Journal of Integrative Agriculture 2026, 25(3): 1099–1113

Available online at ScienceDirect

Journal of Integrative Agriculture



RESEARCH ARTICLE

Unraveling genetic underpinnings of purine content in pork

Cong Huang, Min Zheng, Yizhong Huang, Liping Cai, Xiaoxiao Zou, Tianxiong Yao, Xinke Xie, Bin Yang, Shijun Xiao, Junwu Ma[#], Lusheng Huang[#]

National Key Laboratory for Pig Genetic Improvement and Germplasm Innovation, Jiangxi Agricultural University, Nanchang 330045, China

Highlights

- First study to systematically elucidate the molecular mechanism underlying purine content modulation in pork.
- Two highly significant quantitative trait loci (QTLs) on *Sus scrofa* chromosome 12 (SSC12) were identified, independently influencing guanine content and adenine/hypoxanthine content in pork.
- Provides a robust foundation for selecting pig breeds with reduced purine base content.

Abstract

The significance of purine base content as an important nutrition indicator in foods arises from its potential to trigger hyperuricemia or gout *via* high-purine diet. Livestock meats, including pork, generally contain moderate to high total purine content (TP). Recent research revealed substantial variations within and across pig breeds, implying genetic factors influencing this trait. Thus, this study aimed to unravel the genetic underpinnings governing purine base content in pork. The heritability estimates (h^2) for the four purine traits ranged from 0.14 to 0.35. A total of 14, 36, 19 and 25 quantitative trait loci (QTLs) were identified for guanine, adenine, hypoxanthine, and TP, respectively. Our comprehensive gene set enrichment analysis and gene network analysis revealed 15 promising candidate genes intricately interwoven within diverse purine metabolism pathways, such as purine ribonucleotide metabolic process, purine nucleotide metabolism and transport, and purine salvage pathways, all contributing to TP. Strikingly, most genetic variants significantly associated with TP displayed analogous effects on multiple purine bases. Two distinct and highly significant QTLs ($P<10^{-12}$) emerged on *Sus scrofa* chromosome (SSC) 12: one impacting guanine content and the other concurrently influencing adenine and hypoxanthine levels. The peak of the guanine QTL on SSC12 resided 1.1 kb downstream of the transmembrane protein 238 like (*TMEM238L*) gene and is encapsulated within a genomic segment characterized by the histone modification H3K27me3. Focused fine-mapping for the SSC12 QTL associated with adenine and hypoxanthine levels narrowed its scope to around 172 kb, encompassing the growth arrest specific 7 (*GAS7*) and myosin heavy chain 13 (*MYH13*) genes. However, the observed QTL effect was not attributed to any missense mutations within the two genes. This pioneering study unveils the genetic variations and candidate genes associated with purine content in livestock, laying a robust foundation for the selective breeding of pig lines with reduced purine base content.

Keywords: pig, purine content, genetic architecture, GWAS, QTL

1. Introduction

Gout is the most common inflammatory arthritis, affecting 41 million adults and increasing in prevalence worldwide, and this disease is caused by excess serum uric acid (i.e. hyperuricemia) (Danve *et al.* 2021). Uric acid is the end product of purine metabolism in the human body and is mainly excreted through the kidneys (80%) and guts (20%) to keep homeostasis (So and Thorens 2010). However, this homeostasis can easily be disrupted by excessive consumption of purine-rich foods such as meat and seafood (Choi *et al.* 2004; Wu *et al.* 2019), because the exogenous purine compounds, such as adenosine, can act on related receptors and then be metabolized or internalized into the cell through nucleoside uptake systems (Blanco *et al.*

Received 30 October 2023; Received in revised form 25 January 2024; Accepted 11 March 2024; Available online 13 May 2024
Cong Huang, E-mail: 1714182072@qq.com; [#]Correspondence Junwu Ma, E-mail: mjwjxlab@jxau.edu.cn; Lusheng Huang, E-mail: lushenghuang@hotmail.com

© 2026 CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer review under responsibility of Editorial Board of *Journal of Integrative Agriculture*.
doi: 10.1016/j.jia.2024.05.004

1993). Once the over-produced uric acid cannot be excreted in time, too high a serum uric acid level would result in many diseases, including gout, abdominal obesity, insulin resistance, cardiovascular and kidney disease (Bhole *et al.* 2010; Smith *et al.* 2010; Kanbay *et al.* 2013; Barkas *et al.* 2018). Globally, the escalating incidence of gout is closely correlated with the rising demand for meat products, thereby intensifying the burden on healthcare systems (Smith *et al.* 2011; Soare and Chiurciu 2017; Dehlin *et al.* 2020). In clinical practice, individuals diagnosed with hyperuricemia or gout are often advised to follow a low-purine diet or reduce their meat consumption. However, adhering to these recommendations may compromise their access to vital nutrients present in meat. Therefore, reducing the purine content in meat can enhance its nutritional quality and benefit the health of people susceptible to gout.

While some researchers have explored processing and cooking techniques to diminish purine levels in food, few have investigated the potential of genetic breeding for this purpose. Recently, our study demonstrated substantial variations in total purine content (TP) within and among Chinese native and Western commercial pig breeds (Zheng *et al.* 2018), suggesting the feasibility of cultivating new pig lines with lower TP in muscle. However, the genetic basis of purine content in animal skeletal muscle remains poorly understood, despite the critical role of purine bases (adenine, guanine, hypoxanthine, and xanthine) in vital cellular processes such as RNA and DNA synthesis, regulation of enzyme activity, protein synthesis and function, energy metabolism and transfer, essential coenzyme function, and cell signaling (Duval *et al.* 2013). Genome-wide association studies (GWAS) have successfully identified numerous quantitative trait loci (QTLs) associated with various nutrients or compounds in pork, encompassing protein, fat, lipid, fatty acids, amino acids, carbohydrates, and water (Li *et al.* 2010; Ma *et al.* 2014; Liu *et al.* 2015; Davoli *et al.* 2016; Duan *et al.* 2016; Viterbo *et al.* 2018; Guo *et al.* 2020). In contrast, QTLs for TP and purine base contents in meat have yet to be reported.

To fill this knowledge gap, we established a unique swine heterogeneous stock (HS) by selectively crossing eight breeds, which include four Western commercial pig breeds and four Chinese indigenous pig breeds. This HS population provides a wide range of genetic variations. By leveraging this valuable resource, our study seeks to identify the genetic loci and candidate genes associated with TP and purine base contents in pork muscle. To achieve this, we conducted a comprehensive investigation on the F_6 and F_7 populations of the HS using single-population whole-genome sequence-based GWAS and meta-analysis methods.

2. Materials and methods

2.1. Animals

All animals used in this study were sourced from the sixth and seventh filial generations (F_6 and F_7) of a swine heterogeneous stock (HS) that was developed by crossing eight different pig breeds. The founder breeds of the HS include 4 Western commercial pig breeds (Landrace, Large White, Pietrain, and White Duroc) and 4 Chinese local breeds (Bamaxiang, Erhualian, Laiwu, and Tibetan). Detailed information on the establishment of the HS have been reported previously (Ji *et al.* 2018; Huang *et al.* 2020). The F_6 and F_7 populations were reared under similar management and environmental conditions at the experimental farm of Jiangxi Agricultural University, China. A total of 836 F_6 pigs (448 females and 388 males) and 668 F_7 pigs (288 sows and 380 barrows) were slaughtered at 240±5 days of age in an abattoir of Nanchang Baodi Agricultural Technology Co., Ltd., China.

2.2. Measurement of purine base content in muscle

To determine contents of purine bases, approximately 1 g sample of *Longissimus lumborum* (LL) muscle was taken at 1st–2nd lumbar vertebra on the left side of each carcass within postmortem 45 min and then stored in a refrigerator at –80°C until required. Purine bases were extracted from LL muscle and quantified according to the protocol described by Rong *et al.* (2015). Except for the xanthine content which was too low to be detected, the contents of the other three purine bases (guanine, hypoxanthine, and adenine, units: mg 0.1 kg^{-1}) were determined in the LL muscle samples from 420 F_6 individuals and 665 F_7 individuals. Additionally, TP content was calculated by summing the contents of guanine, hypoxanthine, and adenine in each sample.

2.3. Whole-genome sequencing and genotyping

Genomic DNA of each animal was extracted from the ear or tail tissue using a standard phenol/chloroform protocol and was then dissolved in Tris-EDTA buffer to a final concentration of 100 ng μL^{-1} after being adjusted by a Nanodrop-1000 spectrophotometer (ThermoFisher, USA) with ultraviolet absorption ratios of ~1.8 for 260/280 and 2.0–2.2 for 260/230. The DNA samples were then genotyped using next-generation sequencing (NGS) technology on the Illumina Xten platform (Illumina Inc., San Diego, CA) at Novogene company (Beijing, China). The clean data generated from the raw data was processed by several bioinformatic tools, like BWA (Li and Durbin 2010), SAMTools (Li *et al.* 2009), Picard, Platypus (Rimmer *et al.* 2014), and Beagle v.4 programs (Browning and Browning 2007). Then sixteen randomly selected samples were sequenced twice to compare the genotype concordance between these two datasets for both F_6 and F_7 pig populations. The results showed that the mean ratio of genotype concordance was above 98.3% for SNPs and 97% for Indels in both populations. Data quality control was performed using PLINK software (Chang *et al.* 2015) with the parameters of SNP call rate >90% and minor allele frequency (MAF) >0.03. Consequently, 29,441,528 genetic markers (22,081,146 SNPs and 7,360,382 Indels) in the F_6 pigs and 29,198,737 markers (21,899,053 SNPs and 7,299,684 Indels) in the F_7 pigs were kept for subsequent association analysis.

2.4. Association analysis

GWASs were implemented by using GEMMA (Zhou and Stephens 2012) (version 0.98) software in which a genome-

wide efficient mixed-model association algorithm takes into consideration of sample structure and population stratification. The association test was then performed with phenotype, genotype, and the centered relatedness matrix files and fit into a univariate linear mixed model. The statistical model is shown below:

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

where **y** is an n×1 vector of phenotypic values for *n* individuals; **W** is an n×c incidence matrix of fixed effects of sex and slaughter batch, with c corresponding to the number of levels of fixed effects; α is a c×1 vector containing the value of the intercept and fixed effects of sex and slaughter batch; **x** is an n×1 vector of genotypes of a marker. Three genotypes AA, AB, and BB of a locus were coded as 0, 1, and 2, respectively, where B is the minor allele. β is the effect size of the marker; **u** is an n×1 vector of random polygenic effects with covariance structure assumed to follow N (0, $\mathbf{KV}_g$), where **K** represents an n×n genomic relationship matrix calculated from genetic markers and $\mathbf{V}_g$ is the polygenic additive variance; and **ε** is an n×1 vector of random residuals, **ε~N** (0, $\mathbf{IV}_e$), where I is an n×n identity matrix and $\mathbf{V}_e$ is the residual variance component. The genome-wide significance threshold based on Bonferroni correction is defined as 0.05/N, where N is the number of markers used in genome-wide association analysis, and the chromosome-wide (suggestive) significance threshold is defined as 1/N. However, Bonferroni method is highly conservation and would "overcorrect" for SNPs that are not truly independent. In the present study, with reference to GWASs executed in humans (Randall *et al.* 2010), we chose $P=5\times10^{-8}$ and $P=1\times10^{-6}$ as the genome-wide significance and chromosome-wide significance, respectively.

