. Morrison, T. Connelley, J. G. D. Prendergast, *Nat. Commun.* **2022**, *13*, 910.
- [41] P. Ebert, P. A. Audano, Q. Zhu, B. Rodriguez-Martin, D. Porubsky, M. J. Bonder, A. Sulovari, J. Ebler, W. Zhou, R. Serra Mari, F. Yilmaz, X. Zhao, P. Hsieh, J. Lee, S. Kumar, J. Lin, T. Rausch, Y. Chen, J. Ren, M. Santamarina, W. Höps, H. Ashraf, N. T. Chuang, X. Yang, K. M. Munson, A. P. Lewis, S. Fairley, L. J. Tallon, W. E. Clarke, A. O. Basile, et al., *Science* **2021**, *372*, abf7117.
- [42] T. T. Li, T. Xia, J. Q. Wu, H. Hong, Z. L. Sun, M. Wang, F. R. Ding, J. Wang, S. Jiang, J. Li, J. Pan, G. Yang, J. N. Feng, Y. P. Dai, X. M. Zhang, T. Zhou, T. Li, *Nat. Commun.* **2023**, *14*, 6601.
- [43] J. M. Akey, A. L. Ruhe, D. T. Akey, A. K. Wong, C. F. Connelly, J. Madeoy, T. J. Nicholas, M. W. Neff, *Proc Natl Acad Sci U S A* **2010**, *107*, 1160.
- [44] Y. Feng, N. Xie, F. Inoue, S. Fan, J. Saskin, C. Zhang, F. Zhang, M. E. B. Hansen, T. Nyambo, S. W. Mpoloka, G. G. Mokone, C. Fokunang, G. Belay, A. K. Njamnshi, M. S. Marks, E. Oancea, N. Ahituv, S. A. Tishkoff, *Nat. Genet.* **2024**, *56*, 258.
- [45] W. Zheng, Y. He, Y. Guo, T. Yue, H. Zhang, J. Li, B. Zhou, X. Zeng, L. Li, B. Wang, J. Cao, L. Chen, C. Li, H. Li, C. Cui, C. Bai, Baimakangzhuo, X. Qi, Ouzhuluobu, B. Su, *Genome Biol.* **2023**, *24*, 73.
- [46] R. A. Torres, D. A. Drake, V. Solodushko, R. Jadhav, E. Smith, P. Rocic, D. S. Weber, *Arterioscler Thromb Vasc Biol* **2011**, *31*, 2424.
- [47] Y. Wang, X. Liu, B. Xie, H. Yuan, Y. Zhang, J. Zhu, *Redox Biol.* **2020**, *28*, 101313.
- [48] L. Cocco, L. Manzoli, I. Faenza, G. Ramazzotti, Y. R. Yang, J. A. McCubrey, P. G. Suh, M. Y. Follo, *Adv. Biol. Regul.* **2016**, *60*, 1.
- [49] M. A. S. Pinhal, C. M. Melo, H. B. Nader, *Adv. Exp. Med. Biol.* **2020**, *1221*, 821.
- [50] M. Elkin, N. Ilan, R. Ishai-Michaeli, Y. Friedmann, O. Papo, I. Pecker, I. Vlodavsky, *FASEB J.* **2001**, *15*, 1661.
- [51] P. Zhao, F. Zhao, J. Hu, J. Wang, X. Liu, Z. Zhao, Q. Xi, H. Sun, S. Li, Y. Luo, *Front Physiol* **2022**, *13*, 885444.
- [52] S. Zhu, T. Guo, H. Zhao, G. Qiao, M. Han, J. Liu, C. Yuan, T. Wang, F. Li, Y. Yue, B. Yang, *Front. Genet.* **2020**, *11*, 848.
- [53] J. Hu, M. R. Hameed, N. P. Agaram, K. A. Whiting, L. X. Qin, A. M. Villano, R. B. O'Connor, J. M. Rozenberg, S. Cohen, K. Prendergast, S. Kryeziu, R. L. White, Jr., M. C. Posner, N. D. Socci, M. M. Gounder, S. Singer, A. M. Crago, *Clin. Cancer Res.* **2024**, *30*, 450.
