

1 **PIGOME: An Integrated and Comprehensive Multi-omics**
2 **Database for Pig Functional Genomics Studies**

3

4 Guohao Han^{1,3,4,#}, Peng Yang^{1,2,3,4,5,#}, Yongjin Zhang^{3,4}, Qiaowei Li^{1,3,4,6}, Xinhao
5 Fan^{1,3,4}, Ruipu Chen^{1,3,4}, Chao Yan^{1,3,4}, Mu Zeng^{1,3,4,7}, Yalan Yang^{1,3,4*}, Zhonglin
6 Tang^{1,2,3,4*}

7

8 ¹ *Kunpeng Institute of Modern Agriculture at Foshan, Agricultural Genomics Institute,*
9 *Chinese Academy of Agricultural Sciences, Foshan 528200, China.*

10 ² *School of Life Sciences, Henan University, Kaifeng 475004, China.*

11 ³ *Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key*
12 *Laboratory of Livestock and Poultry Multi-omics of MARA, Agricultural Genomics*
13 *Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124,*
14 *China.*

15 ⁴ *GuangXi Engineering Centre for Resource Development of Bama Xiang Pig, Bama*
16 *547500, China.*

17 ⁵ *Shenzhen Research Institute of Henan University, Shenzhen 518000, China.*

18 ⁶ *School of Veterinary Medicine, University College Dublin, Belfield D04 VIW8*
19 *Dublin 4, Ireland.*

20 ⁷ *Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise*
21 *Breeding, Key Laboratory of Animal Molecular Design and Precise Breeding of*
22 *Guangdong Higher Education Institutes, School of Life Science and Engineering,*
23 *Foshan University, Foshan 528225, China.*

24

25 # Equal contribution.

26 * Corresponding authors.

27 E-mail: tangzhonglin@caas.cn (Tang ZL), yangyalan@caas.cn (Yang YL)

28

29 **Running title:** *Han G et al / PIGOME: Multi-omics Database in Pig*

30 **Abstract**

31 In addition to being a major source of animal protein, pigs are important model for the
32 study of development and diseases in humans. During the past two decades, thousands
33 of high-throughput sequencing studies in pigs have been performed using a variety of
34 tissues from different breeds and developmental stages. However, the multi-omics
35 database specifically used for pig functional genomic research is still limited. Here,
36 we present a user-friendly database of pig multi-omics named PIGOME. PIGOME
37 contains seven types of pig omics datasets, including whole-genome sequencing,
38 RNA-seq, miRNA-seq, ChIP-seq, ATAC-seq, bisulfite-seq, and MeRIP-seq, from
39 6,901 samples and 392 projects with manually curated metadata, integrated gene
40 annotation, and quantitative trait locus information. Furthermore, various ‘explore and
41 browse’ functions have been established for user-friendly access to omics information.
42 PIGOME implemented several tools to visualize genomic variants, gene expression,
43 and epigenetic signals of a given gene in the pig genome, enabling efficient
44 exploration of spatial-temporal expression/epigenetic pattern, function, regulatory
45 mechanism, and associated economic traits. Collectively, PIGOME provides valuable
46 resources for pig breeding and is helpful for human biomedical research. PIGOME is
47 available at <https://pigome.com>.

48

49 **KEYWORDS:** Pig, Multi-omics, Gene expression, Epigenetics, Database

50

51 **Introduction**

52 Pig production accounts for a large proportion of the animal husbandry economy and
53 is one of the mainstays of the global agricultural economy [1,2]. Moreover, pigs have
54 been shown to be an important biomedical model for the study on human
55 development and diseases [3–5]. Local adaptation and artificial selection have
56 resulted in significant phenotypic differences and genetic diversity in pigs [6]. It is an
57 exceptional model to elucidate the underlying mechanisms of traits, such as meat
58 production, litter size, coat color, immune and diseases [7,8]. In the past several
59 decades, with the development of advanced sequencing technologies, massive
60 amounts of high-throughput sequencing data have been generated at multi-omics
61 layers. These massive data provide a better understanding of evolution, selection, trait
62 formation, development, and diseases in the genetic mechanisms and uncover many
63 key variants, genes, and regulatory elements regulating various biological processes
64 and associated with economic traits in pigs [9–11]. Our recent study based on
65 high-resolution DNA methylome and transcriptome in skeletal muscle at 27
66 developmental stages provided insights into the molecular regulation of skeletal
67 muscle development and diversity and uncovered candidate genes, such as *IGF2BP3*
68 and *SATB2*, which contributes to skeletal muscle development, that provides a
69 representative case to integrate multi-omics data facilitating the functional genomics
70 studies of pigs [6,12]. Therefore, it is necessary to integrate multi-omics data to
71 support the scientific discovery of pig genetics and breeding.

72 In particular, a mounting number of high-throughput sequencing studies in pigs
73 have been performed based on a variety of tissues from different breeds and
74 developmental stages [13–17]. However, these datasets are generated from different
75 laboratories and sequencing platforms, making their retrieval, management, standard
76 processing, and visualization time-consuming and difficult [18]. Furthermore, mining
77 and integrated analysis of these datasets to explore the biological functions and
78 regulatory mechanism remains a challenge [19]. Over the past several years, a limited
79 number of specialized pig-related databases have been developed. Recently, IAnimal

80 (<https://ianimal.pro/>) [20] was released, which includes pig multi-omic and genome
81 annotation information. Similarly, ISwine (<http://iswine.iomics.pro/>) [21] contains
82 published pig genome, transcriptome, and QTL data, similar to quantitative trait locus
83 (QTL) information. To date, the Animal Omics Database
84 (<http://animal.nwsuaf.edu.cn/>) contains only pan-genome sequencing datasets.
85 However, the analysis and visualization abilities of these databases are limited (see
86 Table 2). There is still a lack of comprehensive multi-omics databases dedicated to
87 functional genomic research on pigs.