Meta-analysis of GWAS was performed using a z-score method implemented in METAL software (Willer *et al.* 2010), which tests the combined effects and standard errors of effects of each common SNP across the tested populations by taking sample size and direction of genotype effects into account. The Haploview software (Barrett *et al.* 2005) was used to visualize the linkage disequilibrium (LD) blocks. The phenotypic variance explained by a significant GWAS SNP was estimated using the following equation (Falconer 1996): $Var_{SNP}\%=[2p(1-p)\alpha^2/\sigma^2]\times100$, where *p* is the minor allele frequency of the SNP, α is the additive allelic effect of a given SNP, and σ^2 is the phenotypic variance for a trait estimated by the R program (v.3.6.2).

Population stratification was considered one of the major impact factors causing spurious association (Freedman *et al.* 2004), and the genomic inflation factor (λ) was used to assess the negative effects of population stratification (Devlin 1999; Freedman *et al.* 2004; Marchini *et al.* 2004) in our study by constructed quantile-quantile (Q-Q) plot (Pearson and Manolio 2008) using R software. GCTA software was used to estimate the heritability of studied traits (Yang *et al.* 2011).

2.5. Functional enrichment analysis for four purine traits

The candidate genes located within a ±500 kb vicinity of the top SNPs that exceeded the chromosome-wide significance level were selected as input genes for functional annotation analysis, including GO processes and KEGG pathways. The Metascape tool (http://metascape.org) (Zhou *et al.* 2019) was utilized for this analysis. Terms with $P<0.05$, a minimum count 3, and an enrichment factor>1.5 (the enrichment factor defined as the ratio between observed count and the count expected by chance) were collected and grouped into clusters based on their membership similarities. *P*-values were calculated based on accumulative hypergeometric distribution, while *q* values were generated using the Benjamini-Hochberg procedure to correct for multiple testing. To explore the functional relationships among the candidate genes implicated in GO terms linked to purine metabolism, a functional gene network was constructed using HumanNet v.1 (http://www.functionalnet.org/humannet) (Lee *et al.* 2011), a comprehensive genome-scale human gene network comprising 16,243 validated protein-encoding genes with 476,399 functional linkages. The resulting network was visualized by using Cytoscape software (https://cytoscape.org/).

2.6. Fine mapping of highly significant QTLs for purine traits.

Fine-mapping was conducted for the most significant QTL of each studied trait. The process involved the implementation of 3 LOD drop-off method to confirm the confidence interval (CI) based on the *P*-value of top variant of QTLs. Then the variants showing strong LD with the top loci, with a specific focus on missense mutations (already deposited in the dbSNP) and indels in exons, were selected. The nearest genes to these selected variants, as determined by the latest annotation of pig genome, were then explored. Lastly, the possible chromatin states across the region accommodating the CI of the top variant were investigated to screen out the most potentially candidate mutations and genes associated with the different purine base contents. This systematic approach aimed to identify and characterize key genetic elements contributing to the observed traits.

2.7. Correlation analyses between the expression levels of candidate genes and corresponding traits

The expression level of each gene in the *Longissimus* muscle samples from 279 F_6 and 292 F_7 animals were determined using RNA-Seq data. More details about how to proceed with the raw data could refer to Wu *et al.* (2022). In brief, total RNA was extracted from muscle samples using Trizol (Invitrogen, USA). RNA quantity and integrity were assessed using a NanoDrop-1000 spectrophotometer (ThermoFisher Scientific, USA) and concentrations were measured using Qubit® 3.0 fluorometer (Life *Technologies*™). RNA sequencing libraries for each sample (F_6 and F_7) were constructed with the NEBNext® UltraTMR NA Library Prep Kit for Illumina (New England Biolabs, Inc.) according to the manufacturer's manual. After finishing the sequencing processes, the raw RNA reads with the percentage of low quality (Q≤5) bases greater than 50% and N base contents greater than 10% were removed. The quality of RNA-seq data from all individuals was firstly evaluated using FastQC software and the clean reads were then mapped

to the reference genome using the STAR program (Dobin *et al*. 2013). Transcript assembly and quantification were performed using StringTie (Pertea *et al*. 2015) and featureCounts (Liao *et al*. 2014), respectively. The gene expression values were standardized as the fragments per kilobase of exon model per million mapped reads (FPKM) using DESeq2 (Love *et al*. 2014). Both the gene expression levels of candidate genes and phenotypic values (guanine, adenine, hypoxanthine, and TP levels) were adjusted for gender, slaughter batch, and kinship using a robust linear regression model (Lourenco *et al*. 2011). Associations between gene expression levels and the purine traits were evaluated by Pearson correlation coefficient using the *cor* function in R software.

3. Results

3.1. Summary statistics and heritability estimates of four purine traits

Table 1 summarizes the results of descriptive statistics and the heritability estimates (h^2) of four purine traits in the F_6 and F_7 populations. Hypoxanthine is the main component of TP (68%), followed by adenine (22%) and guanine (10%). All phenotype data were normally distributed, except for the adenine and hypoxanthine contents in F_7, which showed slightly right- and left-skewed histograms, respectively (Appendix A). The four purine traits differed significantly between the two populations. The coefficients of variation (CV) for guanine (9.20–9.77%) and TP (5.88–5.97%) were comparable in the F_6 and F_7 pigs, while the CV for adenine content (38.12% *vs*. 8.88%) and hypoxanthine content (16.48% *vs*. 7.22%) were higher in F_7 than in F_6. The heritability estimates (h^2) for guanine content (0.31–0.35), hypoxanthine content (0.22–0.24), and TP (0.20–0.21) were similar in the two populations, while the h^2 for adenine content was twice as high in F_7 than in F_6 (0.30 *vs*. 0.14). The results suggest that the contents of purine bases in muscle are moderately heritable traits.

3.2. Identification of GWAS loci for guanine

After quality control, 29,441,528 and 29,198,737 variants (SNPs and Indels) were retained for association analyses in the F_6 and F_7 pig populations, respectively. The GWAS analyses were not significantly biased by population stratification, as indicated by the Q-Q plots (Appendix B) and the λ values ranging from 1 to 1.05.

Table 1 Descriptive statistics and the heritability estimates (h^2) of four purine traits in F_6 and F_7 pig populations (unit: mg 0.1 kg^{-1})

Trait	F_6 (*n*=420)			F_7 (*n*=665)		
	Mean±SE[2)]	CV[3)]	h^2	Mean±SE	CV	h^2
Guanine	13.17±0.06 a	9.20	0.35	12.27±0.05 b	9.77	0.31
Adenine	25.10±0.11 a	8.88	0.14	33.26±0.49 b	38.12	0.30
Hypoxanthine	95.92±0.34 a	7.22	0.22	81.40±0.52 b	16.48	0.24
TP[1)]	134.19±0.3 a	5.88	0.20	126.92±0.29 b	5.97	0.21

[1)] The content of total purine (TP)=Guanine+Adenine+Hypoxanthine.
[2)] SE, standard error.
[3)] CV, coefficient of variation (%).
Different letters in a row indicate a significant difference between two populations for the same trait (P<0.01).

We identified 8 and 6 QTLs (P<1×10^{-6}) associated with guanine content in the F_6 and F_7 pig populations, respectively (Fig. 1-A and B; Appendix F), which collectively explained about 20% of the total phenotypic variance in each population. Intriguingly, the most significant QTLs in both populations were found to be overlapped in the genomic region from 55.1 to 55.6 Mb on *Sus scrofa* chromosome (SSC) 12, with peak SNPs 12_55179776 (P=8.11×10^{-12}) and 12_55500274 (P=1.11×10^{-19}) in F_6 and F_7, respectively (Appendix F). Furthermore, these SNPs were in moderately linked (r^2=0.5), suggesting that they may represent the same QTL. This was confirmed by the meta-analysis that revealed a much more significant SNP, 12_55495763 (P=1.34×10^{-28}), within the SSC12 QTL interval (Fig. 1-C; Table 2). Additionally, the meta-analysis identified 6 other loci associated with guanine content (Fig. 1-C), which were not detected by the single-population GWAS.

3.3. Identification of GWAS loci for adenine and hypoxanthine

In total, 36 QTLs (P<1×10^{-6}) associated with adenine content were identified, including 8, 17, and 11 QTLs identified by GWAS in the individual F_6 and F_7 population, as well as their meta-analysis, respectively (Fig. 2-A–C; Table 2; Appendix F). The most significant QTL consistently mapped to SSC12 in both populations. However, in the F_6 population, the peak SNP rs331176479 (an intronic variant of ribosomal protein L26 (*RPL26*)/NudE neurodevelopment protein 1 like 1 (*NDEL1*); P=7.96×10^{-9}) was found to be approximately 1.49 Mb away from the lead SNP rs343341698 (a downstream variant of *MYH13*; P=4.51×10^{-9}) identified in the F_7 population; and the LD value between rs331176479 and rs343341698 was very low (r^2=0.03), suggesting that they likely represent distinct QTLs. Interestingly, the meta-analysis also identified an adenine-associated signal (12_53668813; P=2.98×10^{-8}) on SSC12 (Fig. 2-C; Table 2). However, this signal was found to be less significant compared to the lead SNPs (rs331176479 or rs343341698) identified in the individual populations. Therefore, the findings suggest the presence of multiple loci for adenine content in the *RPL26-MYH13* region (about 1.5 Mb) on SSC12. Additionally, in the F_7 population, three additional genome-wide significant loci were discovered near the tumor protein P53 inducible protein 11 (*TP53I11*), *ENSSSCG00000045515*, and ATP binding cassette subfamily C member 5 (*ABCC5*) genes on SSC2, SSC7, and SSC13, respectively.

Regarding hypoxanthine content, the GWAS analysis revealed a single genome-wide significant QTL on SSC13 in F_6 and another one on SSC12 in F_7 (Fig. 2-D and E). The subsequent meta-analysis validated these two QTLs and further identified an additional genome-wide significant QTL on SSC7 (Fig. 2-F), which went undetected in the individual-population GWAS analysis.

