Genomics, Proteomics & Bioinformatics, 2025, **23(1)**, qzaf016
https://doi.org/10.1093/gpbjnl/qzaf016
Advance access publication: 28 February 2025
Database

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

Guohao Han (韩郭皓) [1,2,3,#], Peng Yang (杨朋) [1,2,3,4,5,#], Yongjin Zhang (张永进) [2,3], Qiaowei Li (李巧伟) [1,2,3,6], Xinhao Fan (范新浩) [1,2,3], Ruipu Chen (陈锐朴) [1,2,3], Chao Yan (闫超) [1,2,3], Mu Zeng (曾木) [1,2,3,7], Yalan Yang (杨亚岚) [1,2,3,*], Zhonglin Tang (唐中林) [1,2,3,4,*]

[1]Kunpeng Institute of Modern Agriculture at Foshan, Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences, Foshan 528225, China
[2]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China
[3]GuangXi Engineering Centre for Resource Development of Bama Xiang Pig, Bama 547500, China
[4]School of Life Sciences, Henan University, Kaifeng 475004, China
[5]Shenzhen Research Institute of Henan University, Shenzhen 518000, China
[6]School of Veterinary Medicine, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland
[7]Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding, Key Laboratory of Animal Molecular Design and Precise Breeding of Guangdong Higher Education Institutes, School of Life Science and Engineering, Foshan University, Foshan 528225, China

*Corresponding authors: tangzhonglin@caas.cn (Tang Z), yangyalan@caas.cn (Yang Y).
#Equal contribution.

Handling Editor: Yu Jiang

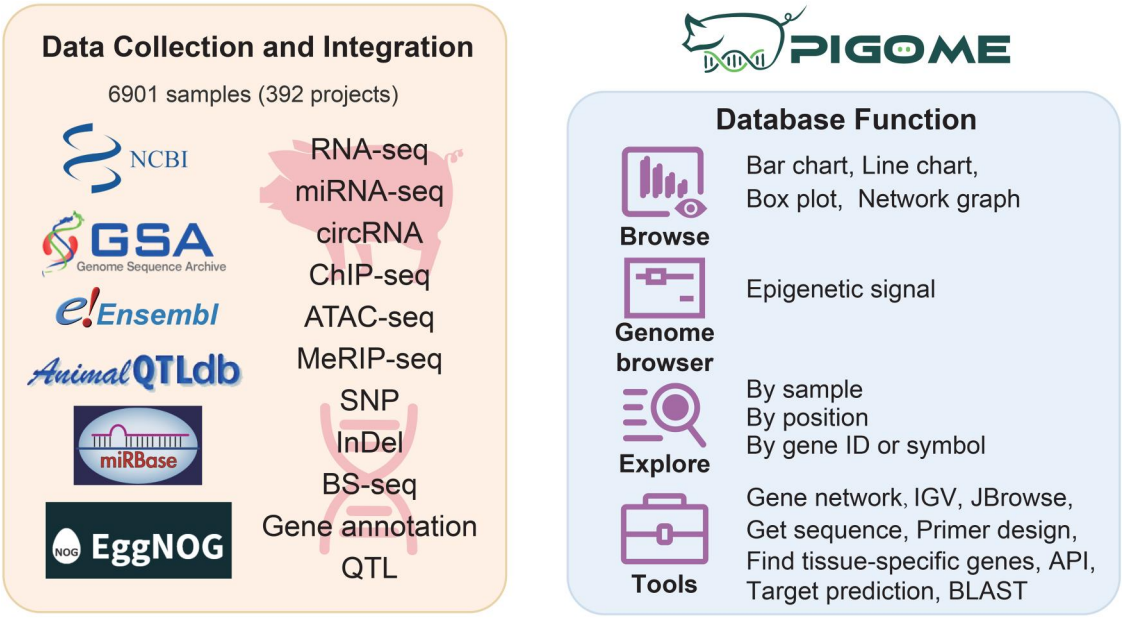
Abstract

In addition to being a major source of animal protein, pigs are an important model for studying development and diseases in humans. Over the past two decades, thousands of high-throughput sequencing studies in pigs have been performed using a variety of tissues from different breeds and developmental stages. However, multi-omics databases specifically designed for pig functional genomics research are still limited. Here, we present PIGOME, a user-friendly database of pig multi-omes. PIGOME currently contains seven types of pig omics datasets, including whole-genome sequencing (WGS), RNA sequencing (RNA-seq), microRNA sequencing (miRNA-seq), chromatin immunoprecipitation sequencing (ChIP-seq), assay for transposase-accessible chromatin sequencing (ATAC-seq), bisulfite sequencing (BS-seq), and methylated RNA immunoprecipitation sequencing (MeRIP-seq), from 6901 samples and 392 projects with manually curated metadata, integrated gene annotation, and quantitative trait locus information. Furthermore, various "Explore" and "Browse" functions have been established to provide user-friendly access to omics information. PIGOME implements several tools to visualize genomic variants, gene expression, and epigenetic signals of a given gene in the pig genome, enabling efficient exploration of spatiotemporal gene expression/epigenetic patterns, functions, regulatory mechanisms, and associated economic traits. Collectively, PIGOME provides valuable resources for pig breeding and is helpful for human biomedical research. PIGOME is available at https://pigome.com.