88 To address these challenges, we developed PIGOME, an integrated and
89 comprehensive web database containing seven types of sequencing data from 6,901
90 datasets and 392 projects, which is currently the most comprehensive omics database
91 for pigs. PIGOME allows researchers to explore and utilize pig multi-omics data
92 easily and effectively. Specifically, PIGOME supports the exploration, analysis, and
93 visualization of genomic variations, expression patterns, regulatory networks, and
94 epigenetic modifications of annotated and predicted pig genes (protein-coding genes
95 (PCGs), microRNAs (miRNAs), and circular RNAs (circRNAs)). PIGOME also
96 supports a tissue-specific analysis tool that allows users to identify the characteristics
97 of genes, miRNAs, and circRNAs in specific tissues. In addition, PIGOME deploys
98 nine tools, such as JBrowse [22], IGV [23], and Sequenceserver [24] to enable users
99 to upload their files to visualize epigenetic signals and perform sequence alignment
100 across the genomes of different pig breeds. In summary, PIGOME is an important
101 resource for pig functional genomics studies and will be of interest to a broad
102 readership in the fields of animal genetics, breeding, and biomedical research.

103

104 **Data collection and database construction**

105 **Data collection**

106 Seven types of high-throughput sequencing data (whole-genome sequencing [WGS] ,
107 transcriptome sequencing [RNA-seq], microRNA sequencing [miRNA-seq],
108 chromatin immunoprecipitation sequencing [ChIP-seq], assay for

109 transposase-accessible chromatin sequencing [ATAC-seq], bisulfite sequencing for
110 DNA methylation analysis [BS-seq] and methylated RNA immunoprecipitation
111 sequencing [MeRIP-seq]) of pigs were collected from NCBI Sequence Read Archive
112 (SRA, <https://www.ncbi.nlm.nih.gov/sra/>) and CNCB Genome Sequence Archive
113 (GSA, <https://ngdc.cncb.ac.cn/gsa/>). BS-seq contains two types of data, including
114 whole genome bisulfite sequencing (WGBS) and reduced representation bisulfite
115 sequencing (RRBS).

116 WGS datasets were employed to construct the pan-genome and identify genomic
117 variants, including single nucleotide polymorphisms (SNPs) and DNA insertion and
118 deletion (InDels), within the genome. RNA-seq and miRNA-seq datasets were used to
119 analyze the expression of mRNA and ncRNAs (such as miRNA, lncRNA, and
120 circRNAs). CHIP-seq datasets were used to identify the modification of CTCF,
121 histone modifications, and POL2 across the genome. The ATAC-seq datasets were
122 used to identify open chromatin regions across the genome. BS-seq and MeRIP-seq
123 datasets were used to analyze genome-wide DNA and RNA methylation, respectively.

124 All these datasets were manually collected with all the relevant metadata for fast
125 and accurate data retrieval and statistical analysis, including project ID, tissue,
126 developmental stage, breed, read number, platform, references, and others of the same
127 kind. For the developmental stage information, referring to our previous paper [12], it
128 was classify into 27 known stages (E33 to D180) and two fuzzy developmental stages
129 ‘Unknown’ and ‘Adult’. Samples with poor data quality (mapping rate < 30% and
130 data volume < 0.15 Gb) were excluded manually. To explore gene functions more
131 conveniently, gene annotation information was integrated from Ensembl 100 [25],
132 miRbase 22.1 [26] and eggNOG 5 [27], while QTL information was integrated from
133 Animal QTLdb 46 [28] (Figure 1).

134

135 **Data processing**

136 The FASTA file of *Sus scrofa* reference genome (build 11.1) and the GTF annotation
137 file (release 100) were downloaded from the Ensembl database. All raw FASTQ files

138 were downloaded from the NCBI SRA and CNGB GSA databases. Fastp (v0.20.0)
139 [29] was used to trim and filter the raw reads. For RNA-seq, HISAT2 (v2.0.5) [30]
140 was used to map to the reference genome. Gene expression quantification in
141 transcripts per kb exon per million mapped reads (TPM) was calculated using
142 StringTie (v1.3.6) [31]. For miRNA-seq, adapters were removed using Cutadapt (v1.8.
143 dev0) [32]. After remove adapters, these reads were then aligned to annotated pig
144 miRNAs gathered from miRBase [26] and the pig reference genome using miRDeep2
145 (v0.1.2) [33]. For circRNAs, HISAT2 (v2.0.5) was used for alignment with the
146 reference genome. Novel circRNAs were identified using CIRIquant (v1.1.1) [34] and
147 Find_circ (v2) [35]. For ATAC-seq and ChIP-seq, all reads were mapped using
148 Bowtie2 (v2.3.5.1) [36]. The peaks were identified using MACS2 (v2.2.6) and
149 annotated using SnpEff (v4.2) [37]. Bismark (v0.23.0) [38] was used to align the
150 reference genome using default parameters. All CpG sites were identified and
151 annotated as previously described [12]. For MeRIP-seq, all reads were mapped using
152 HISAT2 (v2.0.5), and peaks were identified using exomePeak2 (v2) [39] and
153 annotated using SnpEff (v4.2). Picard (v2.25.7) was used to remove duplicate PCR
154 reads for ATAC-seq, ChIP-seq, and MeRIP-seq. BigWig files for genome browser
155 visualization were generated using deepTools (v3.5.1) [40] and bedGraphToBigWig
156 (v4). For WGS, all reads were aligned to the reference genome using BWA (v0.7.12)
157 [41]. The SNP and InDel calling used a Unified Genotyper approach as implemented
158 in the GATK package (v4.1.5.0). Considering that the volume of SNP data is too large,
159 SNPs in intergenic regions was not shown in the website, but could be accessed from
160 the corresponding authors upon reasonable request. Furthermore, tissue-specific genes
161 were identified using the R package TissueEnrich (v3.15) [42] based on the
162 expression matrix of mRNA and ncRNAs. The rcorr function of Hmisc (v5.1-1) in R
163 package was used to calculate the expression correlation between mRNA, miRNA
164 and circRNA to construct the co-expression networks with $r > 0.85$ and P value < 0.01
165 as thresholds. The intersection results of RNAhybrid (v2.1.2) [43] and miRanda