Notably, in the F_7 population, the lead SNP associated with hypoxanthine content (rs328426812; P=8.56×10^{-13}) and the lead SNP for adenine content (rs343341698; P=4.51×10^{-9})

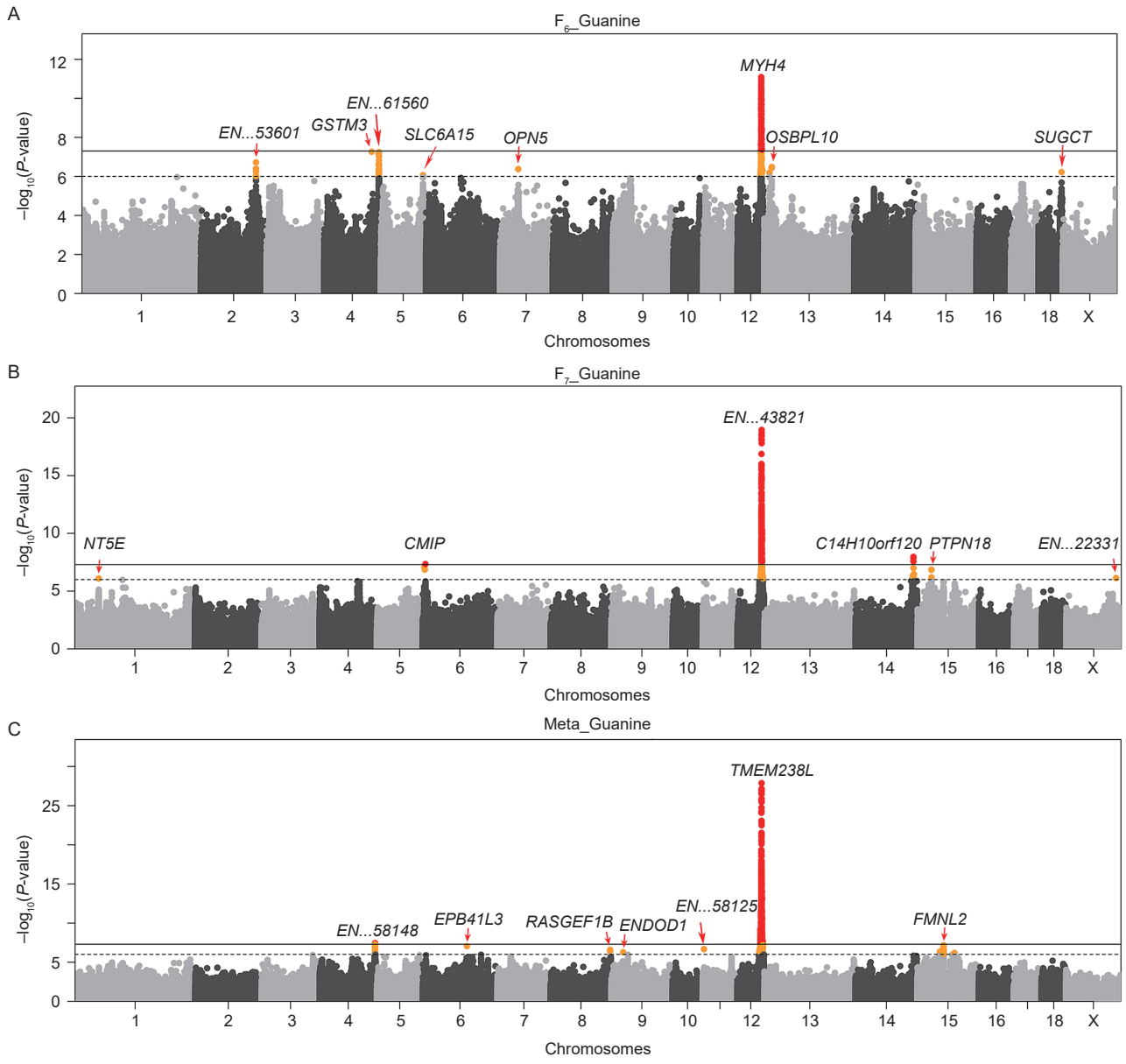


Fig. 1 Manhattan plots illustrating the results of association analyses for guanine content. A and B, GWAS in the F_6 (A) and F_7 population (B). C, meta-analysis. The *y*-axis shows the $-\log_{10}$(*P*-value), and the *x*-axis shows the physical positions of molecule markers on the pig chromosomes. The dashed horizontal line represents the threshold for chromosome-wide significance ($P=1\times10^{-6}$), and the solid line indicates the threshold for genome-wide significance ($P=5\times10^{-8}$). Genes labeled above correspond to the nearest annotated genes to the leading variants.

were both found in close proximity to the *MYH13* gene on SSC12 (Fig. 2-B and E), with a mere 348 bp physical distance and displaying nearly complete LD. This observation suggests the possibility of a shared causal mutation between these two traits. Similarly, an intergenic SNP (rs334270138) of *ABCC5* on SSC13 was significantly associated with both adenine and hypoxanthine contents in the F_7 population ($P<2.55\times10^{-8}$; Fig. 2-B and E; Appendix F).

3.4. Identification of GWAS loci for TP

For TP content, we identified 18 and 5 QTLs in the F_6 and F_7 populations, respectively (Fig. 3-A and B; Appendix F), which explain over 70% of the genetic variance of the trait in each corresponding population. In the F_6 population, three genome-wide significant association signals (rs325098830, 1_242759760, and 6_1309595) were detected for TP, with the closest genes being interaction protein for cytohesin exchange factors 1 (*IPCEF1*), phospholipid phosphatase related 1 (*PLPPR1*) and *U6*, respectively (Fig. 3-A). However, the GWAS loci for TP identified in F_7 were only significant at a suggestive level.

The meta-analysis revealed 6 QTLs for TP (Fig. 3-C; Table 2) with the most significant one being SNP 6_820343 ($P<1.11\times10^{-7}$), located within an intron of the CBFA2/RUNX1 partner transcriptional co-repressor 3 (*CBFA2T3*) gene. Notably, all variants in high LD with 6_820343 ($r^2>0.8$) were also located within the intron of the *CBFA2T3* gene, strongly suggesting that this gene is responsible for the variation in

Table 2 Summary of the significant variants (*P*-value<1×10^{-6}) for the content of four purine traits identified from GWAS meta-analyses[1)]

Trait	Chr[2)]	Lead variants	Pos (bp)[3)]	MAF[4)] (minor/major)	Beta±SE[5)]	Var (%)[6)]	*P*-value	Variant/Element annotation	Nearest gene
Guanine	4	4_126938308	126,938,308	0.186 (TG/T)	0.37±0.07	2.53	3.65e–08	Upstream	*ENSSSCG00000058148*
	6	rs318393618	101,487,177	0.15 (C/A)	–0.42±0.09	2.74	8.74e–08	Intron	*EPB41L3*
	8	rs332648348	136,310,008	0.055 (A/G)	–0.53±0.13	1.78	2.63e–07	Intron	*RASGEF1B*
	9	rs711905525	27,131,432	0.037 (G/C)	0.57±0.13	1.41	5.26e–07	Upstream	*ENDOD1*
	11	11_2908390	2,908,390	0.348 (C/G)	–0.37±0.06	3.78	2.10e–07	Intron	*ENSSSCG00000058125*
	12	12_55495763	55,495,763	0.094 (A/G)	0.98±0.09	9.97	1.34e–28	Downstream	*TMEM238L*
	15	rs344806628	60,147,762	0.04 (C/T)	0.82±0.17	3.15	6.79e–08	Intron	*FMNL2*
Adenine	6	rs709237668	46,839,242	0.436 (AT/A)	–1.88±0.50	1.50	1.19e–07	Downstream	*ZNF607*
	6	6_159649311	159,649,311	0.102 (T/TA)	3.31±0.66	1.73	4.20e–07	Intergenic	*TUT4*
	6	rs345815420	44,728,539	0.457 (G/T)	1.73±0.48	1.28	5.96e–07	Intergenic	*FAM187B*
	7	rs332074701	41,809,920	0.032 (T/C)	7.78±1.56	3.23	8.27e–07	Intron	*ADGRF1*
	7	rs324459067	5,823,577	0.127 (G/T)	–2.01±0.63	0.77	9.98e–07	Intergenic	*ENSSSCG00000050804*
	8	rs339730595	120,478,116	0.043 (A/G)	–4.68±0.98	1.55	5.86e–07	Intergenic	*ENSSSCG00000062660*
	8	rs331307378	121,191,576	0.064 (T/C)	–2.75±0.92	0.78	6.10e–07	Intergenic	*ENSSSCG00000051020*
	10	10_26005630	26,005,630	0.386 (A/AG)	2.81±0.62	3.23	9.55e–07	Intergenic	*ERCC6L2*
	11	rs340441627	75,798,400	0.294 (A/G)	2.29±0.46	1.88	2.14e–07	Intron	*MYO16*
	12	12_53668813	53,668,813	0.027 (AGG/A)	–2.24±1.24	5.29	2.98e–08	Intron	*MYH10*
	12	rs319943339	55,043,915	0.162 (C/T)	1.74±0.60	0.71	3.72e–07	Downstream	*MYH13*
	14	14_30942143	30,942,143	0.033 (TAC/T)	4.86±1.04	1.30	1.23e–07	Intron	*TMEM120B*
	14	14_57722435	57,722,435	0.066 (T/C)	2.55±0.82	0.69	4.53e–07	Intergenic	*ENSSSCG00000049255*
	X	23_112068193	112,068,193	0.311 (G/GT)	–1.16±0.43	0.50	4.64e–07	Intergenic	*GPR101*
	X	rs336184517	111,029,073	0.362 (T/A)	–1.46±0.38	0.85	5.34e–07	Intron	*INTS6L*
Hypoxanthine	1	rs327270655	231,223,669	0.045 (A/G)	6±1.24	1.73	2.96e–07	Intergenic	*PSAT1*
	1	1_220602045	220,602,045	0.116 (A/G)	3.06±0.83	1.07	3.32e–07	Intergenic	*ENSSSCG00000044946*
	2	2_14104095	14,104,095	0.031 (A/C)	4.54±1.38	0.69	7.17e–07	Intergenic	*ENSSSCG00000039933*
	2	2_128870712	128,870,712	0.189 (T/C)	–2.17±0.71	0.81	7.43e–07	Intron	*ENSSSCG00000060012*
	3	rs330990723	125,132,617	0.442 (T/C)	–1.96±0.58	1.06	8.09e–07	Intron	*GREB1*
	6	rs1109137956	66,337,797	0.121 (C/T)	–3.39±0.87	1.37	3.84e–07	Intergenic	*AJAP1*
	6	rs345400833	3,708,813	0.059 (G/A)	4.14±1.13	1.06	4.91e–07	Upstream	*CIBAR2*
	7	rs341290862	43,243,291	0.053 (T/C)	–4.58±1.24	1.18	1.91e–08	Intergenic	*ENSSSCG00000045515*
	12	rs334455944	55,060,526	0.077 (G/A)	–5.21±0.95	8.07	3.43e–10	Intron	*MYH13*
	13	13_18946714	18,946,714	0.417 (C/T)	–2.39±0.56	1.55	1.30e–08	Intron	*GLB1*
	13	rs318515348	16,809,406	0.122 (A/C)	3.68±0.83	1.62	1.32e–08	Intron	*TGFBR2*
	13	rs322963146	12,444,291	0.268 (C/T)	2.23±0.59	1.09	6.34e–08	Intron	*RARB*
	13	rs334270138	121,947,239	0.079 (A/C)	–3.56±1.08	1.03	5.76e–07	Intron	*ENSSSCG00000042409/ ENSSSCG00000062141*
	13	rs321970483	24,122,525	0.487 (G/A)	–1.88±0.58	0.99	7.02e–07	Intergenic	*MOBP*
	13	13_22547155	22,547,155	0.419 (A/G)	2.44±0.59	1.62	4.36e–07	Intron	*ITGA9*
	13	rs340295209	21,446,023	0.267 (C/G)	2.7±0.7	1.60	4.46e–07	Intron	*STAC*
	14	rs345952989	1,453,768	0.052 (A/G)	–4.93±0.97	1.34	7.83e–07	Intergenic	*ENSSSCG00000031962*
	16	rs1113489740	76,345,470	0.155 (A/T)	2.01±0.72	0.59	4.65e–07	Intron	*ADAMTS16*
	18	18_10892415	10,892,415	0.07 (AC/A)	–2.4±1.01	0.42	3.54e–07	Intergenic	*TMEM213*
TP	1	1_12429158	12,429,158	0.03 (T/C)	–2.48±1.10	0.5	3.04e–07	Intergenic	*ENSSSCG00000043123*
	6	6_820343	820,343	0.327 (C/T)	2.13±0.4	2.78	1.11e–07	Intron	*CBFA2T3*
	8	8_127473136	127,473,136	0.032 (A/C)	–3.01±0.89	0.78	8.22e–07	Intergenic	*ENSSSCG00000061497*
	12	12_6931728	6,931,728	0.309 (T/TCTC)	1.64±0.39	1.60	1.42e–07	Intergenic	*RPL38*
	13	rs1109946846	18,674,793	0.478 (G/A)	1.85±0.39	2.38	3.22e–07	Intron	*CNOT10*
	14	rs323501665	134,642,571	0.318 (G/C)	–2.16±0.42	2.82	1.43e–07	Intergenic	*ENSSSCG00000054450*

[1)] It showed the position of the associated variants on the *Sus scrofa* Build 11.1 assembly. The variants that do not possess rsID were named by the author based on the physical positions that they are mapped to the pig reference genome.
[2)] Chr, chromosome.
[3)] Pos, physical position.
[4)] MAF, minor allele frequency.
[5)] Beta, beta estimates. SE, standard errors for beta.
[6)] Var, percentage of phenotypic variations explained by the variants.