Key words: Pig; Multi-omics; Genome; Gene expression; Epigenetics; Database.

Received: 16 February 2023; **Revised:** 29 January 2025; **Accepted:** 25 February 2025.

Graphical abstract



Introduction

Pig production accounts for a large proportion of the animal husbandry economy and is one of the mainstays of the global agricultural economy [1,2]. Moreover, pigs have been shown to be an important biomedical model for studying human development and diseases [3–5]. Local adaptation and artificial selection have resulted in significant phenotypic differences and genetic diversity in pigs [6], providing a unique opportunity to elucidate the underlying mechanisms of key traits, such as meat production, litter size, coat color, immunity, and diseases [7,8]. Over the past two decades, with the development of advanced sequencing technologies, massive amounts of high-throughput sequencing data have been generated at multi-omics levels. These extensive datasets provide a valuable resource for understanding the genetic mechanisms underlying evolution, selection, trait formation, development, and diseases. They also reveal numerous key variants, genes, and regulatory elements that regulate various biological processes and are associated with economically significant traits [9–11]. Our recent studies, based on high-resolution DNA methylome and transcriptome analyses of skeletal muscle at 27 developmental stages, provided insights into the molecular regulation of skeletal muscle development and diversity. We identified candidate genes, such as insulin-like growth factor 2 mRNA-binding protein 3 (*IGF2BP3*) and SATB homeobox 2 (*SATB2*), that contribute to skeletal muscle development [6,12], offering representative examples of how to integrate multi-omics data to facilitate functional genomics studies in pigs. Therefore, it is necessary to integrate multi-omics data to support scientific discoveries of pig genetics and breeding.

In particular, an increasing number of high-throughput sequencing studies in pigs have been performed on a variety of tissues from different breeds and developmental stages [13–17]. However, these datasets are generated from different laboratories and sequencing platforms, making their retrieval, management, standard processing, and visualization time-consuming and difficult [18]. Furthermore, mining and integrated analysis of these datasets to explore biological functions and regulatory mechanisms remain a challenge [19]. Over the past several years, only a limited number of pig-related databases have been developed. Recently, IAnimal (https://ianimal.pro/) [20] was released, which includes pig multi-omics and genome annotation information. Similarly, ISwine (http://iswine.iomics.pro/) [21] contains published pig genomes, transcriptomes, quantitative traits, and annotation information. The Agricultural Animal Omics Database (AAOD, http://animal.nwsuaf.edu.cn/) contains pan-genome sequencing datasets [22]. However, the analysis and visualization capabilities of these databases are limited (**Table 1**). There is still a lack of comprehensive multi-omics databases dedicated to functional genomics research in pigs.

To address these challenges, we developed PIGOME, an integrated and comprehensive web database containing seven types of sequencing data from 6901 datasets and 392 projects, which is currently the most comprehensive omics database for pigs. PIGOME allows researchers to explore and utilize pig multi-omics data easily and effectively. Specifically, PIGOME supports the exploration, analysis, and visualization of genomic variations, gene expression patterns, regulatory networks, and epigenetic modifications for annotated and predicted pig genes [including protein-coding genes (PCGs), long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and circular RNAs (circRNAs)]. PIGOME also includes a tissue-specific analysis tool that allows users to identify gene characteristics in specific tissues. In addition, PIGOME deploys nine tools, such as JBrowse [23], Integrative Genomics Viewer (IGV) [24], and Basic Local Alignment Search Tool (BLAST) [25], to enable users to upload their files to visualize epigenetic signals and perform sequence alignment across the genomes of different pig breeds. In summary, PIGOME is an important database for pig

