# Genome-wide scan and fine-mapping nominate *ANAPC4* as a candidate modulator of Chinese Holstein heifer puberty timing

Cui-li Pan[1,#], Long-gang Ma[1,#], Xue-sha Cao[1], Qi-yan Wang[1], Lu-lu Wang[1], Jiong Wang[1], Ao Wang[1], Ran Li[1], Xi-hong Wang[1] , Yun Ma[2], Guang-lei Liu[3], Kai Zhu[3], Xin-zhe Lv[3], Zhuang-biao Zhang[1,*], Yu Jiang[1,*]

[1] Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A & F University, Yangling 712100, China

[2] College of Animal Science and Technology, Ningxia Key Laboratory of Ruminant Molecular and Cellular Breeding, Ningxia University, Yinchuan, 750021, China

[3] Hebei Pinyuan Biotechnology Co., LTD, Shijiazhuang, 050600, China

[#] These authors contributed equally to this work

[*]Corresponding authors: zhangzhuangbiao18@163.com; yu.jiang@nwafu.edu.cn

## ABSTRACT

Dairy fertility strongly affects farm sustainability because young cows must be raised for many months before they begin producing milk. Age at first service (AFS) is a key driver of reproductive efficiency and rearing costs in dairy systems. We analyzed Chinese Holstein heifers to dissect the genetic architecture of AFS and nominate functional candidates. After low-coverage whole-genome sequencing and genotype imputation, a total of 7 964 217 SNPs were retained for genome-wide association analysis (GWAS) and window-based variance partitioning. Among the seven traits evaluated, AFS exhibited the highest heritability, with a moderate estimate of $h^2$=0.276.





Science Press ZR Zoological Research

Three candidate regions jointly explained 0.91% of the genetic variance. On BTA6, the association peak spans *ZCCHC4* and *ANAPC4*; notably, *ANAPC4* contributes to estrogen-induced oocyte maturation. In blood, cis-expression quantitative trait loci (cis-eQTL) mapping and co-localization analysis provided strong evidence that the GWAS locus colocalizes with *ANAPC4* expression, with the posterior probability of shared causality ($PP_4$) reaching 0.921. SuSiE fine-mapping pinpointed the association to a ~30-kb region comprising a 20-variant credible set that segregates into two high linkage disequilibrium (LD) haplotypes (H0 and H1), with the H1H1 genotype linked to elevated *ANAPC4* expression and a ~2.1-day reduction in AFS. Motif analyses suggested allele-specific TF-binding differences between haplotypes consistent with regulatory effects. Taken together, these findings pinpoint a fine-mapped regulatory haplotype that upregulates *ANAPC4* expression and is associated with earlier AFS, providing actionable markers for genomic selection to advance heifer puberty timing and improve farm profitability.

**Key words:** puberty timing; GWAS; fine-mapping; Chinese Holstein heifers; *ANAPC4*

## INTRODUCTION

AFS is an economically important trait because it marks the onset of a heifer's reproductive life and influences age at first calving (AFC), generation interval, and rearing costs. Although heifers represent the herd's future revenue-generating animals, the non-productive rearing period is a substantial investment with no direct return until first lactation, creating financial pressure for dairy farms. In the United States, average rearing costs were estimated at $2,355 per heifer in total and $2.45 per heifer per day across 26 farms, accounting for as much as 15–20% of total operating expenses (Karszes & Hill, 2020). Strategies that reduce heifer-rearing costs therefore have an outsized impact on dairy profitability.

AFC can be decomposed into three components: AFS, first-to-last insemination interval in

heifers (IFLh), and gestation length in heifers (GLh) (Heise et al., 2018). GLh is relatively fixed—averaging 273 days in Chinese Holstein heifers in this study and 277.8 days in U.S. Holsteins (Norman et al., 2009). By contrast, IFLh reflects sexual maturity and conception ability (Weller et al., 2022). Consequently, the most effective route to optimize AFC is to select heifers that reach puberty earlier while achieving adequate growth, thereby shortening AFS. In line with this rationale, genetic and genomic evaluations have increasingly targeted earlier AFS/AFC and improved insemination/conception probabilities once heifers reach an appropriate size (Ferrari et al., 2024; Kgari et al., 2022; Prakapenka et al., 2023; Stephen et al., 2023).

Because fertility traits in cattle are typically polygenic and environmentally sensitive, GWAS are widely used to discover loci contributing to variation. AFS is a key fertility trait affecting productivity in both beef and dairy systems; however, mapping causal variants has been challenging due to large environmental noise and heterogeneous management. Although several genomic regions and candidate genes for AFC, AFS, and age at puberty have been reported (Prakapenka et al., 2023; Stephen et al., 2023; Weller et al., 2022), cross-population QTL overlap is often limited, underscoring the difficulty of pinpointing shared signals and the need for population-specific studies. To date, candidate regions and genes underlying AFS or age at puberty in Chinese Holstein heifers remain insufficiently characterized.

The growing use of automated activity monitors now provides high-throughput, objective estrus detection and rich behavioral profiles in cattle (Cerri et al., 2021), enabling large-scale and precise AFS phenotyping. Leveraging such phenotypes, the objectives of this study were to: (1) characterize phenotypic distributions and estimate genetic parameters for seven fertility traits in Chinese Holstein heifers; (2) identify genomic regions, variants, and candidate genes influencing AFS using GWAS and integrate evidence from eQTL; and (3) fine-map prioritized candidate

regions. These results provide new insights into the regulation and genetic architecture of AFS in Chinese Holstein heifers and inform selection strategies to improve reproductive efficiency.

## MATERIALS AND METHODS

### Phenotype and genotype data collection and preprocessing

No experimental handling of animals was conducted for this study. Phenotypic and genotypic information was obtained from routine breeding and genomics evaluations in commercial herds. Because only de-identified records and leftover samples collected as part of standard farm management were used, approval from an Institutional Animal Care and Use Committee was not required.

Birth, mating, and calving records of Chinese Holstein heifers collected between 2015 and 2024 were compiled, from which seven fertility-related traits were derived: AFS, AFC, IFLh, conception rate in heifers (CRh), number of services per conception in heifers (NSh), GLh, and birth weight of heifers (BiWh). To illustrate the distributional characteristics of these phenotypes, kernel density estimation (KDE) plots are provided in Supplementary Figure S1. For each trait, outliers were removed.

Genotypes were obtained from 8 574 blood samples subjected to low-coverage whole-genome sequencing (lcWGS; ~1× depth) at MolBreeding Biotech Co., Ltd. Raw reads were quality-filtered with fastp v0.20.0 (Chen et al., 2018) using default parameters and aligned to the bovine reference genome ARS-UCD1.2 with BWA-MEM v0.7.17 (Li & Durbin, 2010). BAM conversion, sorting, and indexing were performed with samtools v1.10 (Li et al., 2009), and PCR duplicates were marked/removed with Picard v2.20.1 (http://broadinstitute.github.io/picard/). Following the previous reference panel and workflow (Zhang et al., 2023), whole-genome imputation yielded 11 106 737 SNPs. Post-imputation quality control excluded variants with minor

allele frequency < 0.05, per-SNP missingness > 0.10, per-sample missingness > 0.10, and Hardy–Weinberg equilibrium $P < 1\times10^{-6}$. After QC, 7,964,217 SNPs remained and were used for downstream analyses, including genetic parameter estimation, GWAS, and fine-mapping (marker density summarized in Supplementary Figure S2).

**Genetic parameter estimation**

Phenotypic and genetic (co)variance components for the heifer fertility traits were estimated by average-information REML (AI-REML) using linear animal models in the BLUPF90 suite (RENUMF90, AIREMLF90, BLUPF90) (Misztal et al., 2014). All genomic analyses used the genomic relationship matrix G computed from post-QC SNPs.

Fixed effects and pre-screening. Candidate fixed effects were farm; year–season (of birth, mating, or calving); sex of calf; gestation-length month; the mating sire; and first-mating month. Years were coded as categorical levels (one level per year), and months were grouped into four seasons (spring, Mar–May; summer, Jun–Aug; autumn, Sep–Nov; winter, Dec–Feb). Fixed effects were pre-screened with a general linear model (GLM), retaining significant terms ($P < 0.05$) for each trait as follows: AFS—farm and year–season of birth; AFC—farm and year–season of mating; IFLh—farm and year–season of mating; NSh—year–season of mating and first-mating month; GL—farm and year–season of calving; BiWh—farm, year–season of calving, sex of calf, and gestation-length.

Univariate animal models were used to estimate heritability and additive genetic variance for each fertility trait separately. Specifically, for each trait we fit:

$$y=\mathbf{X}\beta+\mathbf{Z}a+e, \quad (1)$$

where y is the phenotype vector, β the fixed-effect vector, a the random additive genetic effects, and e the residuals. Distributional assumptions were:

115 $$a \sim N(0,\mathbf{G}\sigma_a^2), \qquad e \sim N(0,\mathbf{I}\sigma_e^2),$$

116 with a and e independent. Heritability was:

117 $$h^2=\frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}\ , (2)$$

118 where $\mathbf{h^2}$ is the heritability, $\boldsymbol{\sigma_a^2}$ is the additive genetic variance, and $\boldsymbol{\sigma_e^2}$ is the residual variance.