TP content (Fig. 3-D). Moreover, we found that the minor allele of SNP 6_820343 significantly increased TP content (*P*=0.0008; Fig. 3-E).

To prioritize candidate genes proximal to GWAS loci associated with purine traits, we performed a GO and a KEGG pathway analysis for genes that were distributed within the ±500 kb of peak variants for all TP QTLs using Metascape. Notably, the analysis revealed the enrichment

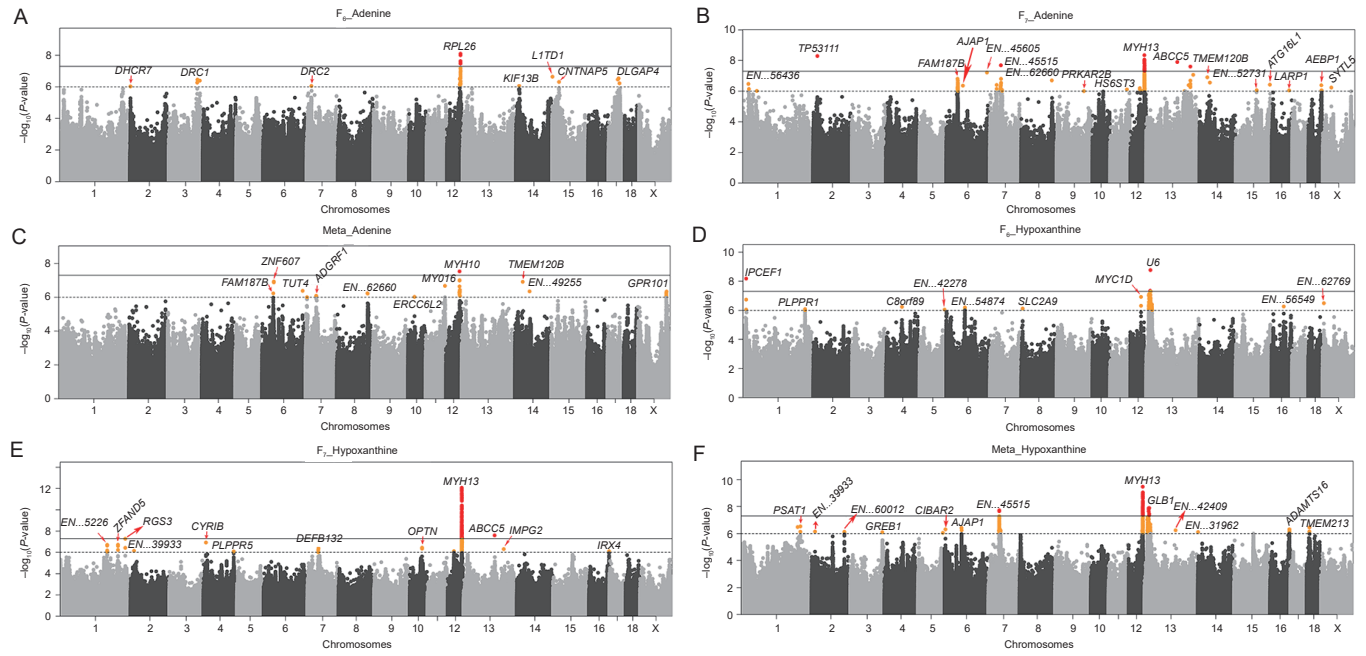


Fig. 2 Manhattan plots of the association analyses for the content of adenine and hypoxanthine. A–F, single-population GWAS and meta-analysis for adenine (A–C) and hypoxanthine content (D–F).

of seven significant GO terms (P<0.05) that are directly related to purine metabolism (Fig. 4-A). These GO terms encompassed various processes such as purine ribonucleotide metabolic process, purine nucleotide biosynthetic process, purine-containing compound metabolic process. Furthermore, the examination of these related GO terms led to the identification of a set of 20 candidate genes (Appendix G).

To gain further insights into the functions and interconnections of these 20 genes, a functional network was constructed using the HumanNet database and Cytoscape. As depicted in Fig. 4-B, the resulting network demonstrated the connectivity among 15 candidate genes, of which 6 (denoted by blue dots) were highlighted along with their connected genes (denoted by green dots) as participants in diverse purine-related pathways (indicated by red words). The results indicate that these 15 genes represent strong candidates for TP. Subsequently, we evaluated the impact of the 11 lead SNPs located adjacent to these 15 genes on TP and the content of three individual purine bases (Appendix G). Notably, two of the peak SNPs (9_72327563 and 10_63740648) showed a negative association with TP as well as each purine base (Appendices C and D). Most of the remaining SNPs influenced TP through their effects on one or two purine bases, with hypoxanthine frequently being involved. Interestingly, SNP 17_10473884 appeared to affect TP by specifically influencing adenine.

3.5. Fine-mapping of the QTL on SSC12 for guanine

To identify potential causal genes/variants underlying the most significant QTL for guanine, which was centered around the peak SNP 12_55495763 on SSC12, we conducted a comprehensive screening process. Within a 25 kb up- and down-stream of 12_55495763, we identified three candidate genes, two protein-coding genes *TMEM238L* and phosphoinositide interacting regulator of transient receptor potential channels (*PIRT*), and a lncRNA *ENSSSCG00000043821* (Fig. 5-A). Furthermore, we observed 277 variants that displayed significant associations (P<1×10^{-6}) with guanine content. Notably, the QTL region, supported by a three logarithm of the odds (LOD) drop-off method, spanned merely 7 kb, wherein the top nine additional variants exhibited strong LD (r^2>0.8) with the SNP 12_55495763. Importantly, all these prominent variants were located in the downstream of *TMEM238L*, except for SNP 12_55500274 resided in the intergenic region between the *TMEM238L* and *ENSSSCG00000043821*. Assuming that the QTL effect did not arise from unexamined structural variants (such as copy number variations, inversions, translocations, large-sequence insertions, and deletions), these identified top variants have the potential to serve as causal variants underlying the QTL.

Considering that these top variants do not induce changes in the coding sequence, we then determined whether they reside within a regulatory region. By utilizing publicly available epigenetic data from skeletal muscle in pigs (Kern *et al.* 2021), we identified a region spanning approximately 27 kb (55.483–55.510 Mb) that is marked by H3K27me3, a well-established epigenetic marker of polycomb-repressed elements, which has been associated with the downregulation of neighboring genes. The presence of these top variants within this regulatory element suggests their potential functional significance, warranting further investigation. In addition, the impact of the SSC12 QTL on guanine content was evaluated by examining the lead SNP 12_55495763. We found that the G/A ((13.33±1.38) mg 0.1 kg^{-1}) and A/A

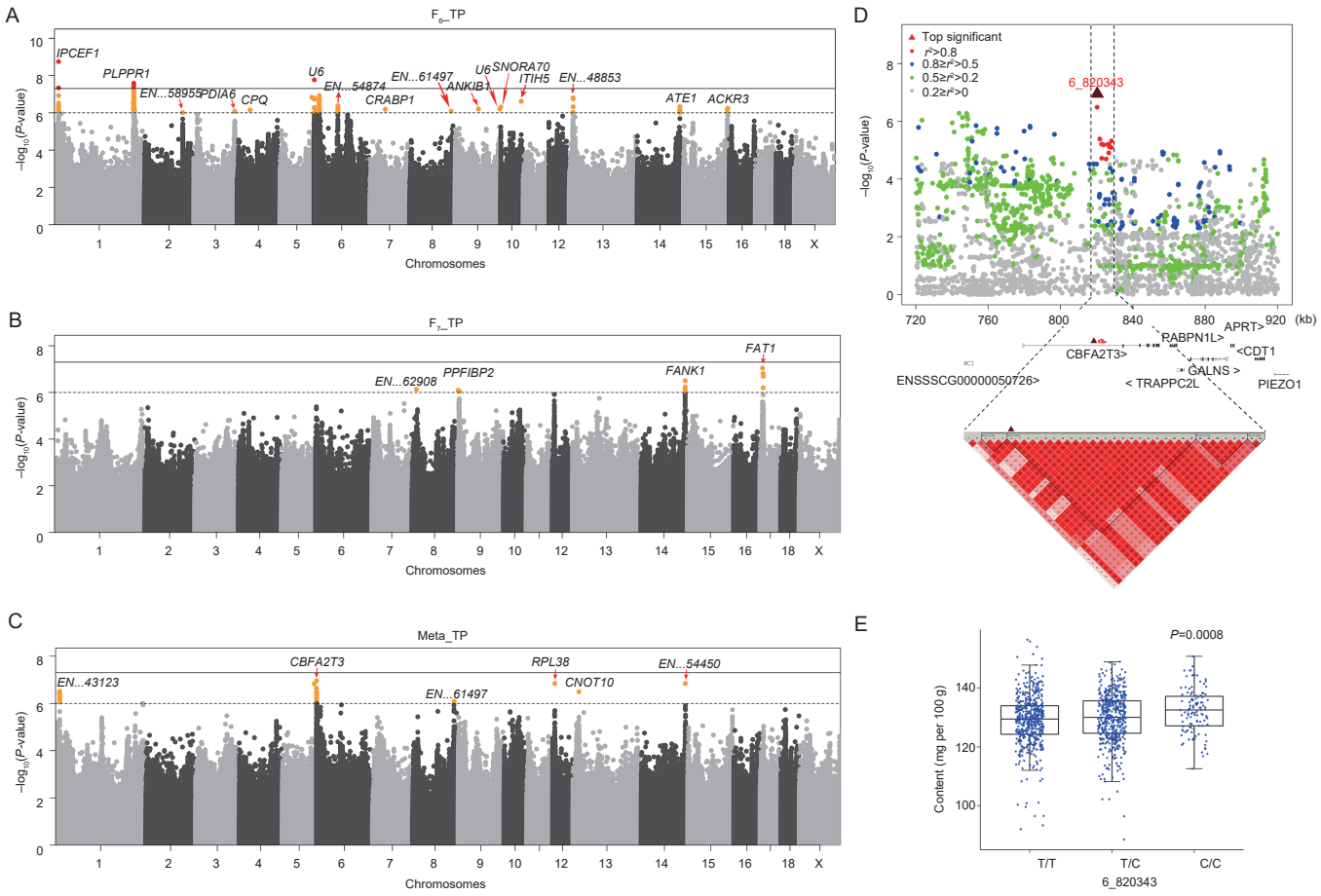


Fig. 3 Results of association analyses for total purine (TP) content and focus on the major QTL on SSC6. A–C, Manhattan plots of GWAS and meta-analysis for TP content. D, regional plot showing the candidate region (SSC6: 720–920 kb) containing the top SNP (6_820343) identified by the meta-analysis. The top SNP is depicted by a brown triangle, and different levels of linkage disequilibrium (LD) between 6_820343 and its surrounding SNPs are indicated in different colors. The positions of lncRNA and protein-coding genes in this region are based on *Sus scrofa* 11.1 genome assembly (release 106). The confidence interval (SSC6: 820–828 kb) was determined through 3-LOD drop-off. E, evaluation of allelic substitution effect of the leading SNP 6_820343 on TP content. *P*-value was calculated by one-way analysis of variance (ANOVA). SSC, *Sus scrofa* chromosome; LOD, logarithm of odds.