- [54] I. H. Jain, S. E. Calvo, A. L. Markhard, O. S. Skinner, T. L. To, T. Ast, V. K. Mootha, *Cell* **2020**, *181*, 716.
- [55] W. H. Tang, S. Wu, T. M. Wong, S. K. Chung, S. S. Chung, *Free Radic Biol Med* **2008**, *45*, 602.
- [56] T. L. Clanton, *J. Appl. Physiol.* **2007**, *102*, 2379.
- [57] Z. Wu, M. Zuo, L. Zeng, K. Cui, B. Liu, C. Yan, L. Chen, J. Dong, F. Shangquan, W. Hu, H. He, B. Lu, Z. Song, *EMBO Rep.* **2021**, *22*, 50827.
- [58] T. Wang, E. Wiater, X. Zhang, J. B. Thomas, M. Montminy, *Proc Natl Acad Sci U S A* **2021**, *118*.
- [59] N. Maeda, T. Funahashi, Y. Matsuzawa, I. Shimomura, *Atherosclerosis* **2020**, *292*, 1.
- [60] Z. Chen, I. L. Ho, M. Soeung, E. Y. Yen, J. Liu, L. Yan, J. L. Rose, S. Srinivasan, S. Jiang, Q. E. Chang, N. Feng, J. P. Gay, Q. Wang, J. Wang, P. L. Lorenzi, L. J. Veillon, B. Wei, J. N. Weinstein, A. K. Deem, S. Gao, G. Genovese, A. Viale, W. Yao, C. A. Lyssiotis, J. R. Marszalek, G. F. Draetta, H. Ying, *Nat. Commun.* **2023**, *14*, 2194.
- [61] J. Han, D. Zheng, P. S. Liu, S. Wang, X. Xie, *Cell Commun Signal* **2024**, *22*, 475.
- [62] H. S. Radomska, C. S. Huettner, P. Zhang, T. Cheng, D. T. Scadden, D. G. Tenen, *Mol. Cell. Biol.* **1998**, *18*, 4301.
- [63] X. Liu, W. Liu, J. A. Lenstra, Z. Zheng, X. Wu, J. Yang, B. Li, Y. Yang, Q. Qiu, H. Liu, K. Li, C. Liang, X. Guo, X. Ma, R. J. Abbott, M. Kang, P. Yan, J. Liu, *Nat. Commun.* **2023**, *14*, 5617.
- [64] C. Li, B. Chen, S. Langda, P. Pu, X. Zhu, S. Zhou, P. Kalds, K. Zhang, M. Bhati, A. Leonard, S. Huang, R. Li, A. Cuoji, X. Wang, H. Zhu, Y. Wu, R. Cuomu, B. Gui, M. Li, Y. Wang, Y. Li, W. Fang, T. Jia, T. Pu, X. Pan, Y. Cai, C. He, L. Wang, Y. Jiang, J. L. Han, et al., *Genomics Proteomics Bioinformatics* **2024**, *22*.
- [65] K. Tsuji-Tamura, M. Ogawa, *Angiogenesis* **2023**, *26*, 523.
- [66] J. Zheng, S. J. Kim, S. Saeidi, S. H. Kim, X. Fang, Y. H. Lee, Y. N. Guillen-Quispe, H. K. C. Ngo, D. H. Kim, D. Kim, Y. J. Surh, *Free Radic Biol Med* **2023**, *194*, 347.
- [67] G. R. Pielberg, A. Golovko, E. Sundström, I. Curik, J. Lennartsson, M. H. Seltenthaler, T. Druml, M. Binns, C. Fitzsimmons, G. Lindgren, K. Sandberg, R. Baumung, M. Vetterlein, S. Strömberg, M. Grabherr, C. Wade, K. Lindblad-Toh, F. Pontén, C. H. Heldin, J. Sölkner, L. Andersson, *Nat. Genet.* **2008**, *40*, 1004.