Table 1 Comparison of PIGOME with existing databases

	PIGOME	IAnimal-pig [20]	Iswine [21]	AAOD [22]
Data type	WGS, RNA-seq (mRNAs, lncRNAs, circRNAs, and AS), miRNA-seq, ChIP-seq, ATAC-seq, BS-seq, and MeRIP-seq	WGS, RNA-seq (mRNAs and lncRNAs), ATAC-seq, and ChIP-seq	WGS and RNA-seq (mRNAs and lncRNAs)	WGS
No. of samples	6901	10,714	4107	12
Data volume (Tb)	49.21	132.92	52.88	3.04
No. of breeds	113	65	23	12
No. of tissues	71	132	95	Not provide
No. of developmental stages	29	549	80	Not provide
Analysis tool	JBrowse2, IGV, Get sequence, Primer design, BLAST, Gene network, Target prediction, Find tissue-specific genes, and API	JBrowse, BLAST, Primer design, Gene network, Gene correlation coefficient, Signal plotter, Signal comparison, Genotype plotter, Enrichment, and API	JBrowse, Primer design, BLAST, and Prioritize	GBrowse, BLAST, and BLAT

Note: AAOD, Agricultural Animal Omics Database; WGS, whole-genome sequencing; RNA-seq, RNA sequencing; miRNA-seq, microRNA sequencing; ChIP-seq, chromatin immunoprecipitation sequencing; ATAC-seq, assay for transposase-accessible chromatin sequencing; BS-seq, bisulfite sequencing; MeRIP-seq, methylated RNA immunoprecipitation sequencing; mRNA, messenger RNA; lncRNA, long non-coding RNA; circRNA, circular RNA; AS, alternative splicing; IGV, Integrative Genomics Viewer; BLAST, Basic Local Alignment Search Tool; API, Application Programming Interface; BLAT, BLAST-Like Alignment Tool. Data collected until October 2024.

functional genomics studies and will be of interest to a broad readership in the fields of animal genetics, breeding, and biomedical research.

Data collection and database construction

Data collection

Seven types of high-throughput sequencing data [whole-genome sequencing (WGS), RNA sequencing (RNA-seq), microRNA sequencing (miRNA-seq), chromatin immunoprecipitation sequencing (ChIP-seq), assay for transposase-accessible chromatin sequencing (ATAC-seq), bisulfite sequencing (BS-seq) for DNA methylation analysis, and methylated RNA immunoprecipitation sequencing (MeRIP-seq)] of pigs were collected from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra/) and the China National Center for Bioinformation (CNCB) Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa/). BS-seq data contain two major types: whole-genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS) data.

WGS datasets were employed to construct the pan-genome and identify genomic variants, including single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels), within the genome. RNA-seq and miRNA-seq datasets were used to analyze the expression of PCGs and non-coding RNAs (ncRNAs), such as miRNAs, lncRNAs, and circRNAs, and to investigate alternative splicing (AS). ChIP-seq datasets were used to identify the binding sites of CCCTC-binding factor (CTCF), histone modifications, and RNA polymerase II (Pol-II) binding across the genome. ATAC-seq datasets were used to identify open chromatin regions across the genome. BS-seq and MeRIP-seq datasets were used to analyze genome-wide DNA methylation and N^6-methyladenosine (m^6A) RNA methylation, respectively.

All these datasets were manually collected with all the relevant metadata to enable fast and accurate data retrieval and statistical analysis, including project identity (ID), tissue, developmental stage, breed, read number, sequencing platform, reference, and other relevant information. The developmental stages, as described in our previous study [12], were classified into 27 defined stages [embryonic day 33 (E33) to postnatal day 180 (D180)] and two less-defined stages, "Unknown" and "Adult". Samples with poor data quality (mapping rate $< 30\%$ and data volume < 0.15 Gb) were excluded manually. To explore gene functions more conveniently, gene annotation information was integrated from Ensembl 100 [26], miRBase 22.1 [27], and eggNOG 5 [28], while quantitative trait locus (QTL) information was integrated from Animal Quantitative Trait Loci Database (Animal QTLdb 46) [29] (**Figure 1**).

Data processing

The FASTA format file of the *Sus scrofa* reference genome (build 11.1) and the gene transfer format (GTF) annotation file (release 100) were downloaded from the Ensembl database. All raw FASTQ format files were downloaded from the NCBI SRA database and the China National GeneBank (CNGB) Sequence Archive (CNSA) database. Fastp (v0.20.0) [30] was used to trim and filter the raw reads. For RNA-seq, Hierarchical Indexing for Spliced Alignment of Transcripts 2 (HISAT2, v2.0.5) [31] was used to map the reads to the reference genome. Gene expression in transcripts per kilobase of exon model per million mapped reads (TPM) was calculated using StringTie (v1.3.6) [32]. To analyze AS, each skipped exon (SE) was quantified using the percent-spliced-in (PSI) metric, which was calculated based on the proportions of long and short splice variants detected through replicate Multivariate Analysis of Transcript Splicing (rMATS, v4.0.2) [33]. The PSI matrix utilized in the current database was adopted from our recent study [34]. For miRNA-seq, adapters were removed using Cutadapt (v1.8.dev0) [35]. After adapter removal, reads were aligned to annotated pig miRNAs gathered from miRBase [27] and the pig reference genome using miRDeep2 (v0.1.2) [36]. For circRNAs, HISAT2 (v2.0.5) was used for alignment with the reference genome. Novel circRNAs were identified using CIRIquant (v1.1.1) [37] and find_circ (v2) [38]. For ATAC-seq and ChIP-seq, all reads were mapped using Bowtie 2 (v2.3.5.1) [39]. The peaks were identified using Model-based Analysis of ChIP-Seq 2 (MACS2, v2.2.6) and annotated using SnpEff (v4.2) [40]. Bismark (v0.23.0) [41] was used to align the