119 The square of the $SE^2$ for the heritability estimates was calculated as:

120 $$SE^2(h^2)=\left[\frac{\hat{\sigma}_a^2}{\hat{\sigma}_p^2}\right]^2\left[\frac{\mathrm{Var}(\hat{\sigma}_a^2)}{(\hat{\sigma}_a^2)^2}+\frac{\mathrm{Var}(\hat{\sigma}_p^2)}{(\hat{\sigma}_p^2)^2}-2\frac{\mathrm{Cov}(\hat{\sigma}_a^2,\hat{\sigma}_p^2)}{\hat{\sigma}_a^2\hat{\sigma}_p^2}\right], (3)$$

121 Genetic correlations ($r_g$) between trait pairs were estimated using a bivariate animal model,

122 which jointly fits two traits to partition additive genetic and residual (co)variance components. The

123 model can be expressed as:

124 $$\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}_1 & 0\\ 0 & \boldsymbol{X}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix}+\begin{bmatrix}\boldsymbol{Z}_1 & 0\\ 0 & \boldsymbol{Z}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\end{bmatrix}+\begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix}\ , (4)$$

125 where $y_1$ and $y_2$ are the matrix of phenotypes; $\beta_1$ and $\beta_2$ are the fixed effect vector of the

126 corresponding phenotype; $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are random additive genetic effect vectors of the

127 corresponding phenotype; $e_1$ and $e_2$ are the random residual effect vector for the first and second

128 phenotypes; X, Z are incidence matrices linking the corresponding phenotypic records to β and a,

129 respectively.

130 The genetic correlation coefficient was calculated as $\mathbf{r}=\frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_i^2\hat{\sigma}_j^2}}$, and the square of the SE for the

131 genetic correlation coefficient was calculated as:

132 $$SE^2(r)=\left[\frac{\hat{\sigma}_{ij}^2}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\right]\left[\frac{\mathrm{Var}(\hat{\sigma}_{ij})}{(\hat{\sigma}_{ij})^2}+\frac{\mathrm{Var}(\hat{\sigma}_i^2)}{4(\hat{\sigma}_i^2)^2}+\frac{\mathrm{Var}(\hat{\sigma}_j^2)}{4(\hat{\sigma}_j^2)^2}-\frac{\mathrm{Cov}(\hat{\sigma}_{ij},\hat{\sigma}_i^2)}{\hat{\sigma}_{ij}\hat{\sigma}_i^2}-\frac{\mathrm{Cov}(\hat{\sigma}_{ij},\hat{\sigma}_j^2)}{\hat{\sigma}_{ij}\hat{\sigma}_j^2}+\frac{\mathrm{Cov}(\hat{\sigma}_i^2,\hat{\sigma}_j^2)}{2\hat{\sigma}_i^2\hat{\sigma}_j^2}\right], (5)$$

133 The variances and covariances of the estimated parameters $(\sigma_i^2,\sigma_{ij})$ were obtained from the

134 inverse of the average information matrix.

**Genome-wide association study**

Regional variance partitioning. We quantified how much additive genetic variance for AFS is captured by local genomic regions using POSTGSF90 (BLUPF90 family) (Aguilar et al., 2019). First, the single-trait genomic animal model described above was fitted to obtain genomic breeding values and the total additive variance $\hat{\boldsymbol{\sigma}}^2_{\mathbf{a}}$. SNP effects were back-solved and aggregated into consecutive, non-overlapping 200-kb windows, which was chosen from LD-decay analysis in this population (Figure 1D).

For window $i$ with $m_i$ SNPs, let $\boldsymbol{Z}_{ij}$ be the centered genotype vector (0/1/2 coded as gene content and centered by $2p_j$) for SNP $j$ and $\alpha_{ij}$ its effect. The genetic value contributed by window $i$ is

$$a_i=\sum_{j=1}^{m_i} Z_{ij}\,\alpha_{ij}\,. \quad (6)$$

The proportion of additive genetic variance explained by window $i$ was

$$PropVar_i=\frac{Var(a_i)}{\hat{\sigma}^2_{\mathrm{a}}}\times 100\ \%, \quad (7)$$

Results are reported as the percentage of additive variance per 200-kb window and visualized in R 4.3.0 (ggplot2).

Single-marker association testing. Genome-wide single-SNP association was performed with a linear mixed model (LMM) that incorporates a genomic relationship matrix (GRM) to correct for kinship among individuals and to help control population stratification:

$$y=X\beta+x_j\alpha_j+\mu+\mathrm{e}, \quad (8)$$

Where $x_j$ is the centered genotype vector for SNP $j$ and $\alpha_j$ its allelic substitution effect; $\mathrm{u}\sim N(0,\mathbf{G}\sigma^2_{\mathbf{u}})$, $\mathrm{e}\sim N(0,\mathbf{I}\sigma^2_{\mathbf{e}})$. SNP effects and standard errors were obtained from the inverse mixed-model equations (BLUPF90/POSTGSF90), and Wald statistics $t_j=\hat{\alpha}_j/SE(\hat{\alpha}_j)$ were converted to two-sided $P$-values. We assessed test-statistic calibration using the genomic inflation factor $\lambda=\mathrm{median}(\chi^2)/0.4549$.

To avoid overly conservative Bonferroni correction under LD, we estimated the effective number of independent tests $N_{eff}$ after genotype QC by LD pruning with PLINK v1.9 (window size 200 kb; LD threshold $r^2$=0.2). The number of variants retained was 86,360. We therefore set

$$P_{GW} = \frac{0.05}{N_{eff}} = \frac{0.05}{86{,}360} = 5.79 \times 10^{-7}, \quad (9)$$

and, for suggestive evidence,

$$P_{SG} = \frac{1}{N_{eff}} = \frac{1}{86{,}360} = 1.16 \times 10^{-5}, \quad (10)$$

Association testing itself used the full, unpruned SNP set; these LD-adjusted thresholds were applied only when declaring significance and for Manhattan plot reference lines.

**Gene annotation and functional enrichment**

Because the SNP with the smallest *P*-value does not necessarily explain the largest proportion of additive variance, we prioritized candidate regions that combined (i) high window-level variance (200-kb, non-overlapping windows from POSTGSF90) with (ii) supporting single-marker evidence. Specifically, we ranked windows by the percentage of additive genetic variance explained and selected the top five windows for follow-up. Adjacent top windows separated by ≤100 kb were merged into a single locus.

Gene annotation. Genes intersecting each candidate locus were annotated with ANNOVAR (Wang et al., 2010; Yang & Wang, 2015) using the Bos taurus ARS-UCD1.2 gene set (http://bovinegenome.org). Variant-to-gene consequences were summarized across standard categories (exonic, splicing, UTR, intronic, upstream ≤2 kb, downstream ≤2 kb, intergenic).

Pathway/GO enrichment. Functional enrichment of candidate-locus genes was performed in R 4.3.0 with clusterProfiler (Wu et al., 2021; Yu et al., 2012), using KEGG and GO (BP/CC/MF). Gene IDs were mapped via org.Bt.eg.db; the background was defined as all genes overlapping any 200-kb window evaluated in the genome scan. Enrichment used a hypergeometric test with

Benjamini–Hochberg correction; terms with FDR < 0.05 were considered significant.

Single gene set enrichment analysis (GSEA). We anchored the analysis on a target gene by ranking all genes according to their Pearson correlation with its expression across samples; using the Bioconductor package GSEABase to build a GeneSetCollection (MSigDB), we computed classical GSEA enrichment scores on this ranked list and assessed significance via permutation to obtain NES and FDR, reporting both positively and negatively enriched pathways.

**Transcriptome sequencing and analysis**

Library preparation and RNA sequencing. Whole blood was collected and total RNA was extracted with TRIzol following the manufacturer's instructions. RNA concentration, purity, and integrity were evaluated using a NanoDrop 2000, the Qubit RNA BR Assay Kit (Cat. Q10211, Invitrogen/Thermo Fisher), and 1% agarose gel electrophoresis. High-quality RNA was used to construct cDNA libraries with the NEBNext Ultra RNA Library Prep Kit for Illumina (Cat. E7530S, New England Biolabs). Sequencing was performed on an Illumina NovaSeq 6000, generating 150-bp paired-end reads.

Quality control, alignment, and quantification. Raw reads were assessed with FastQC v0.11.9 and trimmed using Trim Galore v0.6.6 to remove adapters and low-quality bases/reads. Clean reads were aligned to the Bos taurus ARS-UCD1.2 reference using STAR v2.7.11a (Dobin et al., 2013). Gene abundances were quantified with RSEM v1.3.3 (Li & Dewey, 2011). Principal component analysis (PCA) of expression was performed in R using FactoMineR v2.9. Differentially expressed genes (DEGs) were identified with limma v3.19 (Ritchie et al., 2015), applying Benjamini-Hochberg correction with adjusted $P < 0.05$ and |fold change| > 1.2. Additional expression summaries, k-means clustering, and pairwise comparisons were conducted and visualized in R v4.3.0.