((13.22±1.35) mg 0.1 kg^{-1}) genotypes exhibited significantly higher guanine content (P<2e−16) compared to the G/G ((12.46±1.21) mg 0.1 kg^{-1}) genotype (Fig. 5-B), which suggests that the QTL primarily exhibits a dominant effect in influencing guanine content.

3.6. Fine-mapping of the QTL on SSC12 for adenine and hypoxanthine

As previously mentioned, we believe that the identified QTLs for adenine and hypoxanthine contents on SSC12 in the F_7 population, characterized by peak SNPs rs343341698 (P=6.61×10^{-9}) and rs328426812 (P=5.94×10^{-14}), respectively (Fig. 6-A and B), might have arisen from the same underlying causal mutation. Interestingly, the minor alleles of these lead SNPs were associated with increased adenine content and decreased hypoxanthine content (Fig. 6-C and D), suggesting a reverse effect of this QTL on the variation in the two traits. To identify the potential causative variant(s), we conducted fine-mapping of the QTLs using 3-LOD drop-off method. This analysis revealed an overlapping region of approximately 171 kb (54,924,843–55,096,610 bp) within the two QTL intervals (54,924,531–55,116,425 bp; 54,924,843–55,096,610 bp) (Fig. 6-A and B). This overlapping region is considered the most likely interval for the QTL and encompasses two annotated genes, *GAS7* and *MYH13*. Within these two genes, a total of 1,617 SNPs and 458 indels were identified.

Subsequently, we investigated whether the variants found in the exons of the *GAS7* and *MYH13* genes could potentially be the causative mutation(s). Using the Ensembl Variant Effect Predictor (VEP) tool, we assessed the functional implications of all thirteen missense mutations identified in the myosin heavy chain 3 (*MYH3*) gene (Appendix H). Among these variants, two SNPs, namely rs1110730305 (p.Arg1140His) and rs338479386 (p.Ile1520Thr), emerged as promising candidates due to their relatively higher evolutionary conservation, as indicated by genomic evolutionary rate profiling (GERP) scores greater than 1.5, and their potentially deleterious effect on protein function, as indicated by sorting intolerant from tolerant (SIFT) scores less than 0.05. Moreover, these two SNPs, especially rs1110730305, displayed a stronger association with adenine and hypoxanthine contents than other missense variants.

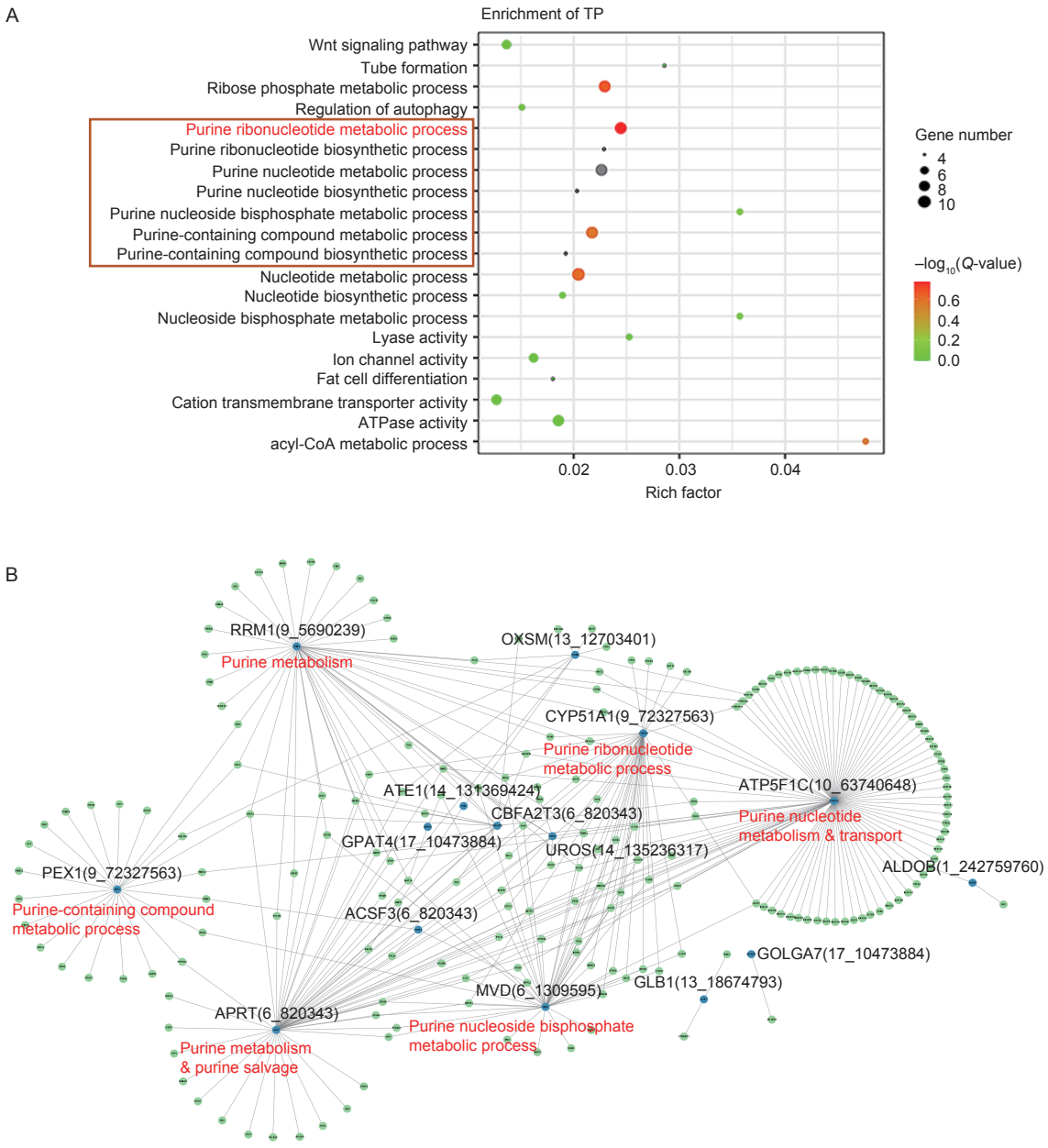


Fig. 4 Enrichment analysis and gene network construction of candidate genes for total purine content (TP). A, enrichment analysis of TP-associated GO terms, including twenty selected terms with seven directly related to purine metabolism. The *y*-axis displays the labels of selected terms, and the *x*-axis denotes the rich factor. The size and color of each bubble reflect the number of candidate genes enriched in the pathway and the level of enrichment significance, respectively. B, gene network constructed using HumanNet and Cytoscape, based on candidate genes and their interacting genes. Blue dots represent node genes and green dots represent predictive connected genes. Red text indicates KEGG pathways or GO biological processes (P<0.01) in which the node genes and its connected genes participate.

It is important to note, however, that the significance of rs1110730305 and rs338479386 were at least three orders of magnitude lower than that of the lead SNPs rs343341698 and rs328426812 (Fig. 6-A and B), and the two missense variants were only in moderate LD with the lead SNPs (r^2≈0.55). To further understand the reasons behind the comparatively lower significance of rs1110730305 and rs338479386 in relation to rs343341698 and rs328426812, haplotypes consisting of the four SNPs were constructed. In the F_7 pigs, three haplotypes, denoted as HAP-1, HAP-2 and HAP-3, were identified with respective frequencies of 0.89, 0.05 and 0.04 (Fig. 6-E). Comparative analysis of the effects of the three haplotypes on the purine phenotypes revealed that HAP-1 most likely carried the wild-type *q* allele of the SSC12 QTL (assuming two alleles, "*Q*" and "*q*", for the QTL), whereas HAP-2 and HAP-3 harbored the mutant *Q* allele (Fig. 6-E and F). Consequently, the alleles of rs343341698 and rs328426812 were found to co-segregate with the QTL allele, while this co-segregation was not observed for rs1110730305 and rs338479386. Therefore, these findings suggest that the two identified missense mutations in *MYH3* were insufficient to account for the observed QTL effect.

To screen candidate genes associated with the content of adenine and hypoxanthine within the SSC12 QTL, we