166 (v3.3a) [44] were used to predict putative targets for mRNAs and circRNAs with E
167 value < -20.

168

169 **Website implementation**

170 PIGOME was built by Thinkphp 6.0.12 (<https://www.thinkphp.cn/>), a mature
171 model-view-controller (MVC) framework, deployed in CentOS 7.9 system. All omics
172 data were stored in MySQL 5.6.50 (<https://www.mysql.com/>). Web interfaces were
173 developed using HTML, CSS, JavaScript and Bootstrap 5.0.2
174 (<https://getbootstrap.com/>). Most of the interactive charts and tables were
175 implemented with ECharts 5.3.1 (<https://echarts.apache.org/>) and Bootstrap Table
176 1.14.2 (<https://bootstrap-table.com/>) (Figure 1). Network proxy services were
177 provided through Nginx 1.20.1 (<https://www.nginx.com/>). We recommend visiting
178 PIGOME using Google Chrome, Microsoft Edge, or Mozilla Firefox.

179

180 **Database content and usage**

181 **Data collection and statistics**

182 At present, PIGOME v1.0 collects 7 types of multi-omics datasets in pigs, including
183 WGS, RNA-seq, miRNA-seq, CHIP-seq, ATAC-seq, BS-seq (WGBS and RRBS) and
184 MeRIP-seq. It contained 6,901 samples from 392 projects, including 113 breeds, 71
185 tissues, and 29 developmental stages. The total clean data reached 49.21 Tb (Table 1
186 and 2). The RNA-seq database represents the most abundant datasets in our database,
187 including 4,217 samples, 74 breeds, 50 tissues, and 29 developmental stages (Table 1).
188 To better interpret the omics data, we integrated 32,452 gene annotations and 29,687
189 QTLs. Gene annotation information records commonly have 22 attributes, including
190 gene symbol, gene type, description, muscle biology, GO, KEGG, CAZy, and PFAM.
191 In addition, the QTL information collected 11 attributes, mainly including position,
192 QTL ID, name, type, trait, and PubMed ID. Additional statistics are summarized on
193 the statistics page (<https://pigome.com/statistics.html>).

194

195 **PIGOME features and functions**

196 PIGOME includes genomics (SNPs, InDels, and genome annotation), epigenomics
197 (chromatin accessibility, histone modifications, DNA/RNA methylation), and
198 transcriptomics (the abundance of mRNA and ncRNAs). In addition, it provides
199 useful and user-friendly functions to help users perform advanced analyses (Figure 1).

200

201 *Browse*

202 Users can easily browse omics data using a Browse tag in the toolbar. After clicking
203 on the omics data type, a summary information related to the data will be displayed.

204 On the summary page of each level of omics data, users can obtain specific statistical
205 data, including sample, gene, and other related information, and freely download
206 these tables and charts. For more details, users can click the icon in the ‘Details’
207 column of a given gene or sample information table on the page, which links to the
208 gene expression page. Basic information about the gene or sample is placed at the top
209 of the gene expression page, with links to external databases. Different gene
210 expression pages contained different sections. Specifically, the gene expression pages
211 of RNA-seq, miRNA-seq, and circRNA can show the gene TPM values in various
212 tissues, breeds, or developmental stages in given tissues, also display the TPM values
213 of subgroups of samples freely selected by users. In addition, to better understand the
214 gene function, it integrates a variety of gene annotations. For better finding the
215 co-regulation between genes, the page shows the gene network of query gene. The
216 expression pattern of a user given gene can be visualized in boxplots, bar charts, and
217 line charts, moreover, the data will be conveniently presented in a table below the
218 graph. Furthermore, the detail page of ATAC-seq, bisulfite-seq, ChIP-seq, and
219 MeRIP-seq provides an IGV genome browser and a table to display information,
220 allowing users to freely explore any genome intervals of each sample. Additionally,
221 users can browse the allele frequency of a certain loci in different breeds on the detail
222 page of SNP and InDel using a bar chart and table. This page also provides QTL
223 information related to this region. In summary, PIGOME has a variety of browsing

224 functions, paving the way for the integration and investigation of different omics in
225 pigs.