Cis expression quantitative trait locus (cis-eQTL) analysis. We mapped eQTLs using tensorQTL (PyTorch backend). Autosomal genotypes were provided in PLINK2 format, and gene-level expression was supplied as a BED file after quantile normalization (QN) across 648 samples. Covariates included 10 genetic principal components (PCs), 10 PEER factors, and age; genotype, expression, and covariate matrices were aligned by sample ID. For cis-eQTL discovery, we tested variants within ±1 Mb of each gene (tensorQTL default) and obtained gene-level empirical *P* values with cis.map_cis. Multiple-testing across genes was controlled using Storey's q-value procedure (FDR=0.05) to define eGenes. For each eGene, independent signals within the cis window were then identified by stepwise conditional analysis with cis.map_independent (FDR=0.05), yielding a nonredundant set of lead variants per gene.

**Fine mapping of the candidate region**

Fine-mapping with SuSiE. We fine-mapped GWAS and eQTL loci using the Sum of Single Effects (SuSiE) regression implemented in susieR. Variants were restricted to autosomal, biallelic SNPs passing QC (MAF ≥ 0.01, INFO ≥ 0.8, missingness ≤ 2%) and aligned to the bovine reference genome ARS-UCD1.2. LD was computed from the discovery cohort genotypes (dosages) used in GWAS; per-locus correlation matrices (R) were constructed after sample/variant alignment and made positive-definite by mild shrinkage, then scaled to correlations. We applied susie_rss to summary statistics (per-variant $z=\beta/\mathrm{SE}$) and the corresponding LD matrix $R$. We set the maximum number of effects L=10, coverage=0.95, estimate_prior_variance=TRUE, residual variance fixed at 1, tol= $1\times10^{-3}$, and max_iter=2000. SuSiE returns posterior inclusion probabilities (PIP) and credible sets (CS) for each independent signal.

Colocalization of GWAS locus and eQTL. We tested whether the GWAS signal at the *ANAPC4* locus shares a causal variant with the gene's cis-eQTL using the COLOC framework (R

227 package coloc). For colocalization, we assembled per-variant summary statistics for (i) the GWAS
228 (effect size $\beta$, standard error, *P* value, minor-allele frequency, sample size) and (ii) cis-eQTLs for
229 *ANAPC4* derived from our tensorQTL pipeline (slope, slope_se, *P* value, MAF, *N*). Variants were
230 harmonized to a common reference build and effect-allele orientation. We restricted analysis to
231 autosomal, biallelic SNPs with MAF $\geq$ 0.01 (and imputation INFO $\geq$ 0.8 where applicable). When
232 the trait was consistent with a single causal signal in the region, we applied coloc.abf (approximate
233 Bayes factors) with default priors. We report posterior probabilities $PP_0$–$PP_4$, where $PP_4$ quantifies
234 support for a shared causal variant. We interpreted $PP_4 \geq 0.8$ as strong evidence of colocalization.
235 To aid interpretation, we also summarized LD between lead variants and verified that the
236 colocalizing set lay within the SuSiE credible intervals.

237 Haplotype analysis and visualization. Within the 30 kb fine-mapped locus, we defined
238 haplotypes by unsupervised clustering of phased haplotype sequences. After standard QC (biallelic
239 SNPs; call rate ≥98%; MAF ≥0.01; HWE $P>1\times10^{-6}$) and phasing, we constructed a haplotype-by-
240 variant matrix (two haplotypes per individual) encoding alternate-allele dosage (0/1). Pairwise
241 haplotype similarity was quantified using Hamming (IBS) distance on the binary matrix. We then
242 performed unsupervised clustering with K=2 (hierarchical agglomerative clustering, Ward's
243 linkage). The two clusters were labeled H0 and H1. Each sample was assigned a diplotype (H0H0,
244 H0H1, H1H1) from the cluster identities of its two phased haplotypes. Haplotype visualization
245 was carried out using an in-house python script.

246 Transcription factor binding prediction. We predicted transcription factor (TF) binding using
247 the MEME Suite (v5.x) on sequences from the fine-mapped locus. For haplotype-aware analyses,
248 we extracted allele-specific FASTA sequences for the H0/H1 haplotypes: 50 bp windows centered
249 on each sentinel/credible variant. For motif scanning and allele-specific effects, we scanned both

haplotypes with FIMO using the union of known (database) and de novo motifs, reporting sites with $P \leq 1\times10^{-4}$ and BH-adjusted q < 0.05. For each variation of the reference or alternative sequence, the binding of its specific transcription factor is defined as a binding site gain/loss event. All analyses used default MEME Suite pseudocounts and convergence settings with FDR control across motifs and windows to account for multiple testing.

Statistical testing and visualization. Analyses were performed in R (v4.x) using ggpubr. Prior to testing, outliers were removed by excluding observations lying outside the range of mean ± 3 standard deviations (SD) for the quantitative variable (e.g., days of AFS or gene expression in TPM) used to compare haplotypes, and rows containing NA values were removed. The haplotype label (H0H0, H0H1, H1H1) was treated as a categorical factor. Between-group comparisons of the outcome across haplotypes used two-sided Wilcoxon rank-sum tests for the following pairwise contrasts. To control for multiple testing, Holm adjustment was applied; adjusted *P*-values were reported, with statistical significance set at α=0.05. For visualization, we generated boxplots stratified by haplotype. This workflow provides both adjusted pairwise significance and distributional summaries.

## RESULTS

### AFS as a GWAS-suitable trait: top-ranking heritability among fertility traits

Descriptive statistics for seven fertility traits shown in Table 1 were calculated after data correction for outliers, where extreme values were excluded. AFS averaged 421.91 ± 21.84 days (range 367–560), with a coefficient of variation (CV) of 5.18%. AFS, AFC, and GLh showed low dispersion (CV < 10%), whereas IFLh, NSh, and CRh were highly variable (CV=186.14%, 60.95%, and 85.07%, respectively).

Variance components and heritabilities are reported in Table 2. AFS showed moderate

heritability ($h^2$=0.276), the highest among the analyzed fertility traits; AFC, IFLh, NSh, CRh, GLh, and BiWh displayed low heritabilities.

Genetic correlations indicated a very strong positive relationship between AFS and AFC ($r_g$=0.93; Supplementary Figure S3), consistent with earlier AFS advancing AFC. AFS showed negative correlations with IFLh and NSh ($r_g$=−0.08 and −0.26), and a low positive correlation with CRh ($r_g$=0.16). Among other traits, NSh and IFLh were moderately positive ($r_g$=0.42), while CRh and NSh were strongly negative ($r_g$=−0.79). Together, the top-ranking heritability of AFS and its favorable genetic relationships support its use as a primary GWAS phenotype in this study.

**Three candidate regions: prioritized by significance and variance explained**

Using the LMM GWAS, we detected 10 suggestive loci for AFS—two on BTA6 and one each on BTA9, BTA10, BTA13, BTA16, BTA18, BTA22, BTA24, and BTA25 (Figure 1A; Supplementary Table S1). The Q–Q plot showed good calibration with a genomic inflation factor λ=1.01, indicating minimal residual stratification (Figure 1B).

To prioritize regions with meaningful effects, we partitioned the additive genetic variance using consecutive 200-kb non-overlapping windows (size chosen from LD-decay analysis; Figure 1D) and computed the proportion of variance explained per window (Figure 1C). In total, 77 windows exceeded 0.1% variance explained (11.2% of all windows; Supplementary Table S2).

Integrating single-marker significance with window-level variance, we retained three candidate regions (containing five 200kb windows) that jointly explain 0.91% of the additive genetic variance. The top region on BTA6, spanning *ZCCHC4-ANAPC4*, explains 0.47% of the additive variance and represents the strongest genome-wide signal.

**_ANAPC4_ as a candidate: involvement in oocyte meiosis and differential expression in low vs. high AFS**

Pathway analysis of candidate-locus genes showed enrichment in progesterone-mediated oocyte maturation, parathyroid hormone synthesis/secretion/action, and selenocompound metabolism (Figure 2A; Supplementary Figure S4). Notably, *ANAPC4* mapped to the progesterone-mediated oocyte maturation pathway, and genes positively correlated with *ANAPC4* expression were significantly enriched for oocyte meiosis (Figure 2B). *ANAPC4* encodes a core subunit of the anaphase-promoting complex/cyclosome (APC/C), an E3 ubiquitin ligase that governs mitotic/meiotic progression, consistent with roles in oocyte maturation and early embryonic development.

To test expression differences, we compared two extreme groups (n=20 per group) that were selected based on the outlier-corrected AFS phenotype and the corresponding GEBV, which were estimated from AFS values adjusted for fixed effects (Figure 2E–F). Using limma (voom) with Benjamini–Hochberg correction (adjusted $P < 0.05$, |FC|>1.2), we identified 1,740 upregulated and 1,777 downregulated genes in the low-AFS group relative to the high-AFS group (Figure 2G; Supplementary Table S3). Five genes located within the GWAS-prioritized regions of BTA6 (*ANAPC4*, *DHX15*, *GUF1*, *SOD3*, *SEPSECS*) were also DEGs; among them, *ANAPC4* was the most significantly upregulated in the low-AFS group.