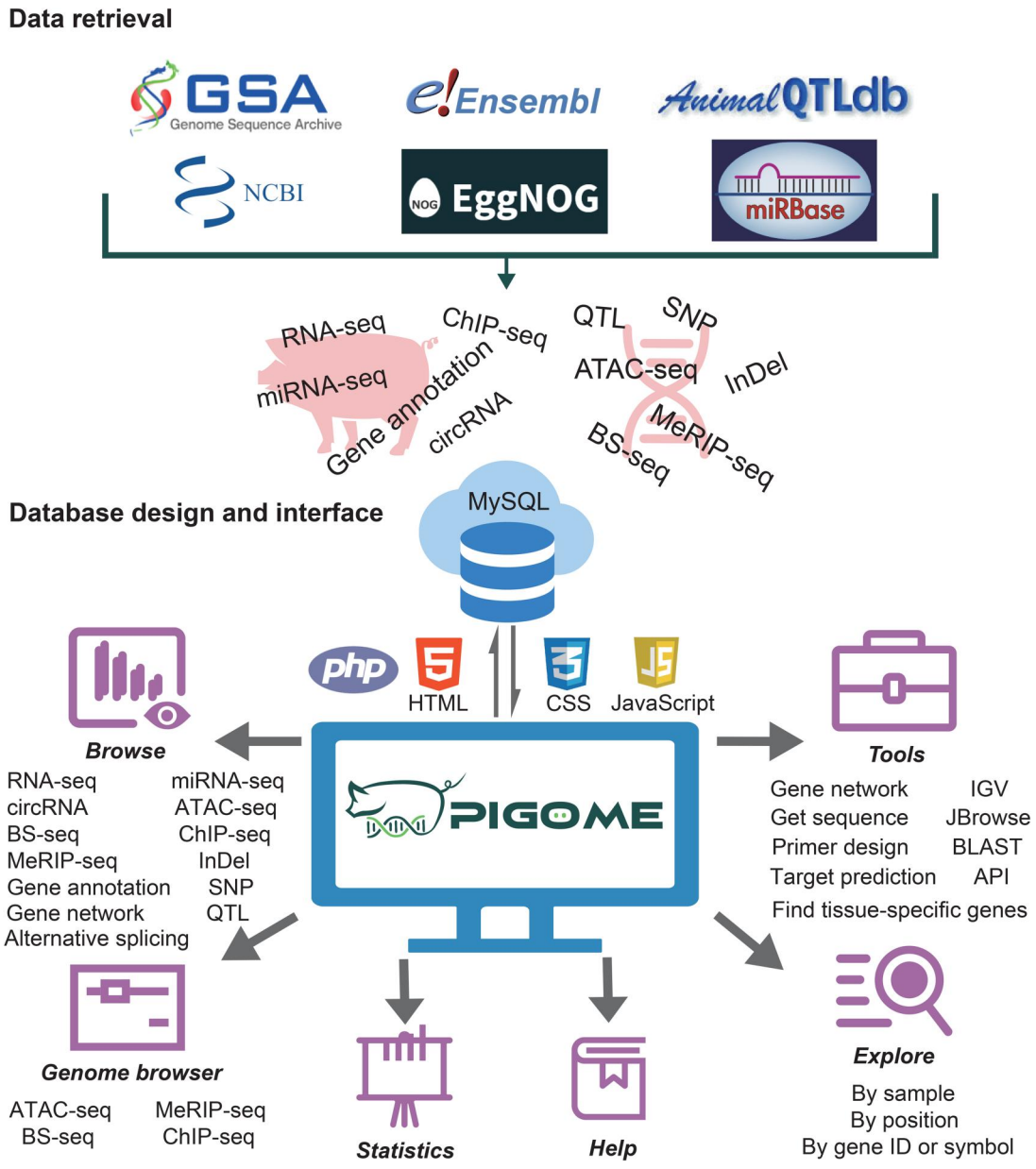


Figure 1 Database content and construction

The current version of PIGOME integrates seven types of omics data, along with gene annotation, QTL, and alternative splicing information for pigs. PIGOME also provides practical functions and analytical tools for browsing, exploring, and visualizing omics data. NCBI, National Center for Biotechnology Information; GSA, Genome Sequence Archive; Animal QTLdb, Animal QTL Database; QTL, quantitative trait locus; RNA-seq, RNA sequencing; miRNA-seq, microRNA sequencing; ChIP-seq, chromatin immunoprecipitation sequencing; ATAC-seq, assay for transposase-accessible chromatin sequencing; BS-seq, bisulfite sequencing; MeRIP-seq, methylated RNA immunoprecipitation sequencing; circRNA, circular RNA; SNP, single nucleotide polymorphism; InDel, insertion and deletion; HTML, HyperText Markup Language; CSS, Cascading Style Sheets; IGV, Integrative Genomics Viewer; BLAST, Basic Local Alignment Search Tool; API, Application Programming Interface.

BS-seq reads to the reference genome using default parameters. All CpG sites were identified and annotated as previously described [12]. For MeRIP-seq, all reads were mapped using HISAT2 (v2.0.5), and peaks were identified using exomePeak2 (v2) [42] and annotated using SnpEff (v4.2). Picard (v2.25.7) was used to remove duplicate polymerase chain reaction (PCR) reads for ATAC-seq, ChIP-seq, and MeRIP-seq. The BigWig files for genome browser visualization were generated using deepTools (v3.5.1) [43] and bedGraphToBigWig (v4). For WGS, all reads were aligned to the reference genome using Burrows-Wheeler Aligner (BWA, v0.7.12) [44]. SNP and InDel calling was performed using the UnifiedGenotyper approach implemented in the Genome Analysis Toolkit (GATK, v4.1.5.0). To achieve high accuracy in variant calling, SNPs and InDels were filtered using the following parameters: QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, or ReadPosRankSum < -8.0. Considering that the volume of SNP data is too large ($\sim$ 148 GB), SNPs in intergenic regions were not shown on the website, but the raw data could be downloaded from the PIGOME database. Furthermore, tissue-specific genes were identified using the R package TissueEnrich (v3.15) [45] based on the expression matrix of PCGs and ncRNAs. The "rcorr" function of the R Hmisc (v5.1-1) package was used to calculate