226

227 *Explore*

228 For convenient usage, PIGOME provides three search engines to explore the whole
229 database, including ‘by gene ID or symbol’, ‘by range’, and ‘by sample’. For ‘by gene
230 ID or symbol’, users can explore by inputting gene ID or symbol. On the ‘by range’
231 page, users can explore by selecting chromosomes and entering the starting position
232 and ending position. On the results page of ‘by gene ID or symbol’ and ‘by range’, all
233 datasets related to a given gene or a given range are integrated and displayed.
234 Additionally, users can obtain more detailed information by clicking on the links in
235 the table. For ‘by sample’, users can fuzzily explore by selecting the dataset and
236 entering SRR ID, Sample ID, or Project ID. Furthermore, the results page of ‘by
237 sample’ displays the relevant sample information and provides relevant links.
238 Importantly, all figures and data of the search results can be downloaded and edited
239 easily.

240

241 *Genome browser*

242 PIGOME embeds a custom genome browser based on JBrowse2 to help users
243 compare and analyze various omics datasets. PIGOME contains information on the
244 sequence and gene annotations from Ensembl. By inputting the genome range or gene
245 ID, users can explore the omics data related to the gene of interest. All tracks were
246 marked according to the type of omics data, tissue, breed, and developmental stage. In
247 the tracking group, the tracks of interest can be displayed by switching the
248 checkboxes.

249

250 *Tools*

251 We have integrated nine practical tools, including IGV, JBrowse, ‘Get Sequence’,
252 ‘Primer Design’, BLAST, ‘Gene Network’, ‘Target prediction’ ‘Find tissue-specific

253 genes', and API (Table 2). For IGV and JBrowse, users can check, verify, and
254 interpret their own sequencing and genome data online. For 'Get Sequence', users can
255 quickly extract the required gene sequence from large number of nucleotide
256 sequences. Then, 'Primer Design' tools [45] can help users to design primers from
257 DNA/RNA sequences of interest for further experimental verification. For BLAST,
258 users can perform an alignment analysis based on their own sequences with 23 pig
259 genomes. 'Gene Network' tool is helpful to find the gene regulatory network formed
260 by the interaction between genes. For 'Target prediction', users can explore the
261 regulatory role of miRNA in gene expression and find potential functional miRNA
262 associated with economical traits. Furthermore, a tool called 'Find tissue-specific
263 genes' was developed, that helps users quickly find tissue-specific genes, miRNAs,
264 and circRNAs based on our massive expression data. Additionally, for API tools,
265 users with basic programming skills can obtain the omics data more flexibly and
266 explore functional genes more effectively. These tools will assist us to better explore
267 the biological mechanisms of various biological processes and important economic
268 traits in pigs.

269

270 Additionally, users can easily find more help from the database through the 'Help tag'
271 in the toolbar.

272

273 **Comparison with the existing databases**

274 To date, several user-friendly databases have been established to aggregate
275 multi-omics datasets in pigs (Table 2). IAnimal [20] is a multi-species and
276 multi-omics database, encompassing four types of pig omics data, including WGS,
277 RNA-Seq, ChIP-Seq and ATAC-Seq. ISwine [21] serves as a professional pig omics
278 database offering access to WGS, RNA-seq, quantitative traits and annotation
279 information. While both IAnimal and ISwine provide valuable sample
280 meta-information, including details on tissue, developmental stage, and breed, they
281 lack secondary classification and correction of this metadata. Consequently,

282 comparative analyses of gene expression regulation between developmental stages or
283 breeds become challenging. The AOD database focuses on providing 12 *de novo*
284 genome assemblies in pigs. However, PIGOME emerges as a standout platform in this
285 landscape. Notably, it boasts the widest array of omics data types and meticulously
286 curated sample meta-information. This refinement facilitates the exploration of gene
287 expression and regulation differences across developmental stages, breeds and tissues
288 (Table 1). Furthermore, PIGOME offers nine practical tools designed to enhance the
289 utilization of multi-omics datasets, a feature comparable to that of IAnimal (Table 2).
290 Additionally, PIGOME facilitates the exploration of tissue-specific mRNA and
291 ncRNA functional genomics in a more convenient manner, as we showed in the next
292 section.

293

294 **Case study**

295 Herein, we provide a case study to verify the usefulness of ‘Find tissue-specific genes’
296 in PIGOME and illustrate how to use PIGOME mining multi-omics information of
297 interested genes. Initially, users can select the option ‘skeletal muscle’ in the ‘find
298 tissue-specifically expressed gene’ section in the tool and thereafter, click the
299 ‘Explore’ button. On the results page, based on substantial expression data, users can
300 find 186 genes that are specifically expressed in skeletal muscle and then click the
301 view icon of ‘*ENSSSCG00000026533*’ to explore more detailed expression
302 information about this gene (Figure 2A). On the expression page, users can obtain the
303 gene symbol of myogenic factor 6 (*MYF6*), also known as *MRF4*, a myogenic
304 regulatory factor involved in myogenesis. In addition, users can first see the related
305 gene annotation and visualize its expression in different tissues using bar, line, or
306 box-plot charts (Figure 2B and C). Importantly, users can also explore the expression
307 trend of genes in skeletal muscles of different breeds and at different developmental
308 stages (Figure 2D), demonstrating the ability to explore potential specific genes.

309 Finally, users can obtain all omics information related to genes using an
310 exploration function. Users can use ‘*ENSSSCG00000026533*’ or ‘*MYF6*’ as the input

311 in ‘Explore by gene ID or symbol’. The results page provides information, including
312 SNP variation, expression abundance, annotation, epigenomics, and QTL related to
313 *MYF6*. More importantly, one circRNA was identified in *MYF6* (Figure 3A).
314 Serendipitously, by clicking the view icon (Figure 3B), it was shown that this
315 circRNA (circ-MYF6), which may be a candidate circRNA that affects the
316 development and growth of skeletal muscle, is also specifically expressed in skeletal
317 muscle. There were 140 peaks identified from the ChIP-seq data, and 24 open
318 chromatin regions were identified from ATAC-seq in *MYF6* (Figure 3C and D).
319 Furthermore, the results showed 3,970 CpG methylation sites in exons, introns, and
320 upstream regions (Figure 3E). Notably, we detected 24 SNPs and InDels in *MYF6*
321 (Figure 3F). These results indicated that PIGOME can be used to explore the potential
322 regulatory mechanisms of these genes.