**A fine mapped haplotype: upregulates _ANAPC4_ expression and associated with AFS**

GWAS signals at the BTA6 peak colocalized with the *ANAPC4* cis-eQTL (Supplementary Table S4), supporting a shared causal mechanism (COLOC $PP_4$=0.921; Figure 3A–C). SuSiE fine-mapping narrowed the association to a ~30-kb interval (Supplementary Tables S5–S6). Within this interval, the variants segregate into two high-LD haplotypes ($r^2 > 0.9$), hereafter referred to as H0

and H1 (Figure 4A–C)**.** Notably, the 20-variant credible set identified in the colocalized region accounts for more than 85% of the cumulative PIP in both the GWAS and the *ANAPC4* eQTL signals (Table 3; Supplementary Figure S5). Motif analysis across the 20 candidates indicated allele-specific TF binding differences between H0 and H1, consistent with regulatory effects (Figure 4D; Supplementary Table S7).

Expression and phenotype aligned with haplotype status: *ANAPC4* expression was ~2.0 TPM higher in H1H1 than H0H0 (Kruskal–Wallis $P$=0.00019; pairwise H0H0 vs H1H1 $P=4.5\times10^{-5}$; Figure 4E), and AFS was shorter by ~2.1 days in H1H1 haplotype (Figure 4F). Together, these results nominate a fine-mapped regulatory haplotype that modulates *ANAPC4* expression at the GWAS locus.

Mechanistically, ANAPC4 is a core subunit of the APC/C E3 ubiquitin ligase complex. It functions as a structural pivot and scaffold that stabilizes the substrate–receptor interface and positions the APC2–APC11 catalytic module. Through CDC20-mediated rotation, ANAPC4 switches APC/C from a low- to a high-activity state, thereby triggering the rapid degradation of Securin and Cyclin B1. This degradation activates Separase, allowing homologous chromosome segregation during meiosis I and sister chromatid separation during meiosis II in oocyte maturation (Figure 4G) (Chang et al., 2014; Hofler et al., 2024).

## DISCUSSION

### AFS as a high-priority fertility trait: leading heritability and strong genetic correlation with AFC

In South African Holsteins, the reported AFS was 16.8 months and AFC 26.7 months (Kgari et al., 2022); in Israeli Holsteins, AFS was 451.6 days (~15.0 months) and AFC 745.1 days (~24.5 months) (Weller et al., 2022); in a Swedish study of purebred Holsteins heifers, AFS was 446.6

days (~14.8 months) (Malchiodi et al., 2014); and in Chinese Holsteins from southern herds, AFS was 467.48 days (~15.5 months) and AFC 781.87 days (~25.7 months) (Zhu et al., 2024). Another investigation reported AFC decreased from 28.1 ± 2.9 to 26.1 ± 3.1 month from 1996 to 2020 in Italian Holsteins (Ferrari et al., 2024). In our study population, AFS averaged 14.0 months (422 days) and AFC averaged 23.8 months (716 days). Both averages are slightly lower than previously reported values, reflecting our stringent data quality control procedures designed to minimize management-related noise. Specifically, extreme records (AFS > 560 days and AFC > 1100 days) were excluded prior to analysis, ensuring that the reported means represent biologically reasonable and well-managed animals. Notably, there remains room for further optimization toward high-performance targets (~21 months AFS and ~24 months AFC) (Hutchison et al., 2017).

To ensure comparability across traits, we therefore treated CRh, NSh and IFLh as continuous variables in the LMM, although some traits are binary or discontinuous traits. This approach is widely adopted in animal genetics, as linear models applied to binary traits on the observed scale provide nearly unbiased estimates and are computationally efficient for large-scale genomic analyses. The heritability of AFS was moderate ($h^2$=0.276±0.025), aligning with several studies: 0.25 in Dutch Friesians (Jansen et al., 1987), 0.299 in Jersey × Red Sindhi (Vinothraj et al., 2016), and 0.22 in Ayrshire/Friesian cows (Mäntysaari et al., 2002); however, it was higher than estimates reported in some Holstein populations (0.02–0.13) (Eghbalsaied, 2011; Jamrozik et al., 2005; Kgari et al., 2022; Raheja et al., 1989) and lower than the 0.58 observed in one Chinese Holstein herd (Hu et al., 2023). By contrast, the other six fertility traits showed low heritabilities (0.001–0.131). In particular, the heritability of AFC ($h^2$ = 0.057) observed here was slightly lower than estimates reported for South African Holsteins (0.08; Kgari et al., 2022), New Zealand multi-breed populations (0.13; Grosshans et al., 1997), and Jersey×Red Sindhi (0.22; Vinothraj et al., 2016),

but higher than that documented for Iranian Holsteins (0.027; Eghbalsaied, 2011).

Heterogeneity in heritability estimates likely reflects differences in management and environment (Mancin et al., 2024; Slagboom et al., 2021), population genetic variation (Notter, 1999), statistical modeling choices (Yadav et al., 2023), potential breed × environment interactions, and data scale/quality.

Genetic correlations further support prioritizing AFS: the AFS–AFC correlation was very high and favorable ($r_g=0.93 \pm 0.03$), consistent with values reported in South African (0.91), Polish (0.98), and Czech (~0.99) (Brzáková et al., 2019; Jagusiak & Zarnecki, 2006; Kgari et al., 2022). These results indicate substantial shared physiological and genetic bases, implying that improvement in AFC can be effectively pursued by selecting on AFS. Accordingly, we focused GWAS on AFS—the trait with the highest heritability in this study—to identify loci that can ultimately inform selection strategies for AFC.

**Pleiotropy of candidate regions: synergistic regulation of growth and reproduction**

To avoid the adverse consequences of premature breeding, heifers must achieve both sexual maturity and sufficient body development before first service. Guidelines typically recommend insemination after reaching specific liveweight/size thresholds, e.g., >370 kg in German Holsteins (mean 406.46±32.90 kg) (Yin & König, 2018), and at approximately 14 months of age with >350 kg body weight and >125 cm withers height in Israel (Weller et al., 2022). In our program, Chinese Holstein heifers were inseminated when >380 kg, >370 days of age, and in estrus based on activity monitoring. Accordingly, AFS is a composite trait integrating growth/development (body weight, stature) and sexual maturity.

We identified three AFS candidate regions (five 200-kb windows). Region 1 on BTA6:43.69–45.01 Mb accounted for ~0.47% of the additive variance for AFS. Prior studies report overlaps

with reproduction traits, including conception rate (Parker Gaddis et al., 2016), calving-to-conception interval (Müller et al., 2017), inseminations per conception and non-return rate (Höglund et al., 2015), and calving ease (Cole et al., 2011). Consistent with the growth component of AFS, the same interval harbored QTL for birth weight (n=14), body weight gain (n=5), birth body length (n=2), and yearling weight (n=1) (Gutiérrez Gil et al., 2012; Lu et al., 2013; Snelling et al., 2010).

Region 2 chiefly overlapped QTL for weight-related traits (birth, weaning, yearling weights; body weight gain) (Snelling et al., 2010), further linking AFS to growth trajectories. Region 3 contained a QTL related to anti-Müllerian hormone (AMH)—a granulosa-cell hormone that limits follicle recruitment and shapes reproductive-tract development (Gobikrushanth et al., 2018). Elevated AMH is often interpreted as indicating a larger ovarian follicle reserve, suggesting that this region may influence oocyte pool size and ductal differentiation.

Taken together, these overlaps are consistent with shared genetic architecture between growth and reproduction and may reflect pleiotropy or tight linkage. From a breeding perspective, selection for earlier, healthy AFS should be balanced against correlated responses in body size and reproductive physiology within multi-trait indices. Although a reduction of 2.1 days may seem small, its cumulative economic benefit at the scale of tens of thousands of animals is significant.

**A *ZCCHC4–ANAPC4* regulatory haplotype linking puberty timing and growth**

The candidate haplotype influencing AFS spans *ZCCHC4-ANAPC4*. Integrating GWAS, colocalization, and fine-mapping evidence, we nominate *ANAPC4* as the likely effector gene for pubertal timing in Holstein heifers. Mammalian oocytes arrest at prophase I until luteinizing hormone (LH) triggers meiotic resumption; as a core subunit of the APC/C E3 ligase, ANAPC4 participates in the tightly regulated transitions that govern oocyte maturation (Furuta et al., 2000;

Hofler et al., 2024). Thus, regulatory variants that increase *ANAPC4* expression could plausibly advance meiosis and puberty onset. Consistent with this model, human studies have associated *ANAPC4* with age at menarche and age at first sexual intercourse (Kichaev et al., 2019; Mills et al., 2021). In our data, the fine-mapped haplotype upregulates *ANAPC4* and associates with earlier AFS, supporting a cis-regulatory mechanism at the GWAS peak.

Although the *ANAPC4* eQTL was initially identified using blood as a proxy tissue, regulatory effects underlying AFS are likely to be mediated in reproductive tissues. Consistent with this notion, expression profiling using cattleGTEx revealed that *ANAPC4* is most highly expressed in oocytes and ovarian granulosa cells (Supplementary Figure S6), supporting a potential role in female reproductive biology and oocyte meiotic maturation. Consistent ATAC-seq peaks were observed across germinal vesicle oocytes, oocytes, and inner cell mass samples, indicating stable chromatin accessibility during key reproductive and early developmental stages (Supplementary Figure S7). The accessible region colocalizes with H3K4me3 ChIP-seq signals in inner cell mass, supporting the presence of an active promoter or proximal regulatory element. In contrast, no corresponding enrichment of the repressive histone mark H3K27me3 was detected in this region. Together, these epigenomic features provide functional context supporting the relevance of the *ANAPC4* to reproductive timing. Moreover, single-cell transcriptomic analyses of our in-house reproductive tissues (ovary, uterus, uterine horn, and oviduct) revealed enriched *ANAPC4* expression in proliferating T cells, lymphatic endothelial cells, and epithelial cells (Supplementary Figure S8). These expression patterns align with the established role of ANAPC4 as a core component of the anaphase-promoting complex/cyclosome (APC/C), which governs cell-cycle progression and proliferative activity. Collectively, these observations suggest that variation at the *ANAPC4* locus may influence reproductive traits through regulation of proliferative and

433 developmental processes within reproductive tissues.