- [68] H. Shao, D. Ganesamoorthy, T. Duarte, M. D. Cao, C. J. Hoggart, L. J. M. Coin, *BMC Bioinformatics* **2018**, *19*, 261.
- [69] E. Kabirova, A. Ryzhkova, V. Lukyanchikova, A. Khabarova, A. Korablev, T. Shnaider, M. Nuriddinov, P. Belokopytova, A. Smirnov, N. V. Khotskin, G. Kontsevaya, I. Serova, N. Battulin, *Nat. Commun.* **2024**, *15*, 4521.
- [70] W. Y. Low, *Animal Research and One Health* **2024**, *2*, 363.
- [71] W. F. Blazak, F. E. Eldridge, *J. Dairy Sci.* **1977**, *60*, 1133.
- [72] A. S. Leonard, D. Crysanto, Z. H. Fang, M. P. Heaton, B. L. Vander Ley, C. Herrera, H. Bollwein, D. M. Bickhart, K. L. Kuhn, T. P. L. Smith, B. D. Rosen, H. Pausch, *Nat. Commun.* **2022**, *13*, 3012.
- [73] A. Mandefro, T. Sisay, Z. Edea, M. R. Uzzaman, K. S. Kim, H. Dadi, *J Anim Sci Technol* **2021**, *63*, 248.
- [74] S. N. Takeshima, C. Corbi-Botto, G. Giovambattista, Y. Aida, *BMC Genet.* **2018**, *19*, 33.

- [75] C. Li, Y. Wu, B. Chen, Y. Cai, J. Guo, A. S. Leonard, P. Kalds, S. Zhou, J. Zhang, P. Zhou, S. Gan, T. Jia, T. Pu, L. Suo, Y. Li, K. Zhang, L. Li, M. Purevdorj, X. Wang, M. Li, Y. Wang, Y. Liu, S. Huang, T. Sonstegard, M. S. Wang, S. Kemp, H. Pausch, Y. Chen, J. L. Han, Y. Jiang, et al., *Mol. Biol. Evol.* **2022**, 39.
- [76] X. Xia, F. Zhang, S. Li, X. Luo, L. Peng, Z. Dong, H. Pausch, A. S. Leonard, D. Crysanto, S. Wang, B. Tong, J. A. Lenstra, J. Han, F. Li, T. Xu, L. Gu, L. Jin, R. Dang, Y. Huang, X. Lan, G. Ren, Y. Wang, Y. Gao, Z. Ma, H. Cheng, Y. Ma, H. Chen, W. Pang, C. Lei, N. Chen, *Genome Biol.* **2023**, 24, 211.
- [77] L. Deng, C. Zhang, K. Yuan, Y. Gao, Y. Pan, X. Ge, Y. He, Y. Yuan, Y. Lu, X. Zhang, H. Chen, H. Lou, X. Wang, D. Lu, J. Liu, L. Tian, Q. Feng, A. Khan, Y. Yang, Z. B. Jin, J. Yang, F. Lu, J. Qu, L. Kang, B. Su, S. Xu, *Natl. Sci. Rev.* **2019**, 6, 1201.
- [78] Q. Qiu, L. Wang, K. Wang, Y. Yang, T. Ma, Z. Wang, X. Zhang, Z. Ni, F. Hou, R. Long, R. Abbott, J. Lenstra, J. Liu, *Nat. Commun.* **2015**, 6, 10283.
- [79] G. L. Semenza, E. A. Rue, N. V. Iyer, M. G. Pang, W. G. Kearns, *Genomics* **1996**, 34, 437.
- [80] C. W. Pugh, P. J. Ratcliffe, *Nat. Med.* **2003**, 9, 677.
- [81] J. Si, J. Guo, X. Zhang, W. Li, S. Zhang, S. Shang, Q. Zhang, *Biol. Direct* **2024**, 19, 45.
- [82] B. Borsari, P. Villegas-Mirón, S. Pérez-Lluch, I. Turpin, H. Laayouni, A. Segarra-Casas, J. Bertranpetit, R. Guigó, S. Acosta, *Genome Res.* **2021**, 31, 1325.