323

324 **Discussion**

325 In the past few decades, researchers have made great efforts in functional genomics
326 research of pigs and accumulate valuable omics data [46]. Compared to earlier
327 released databases for pigs, such as IAnimal [20], ISwine [21], and AOD [47],
328 PIGOME contains the most data types and the most up-to-date multi-omics datasets
329 covering comprehensive meta-information (Table 2). Moreover, PIGOME provides a
330 user-friendly interface for browsing and analyzing omics data via interactive
331 webpages, powerful search engines, and advanced tools. Integrated genomic,
332 transcriptomic, and epigenomic data provide an efficient approach for discovering
333 target genes and loci associated with economic traits and human-related diseases.

334 With the continuous development and innovation of high-throughput sequencing
335 methods, more technologies have been developed, such as scRNA-seq and spatial
336 transcriptomics [48,49]. The amount of omics data in public databases is also
337 increasing. PIGOME will also constantly update new omics types and quantities. In
338 the near future, other types of variations in the pig genome, such as structure
339 variations (SVs), copy number variations (CNVs) and presence/absence variations

340 (PAVs) will be available in PIGOME. We will aim to focus on cutting-edge
341 single-cell sequencing and increase display related data, such as scRNA-seq,
342 scATAC-seq and spatial transcriptome. In addition, PIGOME will update the latest
343 gene annotation, genome-, epigenome- and transcriptome-wide association studies
344 (GWAS, EWAS and TWAS) and applies quantitative trait locus (xQTL) information
345 to help users better understand gene function. Furthermore, we will increase the
346 internal relations among various data in the database and develop rich online tools.
347 Finally, we intend PIGOME to be an important resource for exploring pig functional
348 genomics, and will be of interest to the broad readership in the fields of animal
349 genetics, breeding, and biomedical research.

350

351 **Data availability**

352 PIGOME is available at <https://pigome.com>.

353

354 **CRedit author statement**

355 **Guohao Han:** Methodology, Software, Visualization, Writing - Original Draft. **Peng Yang:**
356 Methodology, Software, Data curation, Formal analysis. **Yongjin Zhang:** Data curation,
357 Formal analysis. **Qiaowei Li:** Formal analysis. **Xinhao Fan:** Formal analysis. **Ruipu Chen:**
358 Formal analysis. **Chao Yan:** Formal analysis. **Mu Zeng:** Formal analysis. **Yalan Yang:**
359 Conceptualization, Project administration, Funding acquisition, Writing - review & editing.
360 **Zhonglin Tang:** Conceptualization, Supervision, Funding acquisition, Writing - review &
361 editing. All authors read and approved the final manuscript.

362

363 **Competing interests**

364 The authors have declared no competing interests.

365

366 **Acknowledgments**

367 We acknowledge the works of all the omics data researchers. This work was
368 supported by the National Key Scientific Research Project [2023YFF1001100], the
369 National Natural Science Foundation of China [U23A20229 and 32172697] and
370 Agricultural Science and Technology Innovation Program [CAAS-ZDRW202006].

371

372 **ORCID**

373 ORCID 0000-0001-8648-7347 (Guohao Han)

374 ORCID 0000-0002-0941-669X (Peng Yang)

375 ORCID 0009-0003-3962-7505 (Yongjin Zhang)

376 ORCID 0000-0002-1878-7863 (Qiaowei Li)

377 ORCID 0009-0004-2208-0133 (Xinhao Fan)

378 ORCID 0009-0005-7501-580X (Ruipu Chen)

379 ORCID 0000-0003-3066-8258 (Chao Yan)

380 ORCID 0009-0003-2375-8666 (Mu Zeng)

381 ORCID 0000-0003-4912-1869 (Yalan Yang)

382 ORCID 0000-0002-4538-4349 (Zhonglin Tang)

383

384 **Reference**

- 385 [1] Zhang S, Wu X, Han D, Hou Y, Tan J, Kim SW, et al. Pork production systems in
386 China: A review of their development, challenges and prospects in green production.
387 *Front Agr Sci Eng* 2021;8:15.
- 388 [2] Wu Y, Zhao J, Xu C, Ma N, He T, Zhao J, et al. Progress towards pig nutrition in the
389 last 27 years. *J Sci Food Agric* 2020;100:5102–10.
- 390 [3] Lunney JK, Van Goor A, Walker KE, Hailstock T, Franklin J, Dai C. Importance of the
391 pig as a human biomedical model. *Sci Transl Med* 2021;13:eabd5758.
- 392 [4] Tu C-F, Chuang C, Yang T-S. The application of new breeding technology based on
393 gene editing in pig industry-A review. *Anim Biosci* 2022;35:791–803.
- 394 [5] Yang H, Wu Z. Genome Editing of Pigs for Agriculture and Biomedicine. *Front Genet*
395 2018;9:360.