434 Because AFS is composite, encompassing both puberty and growth, we also considered
435 *ZCCHC4*. Although the haplotype showed no detectable effect on *ZCCHC4* expression in blood,
436 interrogation of cattleGTEx indicated higher *ZCCHC4* expression in embryo and muscle for
437 H1H1/H1H0 versus H0H0 (Supplementary Figure S9), suggesting tissue-specific regulation
438 relevant to growth. Functionally, *ZCCHC4* encodes an S-adenosylmethionine–dependent rRNA
439 N6-adenosine ($m^6A$) methyltransferase that installs $m^6A$ at 28S-A4220 (helix 65), promoting
440 large-subunit maturation, ribosome assembly, and translation efficiency; loss of this modification
441 rewires translation and suppresses proliferation. Consistent with growth links, *ZCCHC4* has been
442 associated with body mass index (BMI) in humans (Schoeler et al., 2023; Zhu et al., 2020), and a
443 missense variant (rs208065122) in *ZCCHC4* associates with carcass weight and eye muscle area
444 in Hanwoo cattle (Bhuiyan et al., 2018). Moreover, k-means clustering placed *ZCCHC4* with
445 *ANAPC4*, *DHX15*, and *CCDC149* (Cluster 2; Supplementary Figure S10), hinting at coordinated
446 regulation within the locus. Our results suggest a tissue-specific model in which the *ZCCHC4-*
447 *ANAPC4* haplotype contributes to puberty timing via *ANAPC4* and may affect growth via
448 *ZCCHC4*. These observations indicate shared regulatory architecture and possible pleiotropy
449 within the locus, but we do not provide evidence for epistasis, which requires non-additive
450 interactions between distinct loci. Testing epistasis will be a direction for the future work.

451 **CONCLUSION**

452 We mapped three genomic regions associated with AFS in Chinese Holstein heifers that jointly
453 explain 0.91% of the additive genetic variance. Fine-mapping pinpointed a regulatory haplotype
454 at the *ZCCHC4–ANAPC4* locus: carriers showed ~2.0 TPM higher *ANAPC4* expression and ~2.1
455 days earlier AFS, implicating *ANAPC4* as a mechanistic candidate for puberty timing. QTL

overlap and expression patterns suggest pleiotropic, tissue-specific effects spanning reproduction and growth. These findings provide actionable markers for genomic selection to advance heifer puberty, shorten the non-productive rearing period, and enhance farm profitability in Chinese Holsteins.

## DATA AVAILABILITY

The lcWGS data, which require genotype imputation and involve a large sample size, are currently available only as derived VCF files. The raw RNA-seq dataset mainly analysed during the current study is available in the NCBI SRA repository under Accession Number PRJNA1397547. We also uploaded the processed transcriptome expression matrices in figshare (https://figshare.com/s/8a25ba61fd527dfdfba0). All aforementioned data are available from the corresponding author upon reasonable request during peer review.

## SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

## COMPETING INTERESTS

The authors declare that we have no competing interests.

## AUTHORS' CONTRIBUTIONS

YJ and ZBZ conceived and designed the study; LLW, GLL, KZ, and XZL collected the samples; CLP, LGM, XSC, JW, QYW, and AW analyzed the sequencing results, performed GWAS and designed the charts and tables. CLP wrote the manuscript. YJ, ZBZ, XHW, YM, RL provided revisions to the manuscript. All authors read and approved the final version of the manuscript.

## ACKNOWLEDGMENTS

We thank the high-performance computing platform of Northwest A&F University for providing the computing resources.

## FUNDING

This study was funded by National Key R&D Program of China (2022YFF1001200) to Ran Li, ，the National Key R&D Program of China (2022YFF1000100) and the Shaanxi Laboratory Project for Arid Region Agriculture (2024ZY-JCYJ-02-15) to Yu Jiang, the Postdoctoral Fellowship Program of CPSF (GZC20252336) and Shaanxi Postdoctoral Research Foundation (2024BSHSDZZ184) to Zhuangbiao Zhang. The funding bodies played no role in the design of the study, collection, analysis, and interpretation of data and writing the manuscript.

## REFERENCE

Aguilar, I., Legarra, A., Cardoso, F., et al. 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. *Genet Sel Evol, 51*(1), 28.

Bhuiyan, M. S. A., Lim, D., Park, M., et al. 2018. Functional Partitioning of Genomic Variance and Genome-Wide Association Study for Carcass Traits in Korean Hanwoo Cattle Using Imputed Sequence Level SNP Data. *Frontiers in Genetics, 9*.

Brzáková, M., Zavadilová, L., Přibyl, J., et al. 2019. Estimation of genetic parameters for female fertility traits in the Czech Holstein population. *Czech Journal of Animal Science, 64*(5), 199-206.

Cerri, R. L. A., Burnett, T. A., Madureira, A. M. L., et al. 2021. Symposium review: Linking activity-sensor data and physiology to improve dairy cow fertility. *Journal of Dairy Science, 104*(1), 1220-1231.

Chang, L. F., Zhang, Z., Yang, J., et al. 2014. Molecular architecture and mechanism of the anaphase-promoting complex. *Nature, 513*(7518), 388-393.

Chen, S., Zhou, Y., Chen, Y., et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), i884-i890.

Cole, J. B., Wiggans, G. R., Ma, L., et al. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics, 12*, 408.

Dobin, A., Davis, C. A., Schlesinger, F., et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21.

Eghbalsaied, S. 2011. Estimation of genetic parameters for 13 female fertility indices in Holstein dairy cows. *Tropical Animal Health and Production, 43*(4), 811-816.

Ferrari, V., Galluzzo, F., van Kaam, J. B. C. H., et al. 2024. Genetic and genomic evaluation of age at first calving in Italian Holsteins. *Journal of Dairy Science, 107*(5), 3104-3113.

Furuta, T., Tuck, S., Kirchner, J., et al. 2000. EMB-30: an APC4 homologue required for metaphase-to-anaphase transitions during meiosis and mitosis in Caenorhabditis elegans. *Mol Biol Cell, 11*(4), 1401-1419.
Gobikrushanth, M., Purfield, D. C., Colazo, M. G., et al. 2018. The relationship between serum anti-Müllerian hormone concentrations and fertility, and genome-wide associations for anti-Müllerian hormone in Holstein cows. *Journal of Dairy Science, 101*(8), 7563-7574.
Grosshans, T., Xu, Z. Z., Burton, L. J., et al. 1997. Performance and genetic parameters for fertility of seasonal dairy cows in New Zealand. *Livestock production science, 51*(1), 41-51.
Gutiérrez Gil, B., Wiener, P., Williams, J. L., et al. 2012. Investigation of the genetic architecture of a bone carcass weight QTL on BTA 6. *Animal Genetics, 43*(6), 654-661.
Heise, J., Stock, K. F., Reinhardt, F., et al. 2018. Phenotypic and genetic relationships between age at first calving, its component traits, and survival of heifers up to second calving. *J Dairy Sci, 101*(1), 425-432.
Hofler, A., Yu, J., Yang, J., et al. 2024. Cryo-EM structures of apo-APC/C and APC/C(CDH1:EMI1) complexes provide insights into APC/C regulation. *Nat Commun, 15*(1), 10074.
Höglund, J. K., Buitenhuis, B., Guldbrandtsen, B., et al. 2015. Genome-wide association study for female fertility in Nordic Red cattle. *BMC Genetics, 16*(1).
Hu, H. H., Li, F., Mu, T., et al. 2023. Genetic analysis of longevity and their associations with fertility traits in Holstein cattle. *animal, 17*(6), 100851.
Hutchison, J. L., VanRaden, P. M., Null, D. J., et al. 2017. Genomic evaluation of age at first calving. *Journal of Dairy Science, 100*(8), 6853-6861.
Jagusiak, W., Zarnecki, A. 2006. Genetic Evaluation for Fertility Traits in Polish Holsteins. *Proceedings of the 2006 Interbull meeting*(No. 35).
Jamrozik, J., Fatehi, J., Kistemaker, G. J., et al. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. *J Dairy Sci, 88*(6), 2199-2208.
Jansen, J., Van Der Werf, J., De Boer, W. 1987. Genetic Relationships between Fertility Traits for Dairy Cows in Different Parities. *Livestock Production Science*(17), 337-349.
Karszes, J., Hill, L. 2020. Dairy Replacement Program: Cost & Analysis Summer 2019: Department of Animal Science,College of Agricultural and Life Sciences, Cornell University.
Kgari, R. D., Muller, C., Dzama, K., et al. 2022. Estimation of Genetic Parameters for Heifer and Cow Fertility Traits Derived from On-Farm AI Service Records of South African Holstein Cattle. *Animals, 12*(16), 2023.
Kichaev, G., Bhatia, G., Loh, P., et al. 2019. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *The American Journal of Human Genetics, 104*(1), 65-75.
Li, B., Dewey, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323.
Li, H., Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589-595.
Li, H., Handsaker, B., Wysoker, A., et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.
Lu, D., Miller, S., Sargolzaei, M., et al. 2013. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. *Journal of animal science, 91*(8), 3612-3633.
Malchiodi F, Cecchinato A, Bittante G. 2014. Fertility traits of purebred Holsteins and 2- and 3-breed crossbred heifers and cows obtained from Swedish Red, Montbéliarde, and Brown Swiss sires. *Journal of Dairy Science*. 97(12):7916-7926.