- [83] Z. Yang, C. Bai, Y. Pu, Q. Kong, Y. Guo, X. Liu, Q. Zhao, Z. Qiu, W. Zheng, Y. He, Y. Lin, L. Deng, C. Zhang, S. Xu, Y. Peng, K. Xiang, X. Zhang, C. Cui, Y. Pan, J. Xin, Y. Wang, S. Liu, L. Wang, H. Guo, Z. Feng, S. Wang, H. Shi, B. Jiang, T. Wu, X. Qi, et al., *Proc Natl Acad Sci U S A* **2022**, 119, 2200421119.
- [84] J. F. Storz, *Mol. Biol. Evol.* **2021**, 38, 2677.
- [85] X. An, L. Mao, Y. Wang, Q. Xu, X. Liu, S. Zhang, Z. Qiao, B. Li, F. Li, Z. Kuang, N. Wan, X. Liang, Q. Duan, Z. Feng, X. Yang, S. Liu, E. Nevo, J. Liu, J. F. Storz, K. Li, *Nat Ecol Evol* **2024**, 8, 339.
- [86] J. Shi, Z. Jia, J. Sun, X. Wang, X. Zhao, C. Zhao, F. Liang, X. Song, J. Guan, X. Jia, J. Yang, Q. Chen, K. Yu, Q. Jia, J. Wu, D. Wang, Y. Xiao, X. Xu, Y. Liu, S. Wu, Q. Zhong, J. Wu, S. Cui, X. Bo, Z. Wu, M. Park, M. Kellis, K. He, *Nat. Commun.* **2023**, 14, 8282.
- [87] T. Bae, S. P. Hallis, M. K. Kwak, *Exp. Mol. Med.* **2024**, 56, 501.
- [88] X. Ge, Y. Lu, S. Chen, Y. Gao, L. Ma, L. Liu, J. Liu, X. Ma, L. Kang, S. Xu, *Mol. Biol. Evol.* **2023**, 40, msd205.
- [89] I. Said, A. Byrne, V. Serrano, C. Cardeno, C. Vollmers, R. Corbett-Detig, *Proc Natl Acad Sci U S A* **2018**, 115, 5492.
- [90] K. Bilgrav Saether, J. E. Eisfeldt, J. D. Bengtsson, M. Y. Lun, C. M. Grochowski, M. Mahmoud, H. T. Chao, J. A. Rosenfeld, P. Liu, M. Ek, J. Schuy, A. Ameer, H. Dai, J. P. Hwang, F. J. Sedlitz, W. Bi, R. Marom, J. Wincent, A. Nordgren, C. M. B. Carvalho, A. Lindstrand, *Genome Res.* **2024**, 34, 1785.
- [91] D. Nilsson, M. Pettersson, P. Gustavsson, A. Förster, W. Hofmeister, J. Wincent, V. Zachariadis, B. M. Anderlid, A. Nordgren, O. Mäkitie, V. Wirta, M. Käller, F. Vezzi, J. R. Lupski, M. Nordenskjöld, E. S. Lundberg, C. M. B. Carvalho, A. Lindstrand, *Hum Mutat* **2017**, 38, 180.
- [92] J. Guan, Y. Xu, Y. Yu, J. Fu, F. Ren, J. Guo, J. Zhao, Q. Jiang, J. Wei, H. Xie, *Genome Biol.* **2021**, 22, 13.
- [93] O. S. Harringmeyer, H. E. Hoekstra, *Nat Ecol Evol* **2022**, 6, 1965.
- [94] P. J. Hansen, *Anim. Reprod. Sci.* **2004**, 82–83, 349.
- [95] M. D. Littlejohn, K. M. Henty, K. Tiplady, T. Johnson, C. Harland, T. Lopdell, R. G. Sherlock, W. Li, S. D. Lukefahr, B. C. Shanks, D. J. Garrick, R. G. Snell, R. J. Spelman, S. R. Davis, *Nat. Commun.* **2014**, 5, 5861.