- 396 [6] Yang Y, Yan J, Fan X, Chen J, Wang Z, Liu X, et al. The genome variation and
397 developmental transcriptome maps reveal genetic differentiation of skeletal muscle in
398 pigs. *PLoS Genet* 2021;17:e1009910.
- 399 [7] Kyselova J, Tichy L, Jochova K. The role of molecular genetics in animal breeding: A
400 minireview. *Czech J Anim Sci* 2021;66:107–11.
- 401 [8] Kasper C, Ribeiro D, de Almeida AM, Larzul C, Liaubet L, Murani E. Omics
402 application in animal science-A special emphasis on stress response and damaging
403 behaviour in pigs. *Genes* 2020;11:920.
- 404 [9] Long JA. The “omics” revolution: Use of genomic, transcriptomic, proteomic and
405 metabolomic tools to predict male reproductive traits that impact fertility in livestock
406 and poultry. *Anim Reprod Sci* 2020;220:106354.
- 407 [10] Yang Y, Zhou R, Li K. Future livestock breeding: Precision breeding based on
408 multi-omics information and population personalization. *J Integr Agric*
409 2017;16:2784–91.
- 410 [11] Koltjes JE, Cole JB, Clemmens R, Dilger RN, Kramer LM, Lunney JK, et al. A vision
411 for development and utilization of high-throughput phenotyping and big data analytics
412 in livestock. *Front Genet* 2019;10:1197.
- 413 [12] Yang Y, Fan X, Yan J, Chen M, Zhu M, Tang Y, et al. A comprehensive epigenome
414 atlas reveals DNA methylation regulating skeletal muscle development. *Nucleic Acids*
415 *Res* 2021;49:1313–29.
- 416 [13] Kim J-M, Park J-E, Yoo I, Han J, Kim N, Lim W-J, et al. Integrated transcriptomes
417 throughout swine oestrous cycle reveal dynamic changes in reproductive tissues
418 interacting networks. *Sci Rep* 2018;8:5436.
- 419 [14] Jin L, Tang Q, Hu S, Chen Z, Zhou X, Zeng B, et al. A pig BodyMap transcriptome
420 reveals diverse tissue physiologies and evolutionary dynamics of transcription. *Nat*
421 *Commun* 2021;12:3715. <https://doi.org/10.1038/s41467-021-23560-8>.
- 422 [15] Liu Y, Liu Y, Ma T, Long H, Niu L, Zhang X, et al. A splicing mutation in PHKG1
423 decreased its expression in skeletal muscle and caused PSE meat in Duroc × Luchuan
424 crossbred pigs. *Anim Genet* 2019;50:395–8.

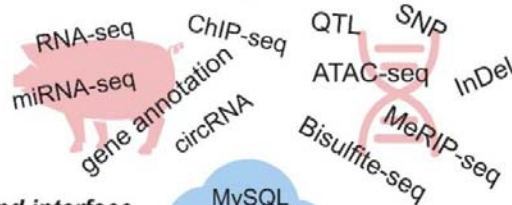
- 425 [16] Zhao Y, Hou Y, Xu Y, Luan Y, Zhou H, Qi X, et al. A compendium and comparative
426 epigenomics analysis of cis-regulatory elements in the pig genome. *Nat Commun*
427 2021;12:2217.
- 428 [17] Kogelman LJA, Cirera S, Zhernakova DV, Fredholm M, Franke L, Kadarmideen HN.
429 Identification of co-expression gene networks, regulatory genes and pathways for
430 obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Med*
431 Genomics 2014;7:57.
- 432 [18] Pezoulas VC, Hazapis O, Lagopati N, Exarchos TP, Goules A, Tzioufas AG, et al.
433 Machine learning approaches on high throughput NGS data to unveil mechanisms of
434 function in biology and disease. *Cancer Genom Proteom* 2021;18:605–26.
- 435 [19] Ghosh M, Sharma N, Singh AK, Gera M, Pulicherla KK, Jeong DK. Transformation of
436 animal genomics by next-generation sequencing technologies: a decade of challenges
437 and their impact on genetic architecture. *Crit Rev Biotechnol* 2018;38:1157–75.
- 438 [20] Fu Y, Liu H, Dou J, Wang Y, Liao Y, Huang X, et al. IAnimal: a cross-species omics
439 knowledgebase for animals. *Nucleic Acids Res* 2022;gkac936.
- 440 [21] Fu Y, Xu J, Tang Z, Wang L, Yin D, Fan Y, et al. A gene prioritization method based
441 on a swine multi-omics knowledgebase and a deep learning model. *Commun Biol*
442 2020;3:502.
- 443 [22] Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, et al. JBrowse: a
444 dynamic web platform for genome visualization and analysis. *Genome Biol* 2016;17:66.
- 445 [23] Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV):
446 high-performance genomics data visualization and exploration. *Briefings Bioinf*
447 2013;14:178–92.
- 448 [24] Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver:
449 A modern graphical user interface for custom BLAST databases. *Mol Biol Evol*
450 2019;36:2922–4.
- 451 [25] Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al.
452 Ensembl 2022. *Nucleic Acids Research* 2022;50:D988–95.