Mancin, E., Gomez Proto, G., Tuliozi, B., et al. 2024. Uncovering genetic parameters and environmental influences on fertility, milk production, and quality in autochthonous Reggiana cattle. *Journal of Dairy Science, 107*(2), 956-977.
Mäntysaari, P., Ojala, M., Mäntysaari, E. A. 2002. Measures of before and after breeding daily gains of dairy replacement heifers and their relationship with first lactation milk production traits. *Livestock production science, 75*(3), 313-322.
Mills, M. C., Tropf, F. C., Brazel, D. M., et al. 2021. Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nature Human Behaviour, 5*(12), 1717-1730.
Misztal, I., Tsuruta, S., Lourenco, D., et al. 2014. BLUPF90 family of programs (pp. 1-161).
Müller, M. P., Rothammer, S., Seichter, D., et al. 2017. Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *Journal of Dairy Science, 100*(3), 1987-2006.
Norman, H. D., Wright, J. R., Kuhn, M. T., et al. 2009. Genetic and environmental factors that affect gestation length in dairy cattle. *Journal of Dairy Science, 92*(5), 2259-2269.
Notter, D. R. 1999. The importance of genetic diversity in livestock populations of the future. *J Anim Sci, 77*(1), 61-69.
Parker Gaddis, K. L., Null, D. J., Cole, J. B. 2016. Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. *Journal of Dairy Science, 99*(8), 6420-6435.
Prakapenka, D., Liang, Z., Da, Y. 2023. Genome-Wide Association Study of Age at First Calving in U.S. Holstein Cows. *International Journal of Molecular Sciences, 24*(8), 7109.
Raheja, K. L., Burnside, E. B., Schaeffer, L. R. 1989. Heifer fertility and its relationship with cow fertility and production traits in Holstein dairy cattle. *J Dairy Sci, 72*(10), 2665-2669.
Ritchie, M. E., Phipson, B., Wu, D., et al. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47.
Schoeler, T., Speed, D., Porcu, E., et al. 2023. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nature Human Behaviour, 7*(7), 1216-1227.
Slagboom, M., Sørensen, A. C., Thomasen, J. R., et al. 2021. Ignoring genotype by environment interaction in the genetic evaluation of dairy cattle reduces accuracy but may increase selection intensity. *Journal of Dairy Science, 104*(12), 12756-12764.
Snelling, W. M., Allan, M. F., Keele, J. W., et al. 2010. Genome-wide association study of growth in crossbred beef cattle. *Journal of Animal Science, 88*(3), 837-848.
Stephen, M. A., Burke, C. R., Steele, N., et al. 2023. Genome-Wide Association Study of age at puberty and its (co)variances with fertility and stature in growing and lactating Holstein-Friesian dairy cattle. *Journal of Dairy Science*.
Vinothraj, S., Subramaniyan, A., Venkataramanan, R., et al. 2016. Genetic evaluation of reproduction performance of Jersey x Red Sindhi crossbred cows. *Vet World, 9*(9), 1012-1017.
Wang, K., Li, M., Hakonarson, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res, 38*(16), e164.
Weller, J. I., Ezra, E., Gershoni, M. 2022. Genetic and genomic analysis of age at first insemination in Israeli dairy cattle. *Journal of Dairy Science, 105*(6), 5192-5205.
Wu, T., Hu, E., Xu, S., et al. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb), 2*(3), 100141.
Yadav, N., Mukherjee, S., Mukherjee, A. 2023. Comparative genetic analysis of frequentist and Bayesian approach for reproduction, production and life time traits showing favourable association of age at first calving in Tharparkar cattle. *Animal Bioscience, 36*(12), 1806-1820.

Yang, H., Wang, K. 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc, 10*(10), 1556-1566.
Yin, T., König, S. 2018. Genetic parameters for body weight from birth to calving and associations between weights with test-day, health, and female fertility traits. *Journal of Dairy Science, 101*(3), 2158-2170.
Yu, G., Wang, L. G., Han, Y., et al. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS, 16*(5), 284-287.
Zhang, Z., Wang, A., Hu, H., et al. 2023. The efficient phasing and imputation pipeline of low‐coverage whole genome sequencing data using a high‐quality and publicly available reference panel in cattle. *Animal Research and One Health, 1*(1), 4-16.
Zhu, K., Wang, H., Chen, Z., et al. 2024. Estimation of genetic parameters for fertility traits in Chinese Holstein of South China. Front. Genet. 14: 1288375.
Zhu, Z., Guo, Y., Shi, H., et al. 2020. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *Journal of Allergy and Clinical Immunology, 145*(2), 537-549.

**Table 1. Descriptive statistics of fertility trait in Chinese Holstein heifer**

| Trait | Number | Mean | SD | CV(%) | Min | Max | Median |
|---|---|---|---|---|---|---|---|
| AFS | 8237 | 421.91 | 21.84 | 5.18 | 367 | 560 | 422 |
| AFC | 8167 | 715.53 | 45.44 | 6.36 | 607 | 1103 | 708 |
| IFLh | 8061 | 20.6 | 38.34 | 186.14 | 0 | 435 | 0 |
| NSh | 8035 | 1.69 | 1.03 | 60.95 | 1 | 8 | 1 |
| CRh | 8060 | 0.58 | 0.49 | 85.07 | 0 | 1 | 1 |
| GLh | 8118 | 273.02 | 10.25 | 3.75 | 206 | 343 | 275 |
| BiWh | 7859 | 36.06 | 6.53 | 18.12 | 12 | 75 | 37 |

AFS, age at first service (days); AFC, age at first calving (days); IFLh, first to last insemination in heifers (days); NSh, number of services per conception in heifers (count); CRh, conception rate in heifers (0/1); GLh, gestation length in heifers (days); BiWh, birth weight of heifers (kg). SD: standard deviation; CV: coefficient of variation; Min: minimum value; Max: maximum value.

.

**Table 2. The variance component and heritability of fertility traits in Chinese Holstein**

630 **heifer**

| Trait | $V_a$ | $V_e$ | SE(Var) | $h^2$ | SE($h^2$) | $P(h^2)$ |
|---|---|---|---|---|---|---|
| AFS | 71.474 | 187.762 | 7.349 | 0.276 | 0.025 | 1.89E-28 |
| AFC | 97.449 | 1604.560 | 28.870 | 0.057 | 0.017 | 6.27E-04 |
| IFLh | 1.372 | 1383.990 | 15.084 | 0.001 | 0.011 | 9.28E-02 |
| NSh | 0.035 | 0.960 | 0.016 | 0.035 | 0.016 | 2.75E-02 |
| CRh | 0.007 | 0.227 | 0.004 | 0.029 | 0.016 | 6.63E-02 |
| GLh | 4.711 | 91.306 | 1.554 | 0.049 | 0.016 | 2.20E-03 |
| BiWh | 2.736 | 18.185 | 0.516 | 0.131 | 0.024 | 3.24E-08 |

631 AFS, age at first service (days); AFC, age at first calving (days); IFLh, first to last insemination in
632 heifers (days); NSh, number of services per conception in heifers(count); CRh, conception rate in
633 heifers (0/1); GLh, gestation length in heifers (days); BiWh, birth weight of heifers (kg). $V_a$:
634 variance of additive genetic effect; $V_e$: variance of residual effect; Var: variance; SE: standard
635 error; $h^2$: heritability.