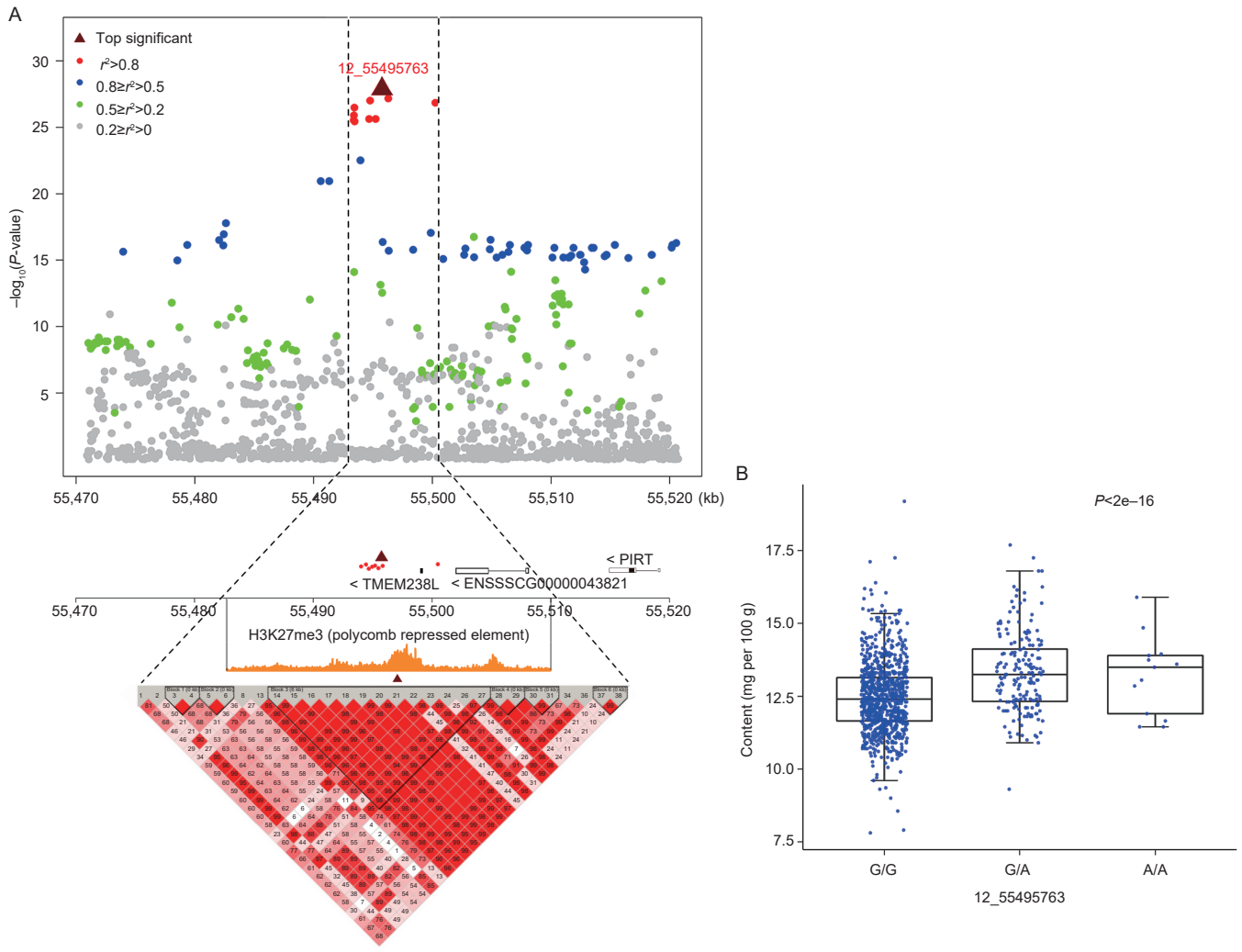


Fig. 5 Regional plot illustrating a major QTL for guanine content on SSC12 identified through meta-analysis. A, QTL candidate region (SSC12: 55,470–55,521 kb) encompasses the top SNP (12_55495763) associated with guanine content. The top SNP is denoted by a red triangle, and different levels of LD between the most significant SNP and surrounding SNPs are indicated in different colors. The arrangement of lncRNAs and protein-coding genes in this region is based on the *Sus scrofa* 11.1 genome assembly (release 106). Epigenetic annotation in this region was obtained from skeletal muscle samples of pigs provided by the University of California. The confidence interval (SSC12: 55,493–55,500 kb, shown between dashed lines) was determined using a logarithm of odds (LOD) drop off of 3. B, evaluation of allelic substitution effect of top SNP (12_55495763). *P*-value was calculated by one-way analysis of variance (ANOVA). SSC, *Sus scrofa* chromosome; LD, linkage disequilibrium.

initially utilized RNA-seq data from muscle samples of the F_7 population (n=290) to detect expression quantitative trait loci (eQTLs). In the genomic region spanning from 54,500 to 56,500 kb, which encompasses a total of 29 protein coding genes, we discovered a cohort of eight candidate genes, myosin heavy chain (*MYH1*), myosin heavy chain (*MYH2*), myosin heavy chain (*MYH4*), myosin heavy chain (*MYH6*), myosin heavy chain (*MYH8*), synthesis of cytochrome C oxidase 1 (*SCO1*), ADP-Ribose/CDP-alcohol diphosphatase, manganese dependent (*ADPRM*), and transmembrane protein 220 (*TMEM220*), which exhibited significant *cis*-eQTL (P<0.05). Then we examined the correlation between mRNA expression levels of these *cis*-regulated genes and the content of adenine and hypoxanthine. The result revealed significant correlations solely between the mRNA expression levels of the *ADPRM*, *SCO1*, and *TMEM220* genes, and the adenine content in F_7 pig, with r=–0.16 (P=0.0073), r=–0.17 (P=0.0031) and r=–0.21 (P=0.0003), respectively (Appendices E and I). However, the significance levels of the *cis*-eQTL peak SNPs for these three genes were considerably lower than that of GWAS peak SNP associated with adenine content, and the LD values between them were quite low (r^2<0.05). These findings indicate that none of these three genes are likely to be the causative gene responsible for the variation in adenine content.

4. Discussion

A harmonious supply of purines is essential for cell growth, proliferation, and survival. Moreover, purines have been shown to play important roles in the physiology of muscles (Cicero *et al.* 2023) and the formation of meat quality. This study pioneers the exploration of the genetic architecture of guanine,

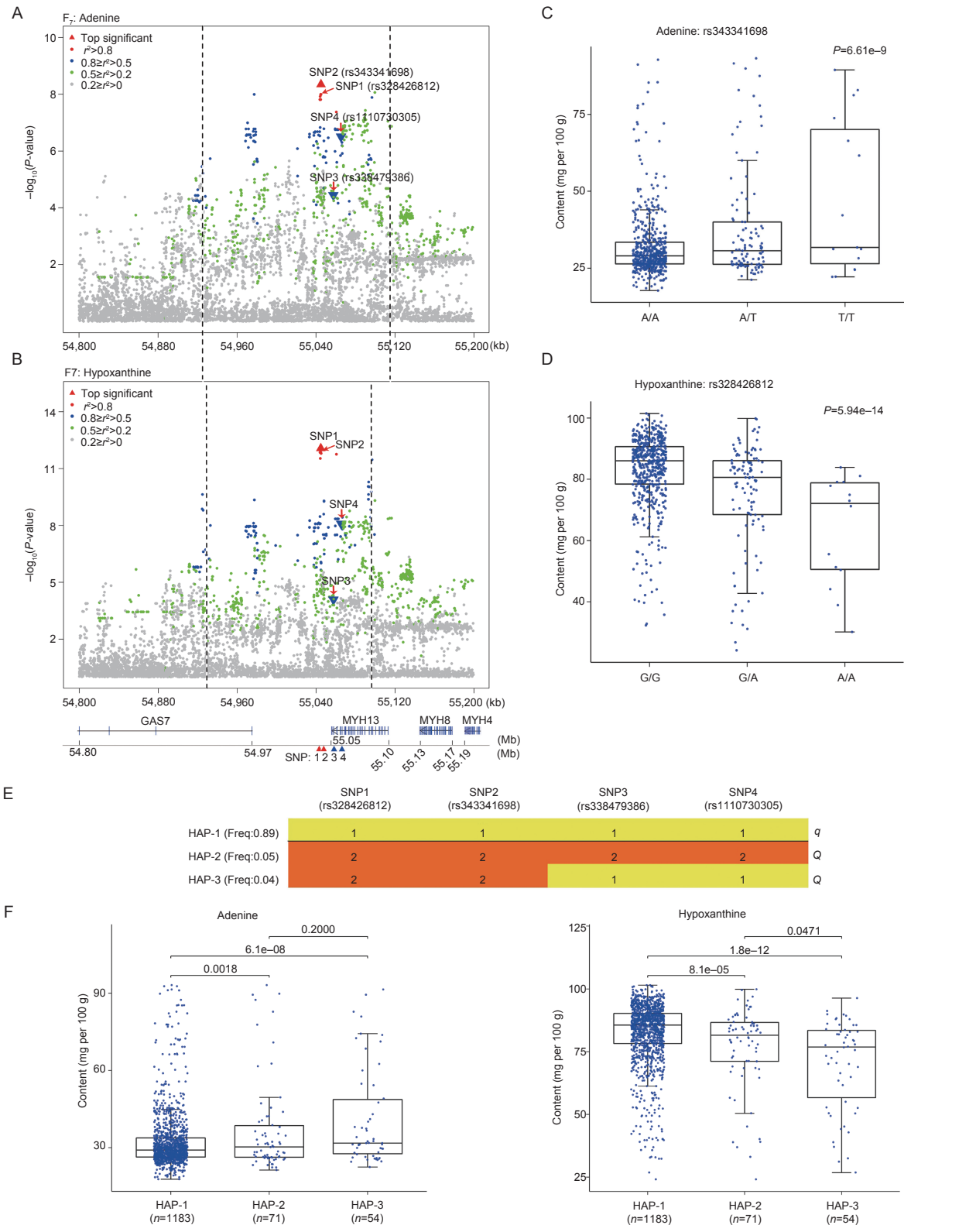


Fig. 6 Effect evaluation of four candidate causal variants for adenine and hypoxanthine. A and B, regional plots (SSC12: 54,800–55,200 kb) displaying the association analysis results for the content of adenine (upper) and hypoxanthine (bottom) in F_7 pig population. The two top SNPs (SNP1: rs328426812; and SNP2: rs343341698) were indicated by red triangles, while two missense SNPs (SNP3: rs338479386; and SNP4: rs1110730305) in *MYH13* were indicated by blue inverted triangles. QTL interval was depicted by two dashed lines, determined through 3-LOD drop-off analysis. C and D, allelic substitution effect of the two top SNPs on purine levels evaluated by one-way ANOVA. E, haplotypes identified in the F_7 pig population, comprising four SNPs (SNP1–4). Major alleles and minor alleles for each SNP marker were denoted as 1 and 2, respectively. The identities and frequencies of haplotypes were given on the left axis, while the estimated effect (*Q* or *q*) was shown on the right axis. F, comparison of the effects of the three haplotypes on the content of adenine (left panel) and hypoxanthine (right panel) in the F_7 pig population. SSC, *Sus scrofa* chromosome; SNP, single nucleotide polymorphism; *MYH13*, myosin heavy chain 13; LOD, logarithm of odds.

adenine, hypoxanthine, and TP contents in swine muscle. The heritability estimates for the four purine traits, calculated in both the F_6 and F_7 pig populations, manifest moderate levels (0.14–0.35), akin to the heritability of many other meat quality attributes (Lopez *et al.* 2020). However, substantial difference in heritability estimates for adenine was observed between F_6 and F_7, contrasting with the other three purine traits (guanine, adenine, and TP). The elevated variability in phenotypic values (Table 1) and a greater count of GWAS top variants associated with adenine content in F_7 may account for its doubled heritability compared to F_6 (0.30 *vs*. 0.14). Notably, our GWAS and meta-analyses revealed the presence of more than 10 QTLs for each purine trait, indicating the multifaceted genetic regulation underlying their concentrations.

While the genes closest to the most significantly associated loci may not always represent the causal factors, they nevertheless present as promising candidates warranting focused analysis, particularly when their gene annotations evince relevance to the studied phenotype. Our observation revealed that a subset of the nearest genes flanking the GWAS peak loci may be intricately involved in purine and nucleotide metabolism, as exemplified by 5′-nucleotidase ecto (*NT5E*), RasGEF domain family member 1B (*RASGEF1B*), and endonuclease domain containing 1 (*ENDOD1*). *NT5E*'s gene product is responsible for catalyzing the conversion of extracellular nucleotides into membrane-permeable nucleosides, while the proteins encoded by *RASGEF1B* and *ENDOD1* have been implicated in the regulation of guanyl-nucleotide exchange factor activity and nucleic acid binding activity, respectively.

Furthermore, it is noted that several genes proximate to GWAS peaks function as transporters of small molecules, as exemplified by solute carrier family 6 member 15 (*SLC6A15*) and solute carrier family 2 member 9 (*SLC2A9*). Specifically, *SLC6A15* facilitates the transportation of neutral amino acids, while *SLC2A9* acts as a high-capacity urate transporter (Caulfield *et al.* 2008) as well as high-affinity, low-capacity glucose and fructose transporter. In the F_6 population, *SLC6A15* emerged as the most proximate candidate gene for the guanine-associated loci on SSC5, whereas *SLC2A9* takes precedence as the closest candidate for hypoxanthine-associated loci on SSC8. Additionally, the *ABCC5* gene, a member of the superfamily of ATP-binding cassette transporters mediating ATP-dependent transport of endogenous metabolites (including cAMP and cGMP), was found in immediate adjacency to the loci for adenine and hypoxanthine content on SSC13 in the F_7 population. Coincidentally, some human association studies have shown that the *SLC2A9*, solute carrier family 22 member 12 (*SLC22A12*) and ATP binding cassette subfamily G member 2 (*ABCG2*) genes (Sulem *et al.* 2011; Tin *et al.* 2011; Okada *et al.* 2012; Kottgen *et al.* 2013; Li *et al.* 2015), which are paralogous to the three aforementioned genes, have impacts on serum urate concentrations, hyperuricamia and gout. This intriguing alignment strongly implies the existence of shared genetic mechanisms of purine metabolism between pigs and humans. Disturbances in purine metabolism within human muscle tissue may correlate with certain disorders, such as Duchenne dystrophy (Camiaa *et al.* 1995). Given the pig as a pertinent model for human diseases, exploring the genetic mechanisms governing purine composition in porcine muscle holds promise for advancing our understanding of the genetic underpinnings of human muscular disorders.