- 453 [26] Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to
454 function. *Nucleic Acids Res* 2019;47:D155–62.
- 455 [27] Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J.
456 eggNOG-mapper v2: Functional annotation, orthology assignments, and domain
457 prediction at the metagenomic scale. *Mol Biol Evol* 2021;38:5825–9.
- 458 [28] Hu Z-L, Park CA, Reecy JM. Bringing the Animal QTLdb and CorrDB into the future:
459 meeting new challenges and providing updated services. *Nucleic Acids Res*
460 2022;50:D956–61.
- 461 [29] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
462 *Bioinformatics* 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
- 463 [30] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and
464 genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;37:907–15.
- 465 [31] Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie
466 enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat*
467 *Biotechnol* 2015;33:290–5.
- 468 [32] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
469 *EMBnet j* 2011;17:10.
- 470 [33] Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately
471 identifies known and hundreds of novel microRNA genes in seven animal clades.
472 *Nucleic Acids Res* 2012;40:37–52.
- 473 [34] Zhang J, Chen S, Yang J, Zhao F. Accurate quantification of circular RNAs identifies
474 extensive circular isoform switching events. *Nat Commun* 2020;11:90.
- 475 [35] Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs
476 are a large class of animal RNAs with regulatory potency. *Nature* 2013;495:333–8.
- 477 [36] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
478 2012;9:357–9.
- 479 [37] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for
480 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs

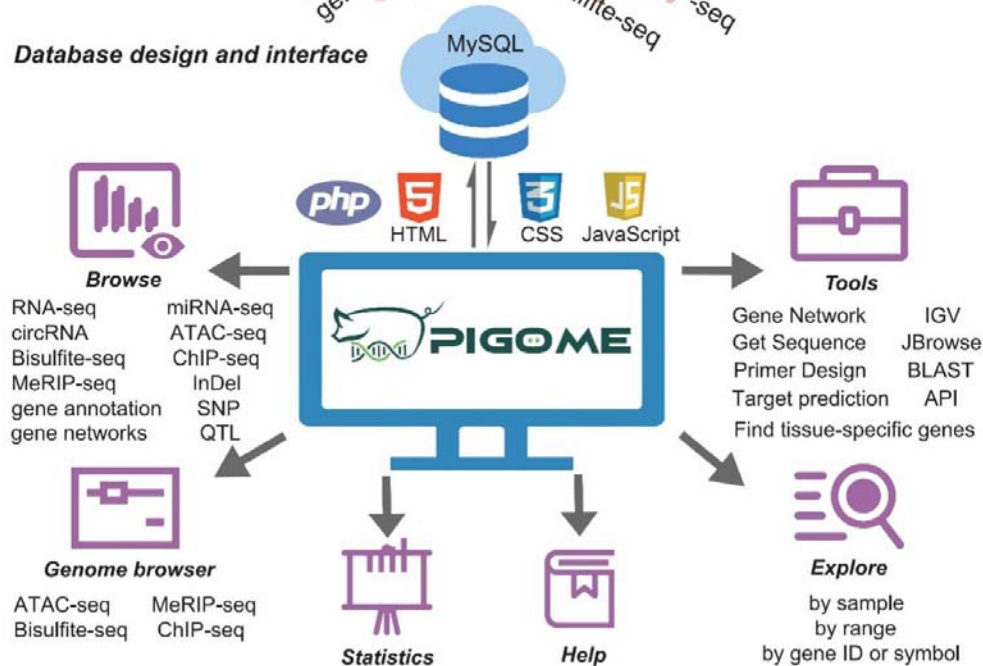
- 481 in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*
482 2012;6:80–92.
- 483 [38] Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for
484 Bisulfite-Seq applications. *Bioinformatics* 2011;27:1571–2.
- 485 [39] Meng J, Lu Z, Liu H, Zhang L, Zhang S, Chen Y, et al. A protocol for RNA
486 methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor
487 package. *Methods* 2014;69:274–81.
- 488 [40] Ramírez F, Dünder F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform
489 for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187–91.
- 490 [41] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM
491 2013.
- 492 [42] Jain A, Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis.
493 *Bioinformatics* 2019;35:1966–7.
- 494 [43] Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and
495 flexible. *Nucleic Acids Res* 2006;34:W451–4.
- 496 [44] Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in
497 *Drosophila*. *Genome Biol* 2003;5:R1.
- 498 [45] Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus,
499 an enhanced web interface to Primer3. *Nucleic Acids Res* 2007;35:W71.
- 500 [46] Davoudi P, Do DN, Colombo SM, Rathgeber B, Miar Y. Application of genetic,
501 genomic and biological pathways in improvement of swine feed efficiency. *Front Genet*
502 2022;13:903733.
- 503 [47] Tian X, Li R, Fu W, Li Y, Wang X, Li M, et al. Building a sequence map of the pig
504 pan-genome from multiple de novo assemblies and Hi-C data. *Sci China Life Sci*
505 2020;63:750–63.
- 506 [48] Kumar KR, Cowley MJ, Davis RL. Next-generation sequencing and emerging
507 technologies. *Semin Thromb Hemost* 2019;45:661–73.
- 508 [49] Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An
509 overview. *Hum Immunol* 2021;82:801–11.