- [96] J. Plassais, J. Kim, B. W. Davis, D. M. Karyadi, A. N. Hogan, A. C. Harris, B. Decker, H. G. Parker, E. A. Ostrander, *Nat. Commun.* **2019**, 10, 1489.
- [97] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, *Cell* **2014**, 159, 1665.
- [98] N. C. Durand, M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, *Cell Syst* **2016**, 3, 95.
- [99] O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, E. L. Aiden, *Science* **2017**, 356, 92.
- [100] N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L. Aiden, *Cell Syst* **2016**, 3, 99.
- [101] F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, *Bioinformatics* **2015**, 31, 3210.
- [102] A. Shumate, S. L. Salzberg, *Bioinformatics* **2021**, 37, 1639.
- [103] G. Perteua, M. Perteua, *F1000Res* **2020**, 9.
- [104] N. Patterson, A. L. Price, D. Reich, *PLoS Genet.* **2006**, 2, 190.
- [105] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, P. C. Sham, *Am. J. Hum. Genet.* **2007**, 81, 559.
- [106] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, *Mol. Biol. Evol.* **2011**, 28, 2731.
- [107] D. H. Huson, D. Bryant, *Nat. Methods* **2024**, 21, 1773.
- [108] D. C. Jeffares, C. Jolly, M. Hoti, D. Speed, L. Shaw, C. Rallis, F. Balloux, C. Dessimoz, J. Bähler, F. J. Sedlitz, *Nat. Commun.* **2017**, 8, 14061.
- [109] P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, *Gigascience* **2021**, 10, giab008.
- [110] A. C. English, V. K. Menon, R. A. Gibbs, G. A. Metcalf, F. J. Sedlitz, *Genome Biol.* **2022**, 23, 271.
- [111] S. Chen, P. Krusche, E. Dolzhenko, R. M. Sherman, R. Petrovski, F. Schlesinger, M. Kirsche, D. R. Bentley, M. C. Schatz, F. J. Sedlitz, M. A. Eberle, *Genome Biol.* **2019**, 20, 291.
- [112] P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, *Bioinformatics* **2011**, 27, 2156.
- [113] A. R. Quinlan, *Curr Protoc Bioinformatics* **2014**, 47.
- [114] B. Langmead, S. L. Salzberg, *Nat. Methods* **2012**, 9, 357.
- [115] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, *Bioinformatics* **2009**, 25, 2078.
- [116] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, *Genome Biol.* **2008**, 9, R137.
- [117] L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, *Mol. Biol. Evol.* **2015**, 32, 268.
- [118] S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, *Nat. Methods* **2017**, 14, 587.
- [119] N. Servant, N. Varoquaux, B. R. Lajoie, E. Viara, C. J. Chen, J. P. Vert, E. Heard, J. Dekker, E. Barillot, *Genome Biol.* **2015**, 16, 259.
- [120] O. Ursu, N. Boley, M. Taranova, Y. X. R. Wang, G. G. Yardimci, W. Stafford Noble, A. Kundaje, *Bioinformatics* **2018**, 34, 2701.
- [121] M. J. Rowley, M. H. Nichols, X. Lyu, M. Ando-Kuri, I. S. M. Rivera, K. Hermetz, P. Wang, Y. Ruan, V. G. Corces, *Mol. Cell* **2017**, 67, 837.
- [122] E. Crane, Q. Bian, R. P. McCord, B. R. Lajoie, B. S. Wheeler, E. J. Ralston, S. Uzawa, J. Dekker, B. J. Meyer, *Nature* **2015**, 523, 240.
- [123] A. Kaul, S. Bhattacharyya, F. Ay, *Nat. Protoc.* **2020**, 15, 991.
- [124] X. Zhou, M. Stephens, *Nat. Genet.* **2012**, 44, 821.
- [125] H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, *Brief Bioinform* **2013**, 14, 178.
- [126] S. S. Dong, W. M. He, J. J. Ji, C. Zhang, Y. Guo, T. L. Yang, *Brief Bioinform* **2021**, 22, bbaa227.