expression correlations between PCGs, miRNAs, and circRNAs to construct co-expression networks with $r > 0.85$ and $P < 0.01$ as thresholds. The intersection results of RNAhybrid (v2.1.2) [46] and miRanda (v3.3a) [47] were used to predict putative targets for mRNAs and circRNAs with E value < -20.

Website implementation

PIGOME was built using ThinkPHP (v6.0.12, https://www.thinkphp.cn/), a mature Model–View–Controller (MVC) framework, deployed in CentOS (v7.9) system. All omics data were stored in MySQL (v5.6.50, https://www.mysql.com/). The web interfaces were developed using HyperText Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript, and Bootstrap (v5.0.2, https://getbootstrap.com/). Most of the interactive charts and tables were implemented with ECharts (v5.3.1, https://echarts.apache.org/) and Bootstrap Table (v1.14.2, https://bootstrap-table.com/) (Figure 1). Network proxy services were provided through NGINX (v1.20.1, https://www.nginx.com/). We recommend visiting PIGOME using Google Chrome, Microsoft Edge, or Mozilla Firefox.

Database content and usage

Data collection and statistics

Currently, PIGOME v1.0 collects 7 types of multi-omics datasets in pigs, including WGS, RNA-seq, miRNA-seq, ChIP-seq, ATAC-seq, BS-seq (WGBS and RRBS), and MeRIP-seq data. It contains 6901 samples from 392 projects, covering 113 breeds, 71 tissues, and 29 developmental stages. The total clean data volume reaches 49.21 Tb (Table 1 and **Table 2**). The RNA-seq datasets are the most abundant in our database, including 4217 samples, 74 breeds, 50 tissues, and 29 developmental stages (Table 2). To better interpret the omics data, we integrated 32,452 gene annotations, 16,932 SE events, and 29,687 QTLs. Gene annotation records commonly have 22 attributes, including gene symbol, gene type, description, muscle biology, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Carbohydrate-Active enZYmes Database (CAZy), and Pfam. The AS data include SE events across multiple tissues and skeletal muscle development stages, providing valuable insights into tissue-specific and temporally regulated AS. In addition, the QTL information collects 11 attributes, mainly including position, QTL ID, name, type, trait, and PubMed ID. Additional statistics are summarized on the statistics page (https://pigome.com/statistics.html).