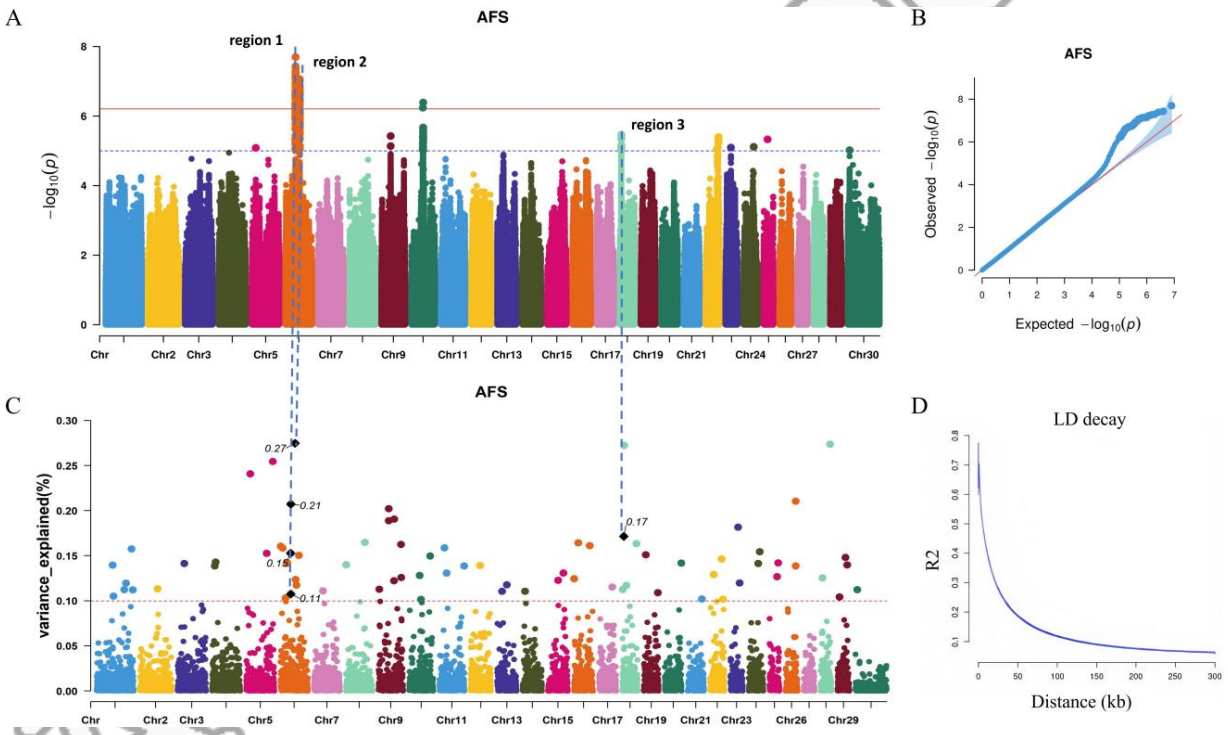
636

637 **Table 3. The results of SusiE fine mapping and eQTL co-localization**

| **Chr:pos** | ***P*_GWAS** | **PIP_GWAS** | ***P*_eQTL** | **PIP_eQTL** | **$PP_4$_coloc** |
|---|---|---|---|---|---|
| 6:44932169 | 7.13E-08 | 0.09 | 1.15E-07 | 0.16 | 0.30 |
| 6:44905170 | 7.12E-08 | 0.09 | 2.29E-07 | 0.08 | 0.17 |
| 6:44919942 | 4.10E-08 | 0.15 | 5.17E-07 | 0.04 | 0.14 |
| 6:44910551 | 1.07E-07 | 0.06 | 2.34E-07 | 0.08 | 0.11 |
| 6:44924418 | 5.69E-08 | 0.11 | 6.57E-07 | 0.03 | 0.08 |
| 6:44930852 | 7.79E-08 | 0.08 | 5.68E-07 | 0.04 | 0.07 |
| 6:44923172 | 1.17E-07 | 0.06 | 9.13E-07 | 0.02 | 0.03 |
| 6:44936395 | 2.50E-07 | 0.04 | 4.26E-07 | 0.05 | 0.03 |
| 6:44927054 | 2.16E-07 | 0.04 | 6.00E-07 | 0.04 | 0.03 |
| 6:44926254 | 1.18E-07 | 0.06 | 1.49E-06 | 0.02 | 0.02 |
| 6:44931905 | 5.57E-07 | 0.03 | 7.70E-07 | 0.03 | 0.01 |
| 6:44918028 | 5.84E-07 | 0.03 | 8.54E-07 | 0.03 | 0.01 |
| 6:44920640 | 2.31E-06 | 0.02 | 8.05E-07 | 0.03 | <0.01 |
| 6:44909660 | 1.23E-04 | <0.01 | 1.34E-07 | 0.14 | <0.01 |
| 6:44936718 | 3.86E-05 | <0.01 | 8.25E-07 | 0.03 | <0.01 |

| Chr:pos | *P*_GWAS | PIP_GWAS | *P*_eQTL | PIP_eQTL | $PP_4$_coloc |
|---|---|---|---|---|---|
| 6:44927644 | 4.79E-05 | <0.01 | 1.04E-06 | 0.02 | <0.01 |
| 6:44922371 | 6.62E-05 | <0.01 | 1.67E-06 | 0.01 | <0.01 |
| 6:44918123 | 7.96E-05 | <0.01 | 1.53E-06 | 0.02 | <0.01 |
| 6:44914879 | 1.38E-04 | <0.01 | 8.50E-07 | 0.03 | <0.01 |
| 6:44912168 | 2.23E-04 | <0.01 | 1.33E-06 | 0.02 | <0.01 |
| sum | - | 0.86 | - | 0.92 | 0.99 |

638 PIP: posterior inclusion probability; $PP_4$: the posterior probability that the GWAS and eQTL
639 signals share the same causal variant.




640

641 **Figure 1. Genome-wide signals for AFS and regional variance partitioning**

642 A: Manhattan plot from LMM-based GWAS. The solid red line marks the LD-adjusted genome-
643 wide threshold; the blue dashed line marks the suggestive threshold. Three prioritized regions are
644 indicated. B: Quantile-quantile plot showing good calibration (λ=1.01). C: Proportion of additive
645 genetic variance explained by consecutive, non-overlapping 200-kb windows across the genome.
646 The black dashed line denotes 0.1% variance explained; vertical guides highlight the three candidate

regions. D: LD decay curve ($r^2$ versus physical distance); the 200-kb window size was chosen based on this decay profile.

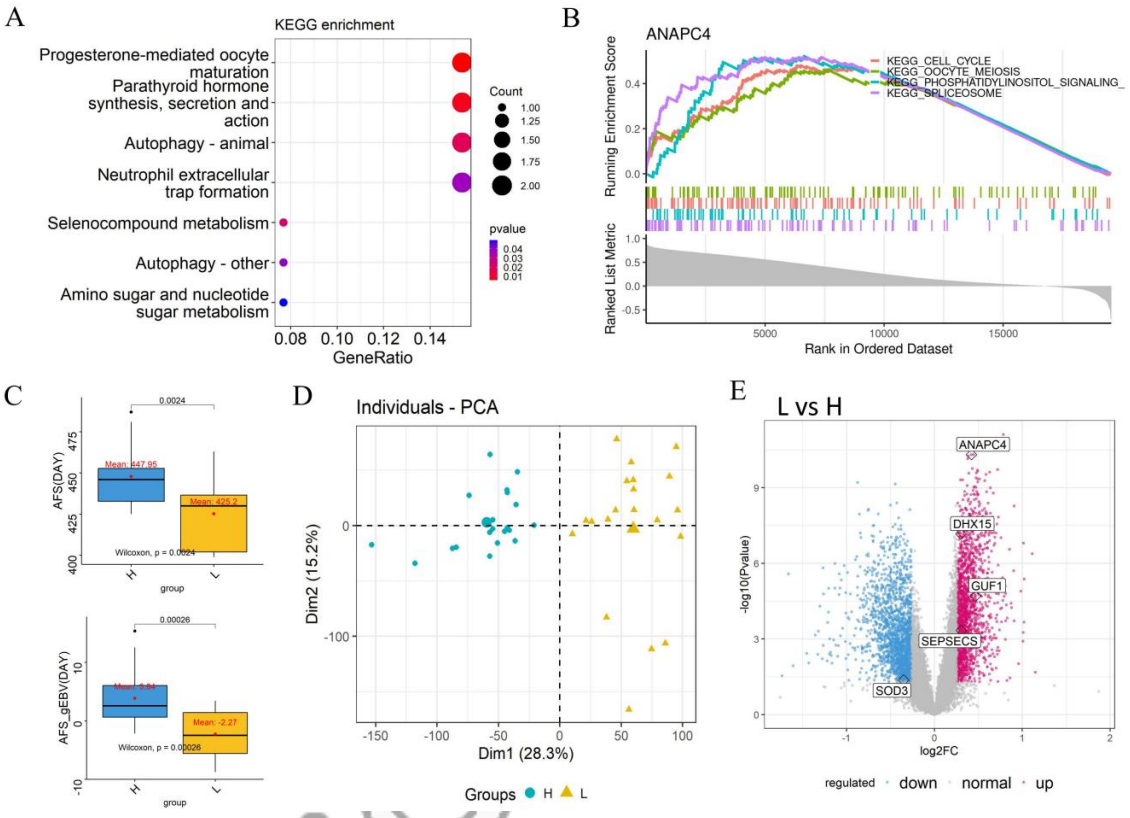
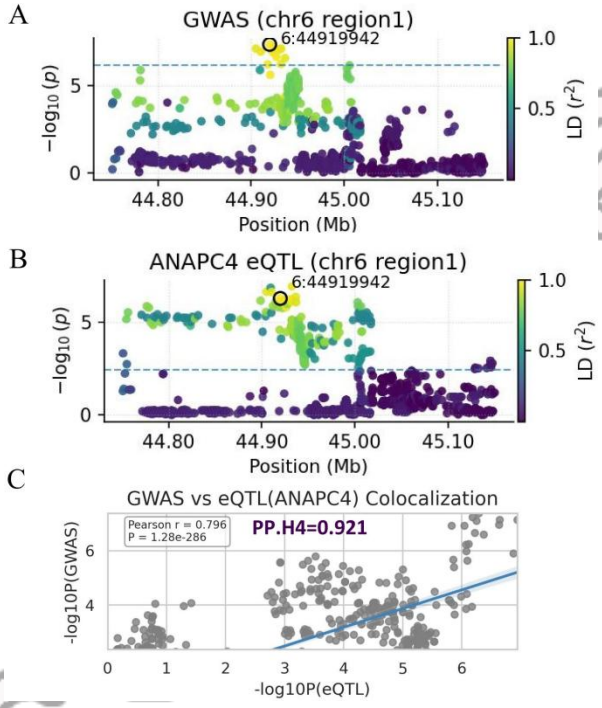