Beyond the identification of the most proximate candidate genes, our approaches involving gene set enrichment analysis and gene network analysis allow for the recognition of additional candidates and their involved pathways associated with the investigated traits. Among the genes implicated in these purine-related GO terms, we identified several genes that encode enzymes engaging in energy metabolism, including ATP Synthase F1 Subunit Gamma (*ATP5F1C*), Mevalonate Diphosphate Decarboxylase (*MVD*), Acyl-CoA Synthetase Family Member 3 (*ACSF3*), and Adenine Phosphoribosyltransferase (*APRT*). To further prioritize the potential causal genes, we employed a comprehensive, large-scale human gene functional interaction network known as HumanNet. Notably, the efficacy of HumanNet in prioritizing candidate disease genes, even in conjunction with GWAS data, has been well-documented (Lee *et al.* 2011). Our findings illuminated a compelling pattern — node genes such as ribonucleotide reductase catalytic subunit M1 (*RRM1*), *APRT*, and *ATP5F1C*, along with their interconnected genes, exhibited marked and significant enrichment in purine-related terms or pathways (P<0.001). Within this network of node genes (15 in total, represented by blue dots in Fig. 4-B), seven genes including *ACSF3*, aldolase, fructose-bisphosphate B (*ALDOB*), *APRT*, arginyltransferase 1 (*ATE1*), glycerol-3-phosphate acyltransferase 4 (*GPAT4*), peroxisomal biogenesis factor 1 (*PEX1*), and *RRM1* had been previously reported to be directly or indirectly involved in various purine-related biological processes, including purine metabolism, uric acid levels, salvage pathway, and the purine nucleotide biosynthesis pathway (Taylor *et al.* 1977; Lanaspa *et al.* 2012; Agrawal *et al.* 2014; Zhang *et al.* 2015; Liang *et al.* 2023). In addition, the assessment of the effects of strongly associated SNPs around the node genes on the concentrations of the three individual purine bases indicate that most of these genes affect the TP by collectively modulating the levels of multiple purines.

The most significant QTLs controlling the levels of individual purine bases were just all mapped within the genomic region spanning from 54,800 to 55,200 kb on SSC12. Thus, we subsequently performed in-depth analyses of these QTLs including fine mapping and identification of potential causal variants and genes. Interestingly, the results of LD analysis among the peak loci for these purines indicate that this candidate region contains a shared genetic determinant exerting opposing effects on adenine and hypoxanthine levels (resulting in an absence of association with TP), alongside an independent genetic factor modulating guanine level. Notably, these QTL loci exhibited considerable effect, with biallelic variants prompt over 20% changes in phenotypic values (Figs. 5-B and 6-C and D).

Further investigation into the guanine-associated QTL region on SSC12 revealed that corresponding peak variants were concentrated within a 6-kb LD block, embedded in an annotated polycomb repressed element, marked by H3K27me3 (Fig. 5-A). This observation leads to the

hypothesis that the causal mutation(s) underlying the guanine QTL may reside within this regulatory element. Such a mutation might regulate guanine level by perturbing the element's function and thereby influencing the expression of its target gene(s). This hypothesis, however, needs further meticulous investigation for validation. Moreover, the guanine QTL peak was positioned upstream of the *TMEM238L* gene, followed by the lncRNA *ENSSSCG00000043821*. *TMEM238L* encodes transmembrane protein 238 like, potentially implicated in inducing apoptosis during endoplasmic reticulum (ER) stress and inhibiting of proliferation and tumorigenicity. In contrast, the function of *ENSSSCG00000043821* remains elusive. Both genes merit careful consideration as plausible candidates for this specific QTL.

Regarding the adenine and hypoxanthine-associated QTL region on SSC12, a clustering of peak SNPs was observed within a segment situated less than 3 kb downstream of the 3′-end of the *MYH13* gene. None of the missense mutations in the refined QTL interval encompassing only *MYH13* and *GAS7* could account for the observed QTL effect, suggesting a likely scenario in which the causal variant(s) underlying the QTL are regulatory in nature. However, neither *MYH13* nor *GAS7* displayed a *cis*-eQTL in the F_7 population. Given that certain regulatory elements, like enhancers, can wield distant regulatory effects on gene expression, the search for candidate genes flanked by this QTL was expanded. However, even with extensive *cis*-eQTL and QTT analysis, a singular, most probable candidate gene remains elusive. This complexity possibly arises due to the presence of multiple members of the myosin heavy chains (Myh) family within this genomic region, encompassing *MYH1*, *MYH3*, *MYH4*, *MYH8*, and *MYH13*. Owing to their high sequence homology (>98.7%), accurately aligning their transcripts from second-generation transcriptome sequencing with these genes proves intricate, thus undermining quantification accuracy. Utilizing third-generation sequencing technology, renowned for long read capability, hold promise in enhancing transcript identification and quantification, further refining *cis*-eQTL mapping. Moreover, in light of myosin's role in ATP hydrolysis for muscle contraction, it is crucial to undertake further investigations to identify the specific Myh gene accountable for the adenine-associated QTL.

In a prior study, we demonstrated a significant correlation between purine composition in meat and various meat quality attributes (Zheng *et al.* 2018; Huang *et al.* 2020). Remarkably, we have noted overlaps between the QTLs associated with purines on SSC12 and those affecting meat redness, intramuscular fat content (IMF) (Cho *et al.* 2019), and muscle fiber characteristics (Huang *et al.* 2022). Consequently, it becomes intriguing to investigate the pleiotropic impacts of the genetic loci influencing purines as identified here, as well as their potential connections to reported GWAS loci for other meat quality traits. Such exploration will contribute to revealing the molecular mechanisms underpinning these QTLs, thus prompting their application in breeding programs.

5. Conclusion

In summary, this investigation has provided unprecedented insights into the genetic architecture of purine composition in pork. All purine traits studied are moderately inheritable. The count of identified QTLs ($P<1\times10^{-6}$) associated with guanine, adenine, hypoxanthine, and TP content amounted to 14, 36, 19, and 25, respectively. Notably, the most significant QTLs for guanine, adenine and hypoxanthine levels, except for TP, have been localized in a 400-kb region on SSC12, which encompasses a cluster of Myh family genes and genetic loci influencing multiple meat quality traits. Application of third-generation sequencing techniques or larger study populations is imperative to pinpoint causal genes and mutations in this region. Among the nearest genes proximal to GWAS peak variants, specific noteworthy candidates emerge. For instance, *NT5E*, engaged in catalyzing purine compound, and *SLC2A9*, involved in glucose and urate transport, hold significant relevance. Furthermore, gene enrichment and network analyses have highlighted a group of even more promising candidate genes for TP, showcasing significant enrichment in various purine-related metabolic pathways.

Acknowledgements

The authors are grateful for the support from the National Natural Science Foundation of China (32272855) and STI 2030-Major Projects, China (2023ZD0404501).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that they did not use AI in the preparation and writing of this manuscript.

Ethical approval

All animal procedures were performed in accordance with the Guidelines for the Care and Use of Experimental Animals approved by the State Council of the People's Republic of China. The ethics committee of Jiangxi Agricultural University approved this entire study (GNDF2014-018).

Appendices associated with this paper are available at https://doi.org/10.1016/j.jia.2024.05.004

References

Agrawal A A, McLaughlin K J, Jenkins J L, Kielkopf C L. 2014. Structure-guided U2AF65 variant improves recognition and splicing of a defective pre-mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 17420–17425.

Barkas F, Elisaf M, Liberopoulos E, Kalaitzidis R, Liamis G. 2018. Uric acid and incident chronic kidney disease in dyslipidemic individuals. *Current Medical Research and Opinion*, **34**, 1193–1199.

Barrett J C, Fry B, Maller J, Daly M J. 2005. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Bhole V, Choi J W, Kim S W, de Vera M, Choi H. 2010. Serum uric acid levels and the risk of type 2 diabetes: A prospective study. *The American Journal of Medicine*, **123**, 957–961.

Blanco J, Canela E J, Sayos J, Mallol J, Llyis C, Franco R. 1993. Adenine nucleotides and adenosine metabolism in pig kidney proximal tubule membranes. *Journal of Cellular Physiology*, **157**, 77–83.

Browning S R, Browning B L. 2007. Rapid and accurate haplotype

phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, **81**, 1084–1097.

Camiaa F, Novo-Rodriguez M I, Rodriguez-Segade S, Castro-Gago M. 1995. Purine and carnitine metabolism in muscle of patients with Duchenne muscular dystrophy. *Clinica Chimica Acta*, **243**, 151–164.

Caulfield M J, Munroe P B, O'Neill D, Witkowska K, Charchar F J, Doblado M, Evans S, Eyheramendy S, Onipinla A, Howard P, Shaw-Hawkins S, Dobson R J, Wallace C, Newhouse S J, Brown M, Connell J M, Dominiczak A, Farrall M, Lathrop G M, Samani N J, *et al.* 2008. SLC2A9 is a high-capacity urate transporter in humans. *PLoS Medicine*, **5**, e197.

Chang C, Chow C, Tellier L, Vattikuti S, Purcell S, Lee J. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.

Cho I, Park H, Ahn J, Han , Lee J, Lim H, Yoo C, Jung E, Kim D, Sun W, Ramayo-Caldas Y, Kim S, Kang Y, Kim Y, Shin H, Seong P, Hwang I, Park B, Hwang S, Lee S, *et al.* 2019. A functional regulatory variant of MYH3 influences muscle fiber-type composition and intramuscular fat content in pigs. *PLoS Genetics*, **15**, e1008279.

Choi H K, Atkinson K, Karlson E W, Willett W, Curhan G. 2004. Purine-rich foods, dairy and protein intake, and the risk of gout in men. *The New England Journal of Medicine*, **350**, 1093–1103.

Cicero A F G, Fogacci F, Di Micoli V, Angeloni C, Giovannini M, Borghi C. 2023. Purine metabolism dysfunctions: experimental methods of detection and diagnostic potential. *International Journal of Molecular Sciences*, **24**, 1–14.

Danve A, Sehra S T, Neogi T. 2021. Role of diet in hyperuricemia and gout. *Best Practice & Research in Clinical Rheumatology*, **35**, 101723.

Davoli R, Ros-Freixedes R, Gol S, Pena R N, Tor M, Ibáñez-Escriche N, Dekkers J C M, Estany J. 2016. Genome-wide association study singles out SCD and LEPR as the two main loci influencing intramuscular fat content and fatty acid composition in Duroc pigs. *PLoS ONE*, **11**, e0152496.

Dehlin M, Jacobsson L, Roddy E. 2020. Global epidemiology of gout: Prevalence, incidence, treatment patterns and risk factors. *Nature Reviews Rheumatology*, **16**, 380–390.