510 **Figure legends**

Data retrieval



Database design and interface



511

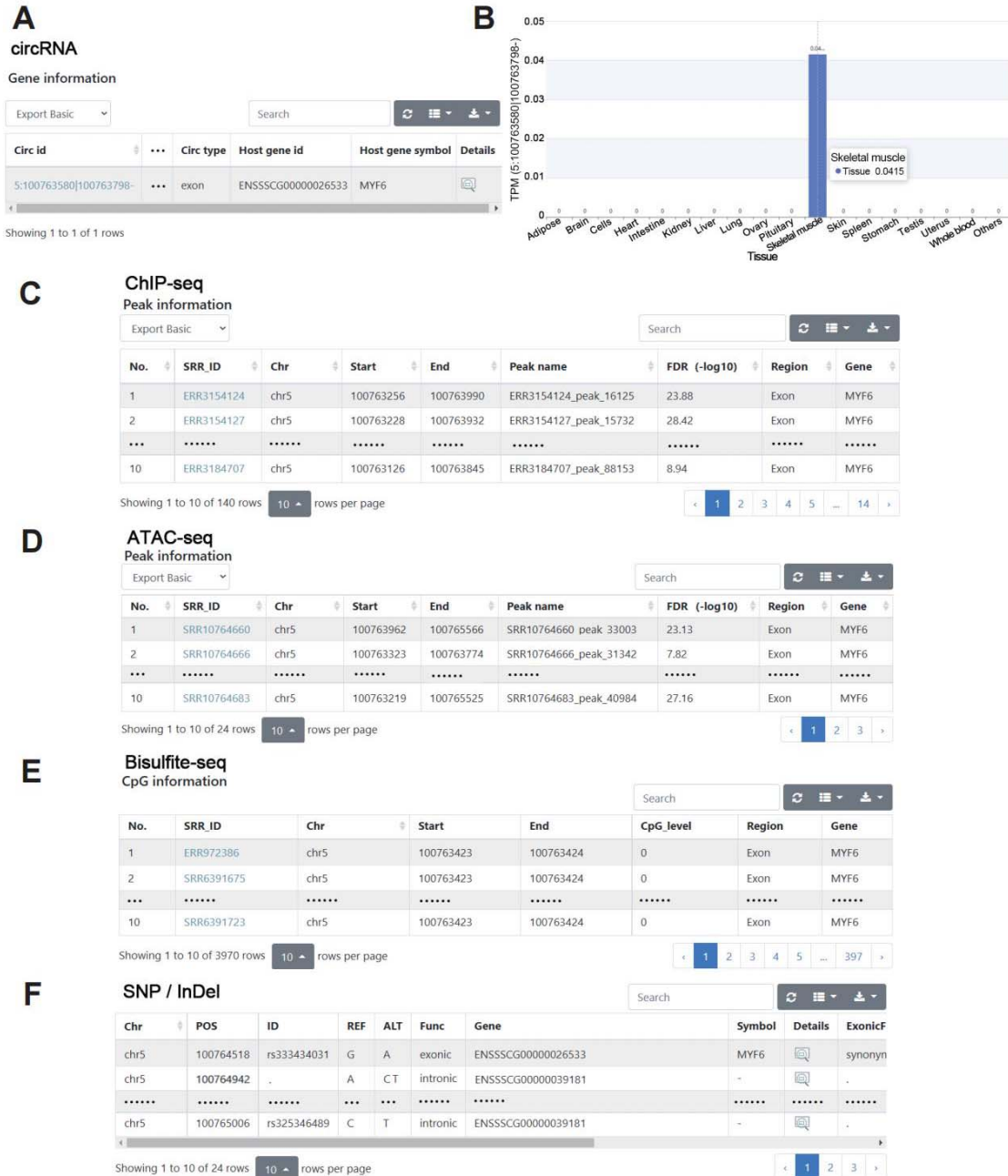
512 **Figure 1 Database contents and construction**

513 The present version of PIGOME contains 7 types of omics data, gene annotation and

514 QTL information in pigs. PIGOME also contains practical functions and analytical

515 tools to browse, explore and visualize omics data. QTL, quantitative trait locus.

516



523 **Figure 3 Explore the function and regulation of MYF6 by PIGOME**

524 A. The circRNA results are related to MYF6. B. Visualization results of circMYF6

525 expression in different tissues. C-F. The ChIP-seq (C), ATAC-seq (D), Bisulfite-seq

526 (E), SNPs (F), and InDels (F) results related to MYF6, respectively.

527

528 **Table 1 Summary of omics data in PIGOME database**

Type	Sample	Project	Tissue	Breed	Developmental stage	Data amount	Characteristic
RNA-seq	4217	268	50	74	29	33.92 Tb	31,908 genes
miRNA-seq	995	78	39	32	29	269.35 Gb	544 miRNAs
ATAC-seq	58	5	13	6	5	1.26 Tb	2,884,709 peaks
Bisulfite-seq	309	20	26	13	9	2.17 Tb	34,560,764 CpG sites
ChIP-seq	388	16	22	6	7	2.75 Tb	21,318,546 peaks
MeRIP-seq	47	5	5	5	10	174.32 Gb	424,376 peaks
WGS	887	-	-	53	-	8.67 Tb	12,074,987 SNPs

529 *Note:* WGS, whole genome sequencing.

530

531 **Table 2 Comparison of PIGOME with other databases**

	PIGOME	IAnimal-pig	Iswine	AOD
Data type	WGS, RNA-seq, miRNA-seq, ChIP-seq, ATAC-seq, BS-seq, MeRIP-seq	WGS, RNA-seq, ATAC-seq, ChIP-seq,	WGS, RNA-seq	WGS
Sample	6901	10714	4107	12
Data volume (TB)	49.21	132.92	Not provide	Not provide
Breed	113	Statistical difficulties	~23	12
Tissue	71	~130	95	Not provide
Developmental stage	29	Statistical difficulties	~80	Not provide
Analysis tools	Jbrowse2, IGV, Get Sequence, Primer Design, BLAST, Gene Network, Target Prediction, Find Tissue-specific Genes, API	Jbrowse, BLAST, Primer, Gene Network, Gene Correlation Coefficient, Signal Plotter, Signal Comparison, Genotype Plotter, Enrichment, API	Jbrowse, Primer, BLAST, Prioritize	GBrowse, BLAST, BLAT

532 *Note:* WGS, whole genome sequencing. Data collected until December 2023.