PIGOME features and functions

PIGOME includes genomic (SNPs, InDels, and genome annotation), epigenomic (chromatin accessibility, histone modifications, and DNA/RNA methylation), and transcriptomic (mRNAs, ncRNAs, and AS) data. In addition, it provides useful and user-friendly tools to help users perform advanced analyzes (Figure 1).

Browse

Users can easily browse omics data using a "Browse" tag in the toolbar. After clicking on the omics data type, summary information related to the data will be displayed. On the summary page of each level of omics data, users can obtain specific statistical data, including sample, gene, and other related information, and freely download these tables and charts. For more details, users can click the icon in the "Details" column of a given gene or sample information table on the page, which links to the gene expression page. Basic information about the gene or sample is displayed at the top of the gene expression page, with links to external databases. Different gene expression pages contain different sections. Specifically, the gene expression pages for RNA-seq, miRNA-seq, and circRNA data can show the TPM values in various breeds or developmental stages in given tissues, and also display the TPM values of subgroups of samples freely selected by users. In addition, to better understand the gene function, PIGOME integrates a variety of gene annotations. To better find the co-regulation between genes, the page shows the gene network of the queried gene. The expression pattern of a given gene, as provided by users, can be visualized through box plots, bar charts, and line charts. The data will be conveniently presented in a table below the graph. The AS page enables users to explore the dynamics of SEs across various tissues in Luchuan and Duroc pigs, as well as across 27 developmental stages of the skeletal muscle in Tongcheng pigs. Furthermore, the detail pages for ATAC-seq, BS-seq, ChIP-seq, and MeRIP-seq data provide an IGV genome browser and a table to display information, allowing users to freely explore any genome intervals of each sample. Additionally, users can browse the allele frequency of certain loci in different breeds on the detail page for SNPs and InDels using a bar chart and table. This page also provides QTL information related to this region. In summary, PIGOME has a variety of browsing functions, paving the way for the integration and investigation of different omics data in pigs.

Explore

For convenient usage, PIGOME provides three search engines to explore the entire database, including "by gene ID or symbol", "by position", and "by sample". For "by gene ID or

Table 2 Summary of omics data in PIGOME database

Data type	No. of samples	No. of projects	No. of tissues	No. of breeds	No. of developmental stages	Data volume	Data content
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 accessible chromatin regions
BS-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 genomic regions
MeRIP-seq	47	5	5	5	10	174.32 Gb	424,376 RNA methylation sites
WGS	887	–	–	53	–	8.67 Tb	12,074,987 SNPs and 5,349,818 InDels

Note: miRNA, microRNA; SNP, single nucleotide polymorphism; InDel, insertion and deletion.

symbol", users can search by entering Ensembl gene ID or official gene symbol. On the "by position" page, users can search by selecting a chromosome and entering the start and end positions. On the results pages for both "by gene ID or symbol" and "by position", all omics datasets related to a given gene or genomic region are integrated and displayed. Additionally, users can obtain more detailed information by clicking links in the table. For "by sample", users can fuzzily search by selecting the dataset and entering Sequence Read Archive Run (SRR) ID, sample ID, or project ID. Furthermore, the results page for "by sample" displays relevant sample information with associated links. Importantly,

Figure 2 Find tissue-specific genes module in PIGOME

A. The results of finding tissue-specific genes in skeletal muscle. **B**. Basic and annotation information related to ENSSSCG00000026533 (*MYF6*). **C**. The expression of *MYF6* in different tissues. **D**. Expression trend of *MYF6* in skeletal muscle at different developmental stages. TF, transcription factor; COG, cluster of orthologous groups; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, KEGG orthology; CAZy, Carbohydrate-Active enZYmes Database; TPM, transcripts per kilobase of exon model per million mapped reads; E, embryonic day; D, postnatal day.

all figures and data from the search results can be downloaded freely and edited easily.

Genome browser

PIGOME integrates a custom genome browser based on JBrowse2 to help users compare and analyze various omics datasets. PIGOME contains gene annotation information from Ensembl. By inputting a genome range, gene ID, or gene symbol, users can explore the omics data related to the gene of interest. All tracks are labeled according to the type of omics data, tissues, breeds, and developmental stages. In the track group, users can display tracks of interest by toggling the checkboxes.