Devlin B. 1999. Genomic control for association studies. *Biometrics*, **55**, 997–1004.

Dobin A, Davis C A, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras T R. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Duan Y, Guo Q, Wen C, Wang W, Li Y, Tan B, Li F, Yin Y. 2016. Free amino acid profile and expression of genes implicated in protein metabolism in skeletal muscle of growing pigs fed low-protein diets supplemented with branched-chain amino acids. *Journal of Agricultural and Food Chemistry*, **64**, 9390–9400.

Duval N, Luhrs K, Wilkinson T G, Baresova V, Skopova V, Kmoch S, Vacano G N, Zikanova M, Patterson D. 2013. Genetic and metabolomic analysis of AdeD and AdeI mutants of *de novo* purine biosynthesis: Cellular models of *de novo* purine biosynthesis deficiency disorders. *Molecular Genetics and Metabolism*, **108**, 178–189.

Falconer D S, Mackay Trudy F C. 1996. *Introduction to Quantitative Genetics*. Longman Group Ltd. Publishing, Edinburgh Gate, UK. pp. 1-479.

Freedman M L, Reich D, Penney K L, McDonald G J, Mignault A A, Patterson N, Gabriel S B, Topol E J, Smoller J W, Pato C N, Pato M T, Petryshen T L, Kolonel L N, Lander E S, Sklar P, Henderson B, Hirschhorn J N, Altshuler D. 2004. Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, **36**, 388–393.

Guo T, Gao J, Yang B, Yan G, Xiao S, Zhang Z, Huang L. 2020. A whole genome sequence association study of muscle fiber traits in a White Duroc×Erhualian F_2 resource population. *Asian–Australasian Journal of Animal Sciences*, **33**, 704–711.

Huang C, Zheng M, Huang Y, Liu X, Zhong L, Ji J, Zhou L, Zeng Q, Ma J, Huang L. 2020. The effect of purine content on sensory quality of pork. *Meat Science*, **172**, 108346.

Huang Y, Cai L, Duan Y, Zeng Q, He M, Wu Z, Zou X, Zhou M, Zhang Z, Xiao S, Yang B, Ma J, Huang L. 2022. Whole-genome sequence-based association analyses on an eight-breed crossed heterogeneous stock of pigs reveal the genetic basis of skeletal muscle fiber characteristics. *Meat Science*, **194**, 108974.

Ji J, Zhou L, Huang Y, Zheng M, Liu X, Zhang Y, Huang C, Peng S, Zeng Q, Zhong L, Yang B, Li W, Xiao S, Ma J, Huang L. 2018. A whole-genome sequence based association study on pork eating quality traits and cooking loss in a specially designed heterogeneous F_6 pig population. *Meat Science*, **146**, 160–167.

Kanbay M, Segal M, Afsar B, Kang D H, Rodriguez-Iturbe B, Johnson R J. 2013. The role of uric acid in the pathogenesis of human cardiovascular disease. *Heart*, **99**, 759–766.

Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, Saelao P, Waters S, Xiang R, Chamberlain A, Korf I, Delany M E, Cheng H H, Medrano J F, Van Eenennaam A L, Tuggle C K, Ernst C, Flicek P, Quon G, Ross P, *et al.* 2021. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nature Communication*, **12**, 1821.

Kottgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, Pistis G, Ruggiero D, O'Seaghdha C M, Haller T, Yang Q, Tanaka T, Johnson A D, Kutalik Z, Smith A V, Shi J, Struchalin M, Middelberg R P, Brown M J, Gaffo A L, *et al.* 2013. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nature Genetics*, **45**, 145–154.

Lanaspa M A, Sanchez-Lozada L G, Choi Y J, Cicerchi C, Kanbay M, Roncal-Jimenez C A, Ishimoto T, Li N, Marek G, Duranay M, Schreiner G, Rodriguez-Iturbe B, Nakagawa T, Kang D H, Sautin Y Y, Johnson R J. 2012. Uric acid induces hepatic steatosis by generation of mitochondrial oxidative stress: Potential role in fructose-dependent and -independent fatty liver. The *Journal of Biological Chemistry*, **287**, 40732–40744.

Lee I, Blom U M, Wang P I, Shim J E, Marcotte E M. 2011. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Research*, **21**, 1109–1121.

Li C, Li Z, Liu S, Wang C, Han L, Cui L, Zhou J, Zou H, Liu Z, Chen J, Cheng X, Zhou Z, Ding C, Wang M, Chen T, Cui Y, He H, Zhang K, Yin C, Wang Y, *et al.* 2015. Genome-wide association analysis identifies three new risk loci for gout arthritis in Han Chinese. *Nature Communication*, **6**, 7041.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li H D, Lund M S, Christensen O F, Gregersen V R, Henckel P, Bendixen C. 2010. Quantitative trait loci analysis of swine meat quality traits. *Journal of Animal Science*, **88**, 2904–2912.

Liang J S, Hung K L, Lin L J, Ong W P, Keng W T, Lu J F. 2023. Novel PEX1 mutations in fibroblasts from children with Zellweger spectrum disorders exhibit temperature sensitive characteristics. *Epilepsy & Behavior*, **145**,1–9.

Liao Y, Smyth G K, Shi W. 2014. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

Liu X, Xiong X, Yang J, Zhou L, Yang B, Ai H, Ma H, Xie X, Huang Y, Fang S, Xiao S, Ren J, Ma J, Huang L. 2015. Genome-wide association analyses for meat quality traits in Chinese Erhualian pigs and a Western Duroc×(Landrace×Yorkshire) commercial population. *Genetics Selection Evolution*, **47**, 44.

Lopez B I, Santiago K G, Lee D, Cho Y, Lim D, Seo K. 2020. Single-step genomic evaluation for meat quality traits, sensory characteristics, and fatty-acid composition in Duroc pigs. *Genes* (Basel), **11**, 1–13.

Lourenco V M, Pires A M, Kirst M. 2011. Robust linear regression methods in association studies. *Bioinformatics*, **27**, 815–821.

Love M I, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Ma J, Yang J, Zhou L, Ren J, Liu X, Zhang H, Yang B, Zhang Z, Ma H, Xie X, Xing Y, Guo Y, Huang L. 2014. A splice mutation in the *PHKG1* gene causes high glycogen content and low meat quality in pig skeletal muscle. *PLoS Genetics*, **10**, e1004710.

Marchini J, Cardon L R, Phillips M S, Donnelly P. 2004. The effects of human population structure on large genetic association studies. *Nature Genetics*, **36**, 512–517.

Okada Y, Sim X, Go M J, Wu J Y, Gu D, Takeuchi F, Takahashi A, Maeda S, Tsunoda T, Chen P, Lim S C, Wong T Y, Liu J, Young T L, Aung T, Seielstad M, Teo Y Y, Kim Y J, Lee J Y, Han B G, *et al.* 2012. Meta-analysis identifies multiple loci associated with kidney

function-related traits in east Asian populations. *Nature Genetics*, **44**, 904–909.
Pearson T A, Manolio T A. 2008. How to interpret a genome-wide association study. *JAMA*, **299**, 1335–1344.
Pertea M, Pertea G M, Antonescu C M, Chang T C, Mendell J T, Salzberg S L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**, 290–295.
Randall C, Johnson G W N, Jennifer L Troyer, James A Lautenberger, Bailey D Kessing, Cheryl A Winkler, Stephen J O'Brien. 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, **11**, 1–6
Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg S R F, Consortium W G S, Wilkie A O M, McVean G, Lunter G. 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics*, **46**, 912–918.
Rong S, Zou L, Zhang Y, Zhang G, Li X, Li M, Yang F, Li C, He Y, Guan H, Guo Y, Wang D, Cui X, Ye H, Liu F, Pan H, Yang Y. 2015. Determination of purine contents in different parts of pork and beef by high performance liquid chromatography. *Food Chemistry*, **170**, 303–307.
Smith E U, Diaz-Torne C, Perez-Ruiz F, March L M. 2010. Epidemiology of gout: An update. *Best Practice & Research in Clinical Rheumatology*, **24**, 811–827.
Smith H S, Bracken D, Smith J M. 2011. Gout: Current insights and future perspectives. *The Journal of Pain*, **12**, 1113–1129.
So A, Thorens B. 2010. Uric acid transport and disease. *The Journal of Clinical Investigation*, **120**, 1791–1799.
Soare E, Chiurciu I A. 2017. Study on the pork market worldwide. *Scientific Papers Series Management*, *Economic Engineering in Agriculture and Rural Development*, **17**, 321–326.
Sulem P, Gudbjartsson D F, Walters G B, Helgadottir H T, Helgason A, Gudjonsson S A, Zanon C, Besenbacher S, Bjornsdottir G, Magnusson O T, Magnusson G, Hjartarson E, Saemundsdottir J, Gylfason A, Jonasdottir A, Holm H, Karason A, Rafnar T, Stefansson H, Andreassen O A, *et al.* 2011. Identification of low-frequency variants associated with gout and serum uric acid levels. *Nature Genetics*, **43**, 1127–1130.
Taylor M W, Pipkorn J H, Tokito M K, Pozzatti Jr R O. 1977. Purine mutants of mammalian cell lines: Iii control of purine biosynthesis in adenine phosphoribosyl transferase mutants of CHO cells. *Somatic Cell Genetics*, **3**, 195–206.
Tin A, Woodward O M, Kao W H L, Liu C-T, Lu X, Nalls M A, Shriner D, Semmo M, Akylbekova E L, Wyatt S B, Hwang S J, Yang Q, Zonderman A B, Adeyemo A A, Palmer C, Meng Y, Reilly M, Shlipak M G, Siscovick D, Evans M K, Rotimi C N, *et al.* 2011. Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Human Molecular Genetics*, **20**, 4056–4068.
Viterbo V S, Lopez B I M, Kang H, Kim H, Song C W, Seo K S. 2018. Genome wide association study of fatty acid composition in Duroc swine. *Asian–Australasian Journal of Animal Sciences*, **31**, 1127–1133.
Willer C J, Li Y, Abecasis G R. 2010. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
Wu B, Roseland J M, Haytowitz D B, Pehrsson P R, Ershow A G. 2019. Availability and quality of published data on the purine content of foods, alcoholic beverages, and dietary supplements. *Journal of Food Composition and Analysis*, **84**, 1–8.
Wu Z, Gong H, Zhou Z, Jiang T, Lin Z, Li J, Xiao S, Yang B, Huang L. 2022. Mapping short tandem repeats for liver gene expression traits helps prioritize potential causal variants for complex traits in pigs. *Journal of Animal Science Biotechnology*, **13**, 8.
Yang J, Lee S H, Goddard M E, Visscher P M. 2011. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, **88**, 76–82.
Zhang F, Patel D M, Colavita K, Rodionova I, Buckley B, Scott D A, Kumar A, Shabalina S A, Saha S, Chernov M, Osterman A L, Kashina A. 2015. Arginylation regulates purine nucleotide biosynthesis by enhancing the activity of phosphoribosyl pyrophosphate synthase. *Nature Communication*, **6**, 7517.
Zheng M, Huang Y, Ji J, Xiao S, Ma J, Huang L. 2018. Effects of breeds, tissues and genders on purine contents in pork and the relationships between purine content and other meat quality traits. *Meat Science*, **143**, 81–86.
Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821–824.
Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi A H, Tanaseichuk O, Benner C, Chanda S K. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communication*, **10**, 1523.

Executive Editor-in-Chief Kui Li
Managing Editor Han Zhang