**Figure 2. Functional enrichment and gene differential expression analysis**

A: KEGG pathway enrichment of the candidate genes. The size of each bubble represents the number of genes enriched in the pathway, and the color intensity indicates the statistical significance. The X-axis represents the enrichment score, while the Y-axis shows the KEGG pathway terms. B: Single-gene GSEA anchored on *ANAPC4*: genes ranked by correlation with *ANAPC4* expression; running enrichment score curves shown for top pathways (e.g., cell cycle, oocyte meiosis). Vertical ticks mark pathway members along the ranked list. C: Boxplots comparing AFS phenotype and AFS GEBV between the high (H) and low (L) groups; *P* values from Wilcoxon tests are indicated. D: PCA of RNA-seq samples based on normalized gene

660 expression, colored by H/L groups. E: Volcano plot of L vs H differential expression (FDR < 0.05,
661 |FC| > 1.2). The five marked genes *ANAPC4, GUF1, DHX15, SOD3* and *SEPSECS* are all located
662 in Region 1 of BTA6. Light blue and pink denote significantly down- and up-regulated genes,
663 respectively; selected genes from GWAS-prioritized regions were labeled.
664




665

666 **Figure 3. Colocalization of the GWAS signal with the *ANAPC4* cis-eQTL on BTA6**

667 A: GWAS association plot for AFS in region 1 of BTA6. B: cis-eQTL plot for *ANAPC4* expression
668 in blood (n=648) across the same interval. In (A–B), points are colored by LD ($r^2$) with the lead
669 variant (BTA6:44,919,942, circled); the horizontal dashed line marks the suggestive threshold. C:
670 GWAS-eQTL colocalization. Each dot is a variant present in both analyses; the blue line shows the
671 linear fit.

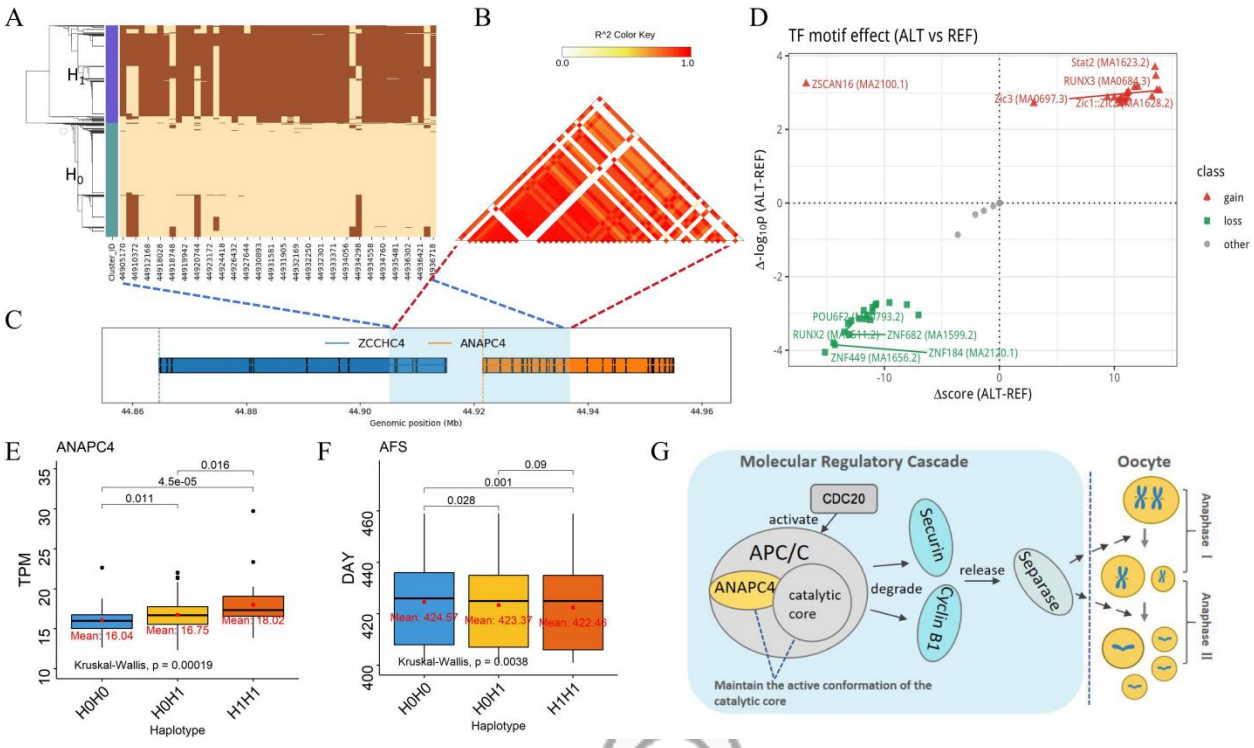



**Figure 4. A fine-mapped haplotype at *ZCCHC4–ANAPC4* that upregulates *ANAPC4* and associates with AFS**

A: Haplotype heat map in the BTA6 peak, showing two high-LD haplotypes (H0, H1). B: LD block structure of the same interval (pairwise $r^2$; warmer colors represent stronger LD). C: Genomic context of the fine-mapped region on BTA6 spanning *ZCCHC4* and *ANAPC4*.. D: Transcription-factor motif volcano plot of allele-specific effects. Each point is one TF motif comparing ALT vs REF sequences. Upper-right quadrant indicates stronger and more significant matches in ALT (motif gain); lower-left indicates loss in ALT relative to REF. Red triangles represent gain; green squares represent loss; grey circles represent non-significant/other. E: Boxplots of *ANAPC4* expression (TPM) by diplotype. Group sizes: n(H0H0)=170, n(H0H1)=359, n(H1H1)=119.. F: Boxplots of AFS (days) by diplotype. Group sizes: n(H0H0)=2331, n(H0H1)=4108, n(H1H1)=1597. G: Proposed mechanistic model of ANAPC4.

## SUPPLEMENTARY DATA

**Supplementary Figures**:



**Supplementary Tables**:

# GWAS 与精细定位鉴定 *ANAPC4* 为影响中国荷斯坦牛青春期的候选调控基因

潘翠丽[1#]，马龙刚[1#]，曹雪莎[1]，王琪艳[1]，王璐璐[1]，王炯[1]，汪傲[1]，李冉[1]，王喜宏[1]，马云[2]，刘光磊[3]，朱凯[3]，吕昕哲[3]，张壮彪[1*]，姜雨[1*]

[1] 陕西省动物遗传育种与繁殖重点实验室，西北农林科技大学动物科技学院，陕西杨凌 712100，中国

[2] 宁夏大学动物科技学院，宁夏反刍动物分子与细胞育种重点实验室，宁夏银川 750021，中国

[3] 河北品元生物科技有限公司，河北石家庄 050600，中国

[#]标注作者对本研究贡献相同

*通讯作者：zhangzhuangbiao18@163.com；yu.jiang@nwafu.edu.cn

## 摘要

奶牛繁殖效率对牧场可持续经营具有重要影响，青年母牛在开始产奶之前需要经历较长时间的饲养阶段。初配日龄（Age at First Service，AFS）是影响奶牛繁殖效率和育成成本的关键指标之一。本研究以中国荷斯坦牛为研究对象，旨在解析 AFS 的遗传结构并鉴定潜在的功能候选基因。通过低深度全基因组重测序结合基因型填充，共获得 7,964,217 个高质量 SNP，用于全基因组关联分析（GWAS）及基于窗口的遗传方差分析。在评估的七个繁殖性状中，AFS 的遗传力最高（$h^2 = 0.276$）。三个候选基因组区域共同解释了 0.91%的遗传方差，位于 BTA6 的显著关联信号覆盖 *ZCCHC4* 和 *ANAPC4* 区域，*ANAPC4* 参与雌激素诱导的卵母细胞成熟过程。GWAS 信号与 *ANAPC4* 血液顺式表达数量性状位点（cis-eQTL）共定位的后验概率为 0.921，提示其共享因果变异。进一步的 SuSiE 精细

定位将关联信号锁定在约 30 kb 的基因组区域内，该区域的 20 个变异组成两个高度连锁不平衡（LD）的单倍型（H0 和 H1）。其中，H1H1 单倍型 *ANAPC4* 表达显著升高，AFS 平均提前约 2.1 天。转录因子结合预测显示不同单倍型之间存在等位基因特异性的转录因子结合差异。综上，本研究鉴定并精细定位了一个上调 *ANAPC4* 并与 AFS 提前相关的调控单倍型，为利用基因组选择优化青年牛青春期时间、提高奶牛生产效率和牧场经济效益提供候选分子标记。

关键词：青春期时间；GWAS；精细定位；中国荷斯坦青年牛；*ANAPC4*