Tools

We have integrated nine practical tools, including IGV, JBrowse, "Get sequence", "Primer design", BLAST, "Gene network", "Target prediction", "Find tissue-specific genes", and Application Programming Interface (API) (Table 1). For IGV and JBrowse, users can check, verify, and interpret their own sequencing and genome data online. The "Get sequence" tool allows users to quickly extract the required gene sequence from a large number of nucleotide sequences. Then, "Primer design" tool [48] can help users design primers from DNA/RNA sequences of interest for further experimental verification. With BLAST, users can perform an alignment analysis using their own sequences against 23 pig reference genomes. The "Gene network" tool is useful for identifying gene regulatory networks formed by the interactions between genes. For "Target prediction" tool, users can explore the regulatory role of a given miRNA in gene expression and find potential functional miRNAs associated with economic traits. Furthermore, a tool called "Find tissue-specific genes" was developed, which helps users quickly find tissue-specific genes, miRNAs, and circRNAs based on our massive RNA-seq and miRNA-seq data. Additionally, for the API tool, users with basic programming skills can obtain the omics data more flexibly and explore functional genes more effectively. These tools will assist users to better explore the biological mechanisms of various biological processes and important economic traits in pigs.

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

Comparison with existing databases

To date, several user-friendly databases have been established to aggregate multi-omics datasets in pigs (Table 1). IAnimal [20] is a multi-species and multi-omics database, encompassing four types of pig omics data, including WGS, RNA-seq (mRNAs and lncRNAs), ChIP-seq, and ATAC-seq data. ISwine [21] serves as a professional pig omics database, offering access to WGS data, RNA-seq (mRNAs and lncRNAs) data, quantitative traits, and annotation information. While IAnimal and ISwine both provide valuable sample meta-information, including details on tissue, developmental stage, and breed, they lack secondary classification and correction of this metadata. Consequently, comparative analysis of gene expression regulation between developmental stages or breeds becomes challenging. The AAOD database [22] focuses on providing 12 *de novo* genome assemblies in pigs. PIGOME, however, emerges as a standout platform in this landscape. Notably, it boasts the widest array of omics data types, including WGS (SNPs and InDels), RNA-seq (mRNAs, lncRNAs, circRNAs, and AS), miRNA-seq, ChIP-seq, ATAC-seq, BS-seq, and MeRIP-seq data (Table 2). Specifically, PIGOME uniquely provides a leading resource for RNA regulation research, offering data on over 500 miRNAs, 150,000 circRNAs, and 16,000 SE events. Furthermore, PIGOME provides meticulously curated and well-organized meta-information on sample data. It supports data from 113 breeds, significantly surpassing other databases and enabling more comprehensive analyses of breed-specific traits and gene regulation. The sample details on developmental stages and tissue organization have been rigorously corrected and refined, ensuring high accuracy. This meticulous curation offers invaluable insights for users aiming to investigate traits with greater precision across diverse breeds, developmental stages, and tissues. Additionally, PIGOME offers nine practical tools designed to enhance the utilization of multi-omics datasets, a feature comparable to that of IAnimal (Table 1). Notably, PIGOME includes useful tools such as "Target prediction", "Find tissue-specific genes", "Get sequence", and "IGV", which are unavailable in other databases. PIGOME facilitates the functional genomics exploration of tissue-specific PCGs and ncRNAs in a more convenient manner, as shown in the next section.

Case study

Herein, we provide a case study to verify the usefulness of the "Find tissue-specific genes" tool in PIGOME and illustrate how to use PIGOME to mine multi-omics information for genes of interest. Initially, users can select the option "skeletal muscle" in the "Find tissue-specifically expressed genes" section of the tool and then click the "Explore" button. On the results page, based on substantial expression data, users can find 186 high-confidence genes that are specifically expressed in skeletal muscle, and can then click the view icon of "ENSSSCG00000026533" to explore more detailed expression information (**Figure 2A**). On the expression page, users can obtain the gene symbol of myogenic factor 6 (*MYF6*), also known as *MRF4*, which encodes a myogenic regulatory factor involved in myogenesis. In addition, users can first view the related gene annotation and visualize its expression in different tissues using bar charts, line charts, or box plots (Figure 2B and C). Importantly, users can also explore the expression trend of this gene in skeletal muscles of different breeds and at different developmental stages (Figure 2D), demonstrating the ability of PIGOME to explore potential tissue-specific genes.

Finally, users can obtain all omics information related to genes using the exploration function. Users can use "ENSSSCG00000026533" or "*MYF6*" as the input in "Explore by gene ID or symbol". The results page provides information, including SNP and InDel variations, gene expression abundance, AS, annotation, epigenetic modifications, and QTLs related to *MYF6*. More importantly, a circRNA derived from the *MYF6* locus was identified (**Figure 3A**). Interestingly, by clicking the view icon (Figure 3B), data revealed that this circRNA (*circ-MYF6*), which is specifically expressed in skeletal muscle, may be a candidate circRNA that affects the development and growth of skeletal muscle. A total of 140 peaks were identified from the ChIP-seq data, along with ATAC-seq-detected open chromatin regions in the *MYF6* gene (Figure 3C and D). Furthermore, the results showed 3970 CpG methylation sites (Figure 3E) as well as 24 SNPs and InDels (Figure 3F) distributed across the exons, introns, and upstream regions of *MYF6*. These results indicate that PIGOME can be used to explore the potential regulatory mechanisms of candidate genes.

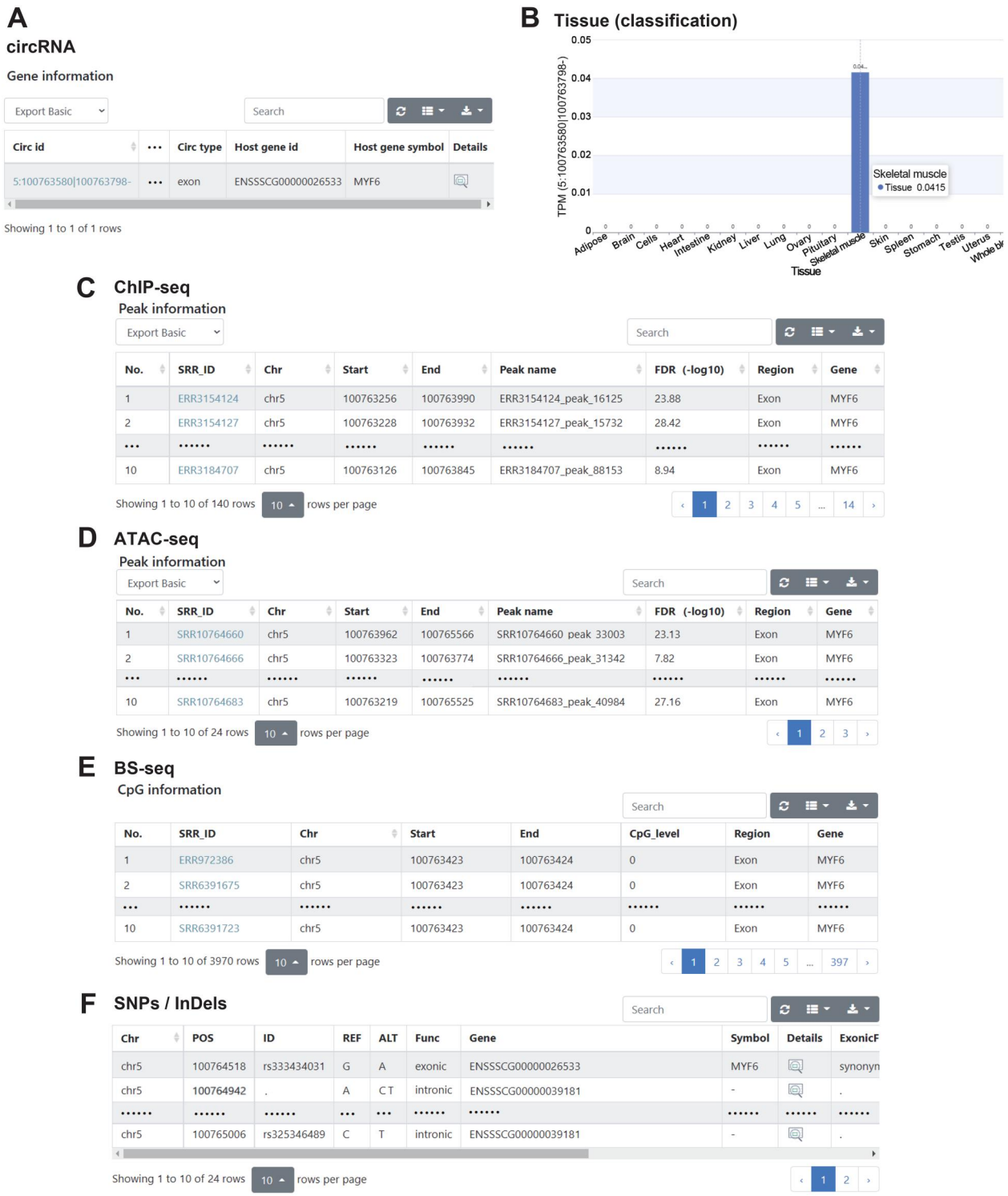


Figure 3 Exploration of the function and regulation of *MYF6* by PIGOME

A. The circRNA results related to *MYF6*. **B**. Visualization results of *circ-MYF6* expression across different tissues. **C**.–**F**. The ChIP-seq (C), ATAC-seq (D), BS-seq (E), and SNPs/InDels (F) results related to *MYF6*, respectively. SRR, Sequence Read Archive Run; FDR, false discovery rate.

Discussion

Over the past few decades, researchers have made great efforts in functional genomics research of pigs and have accumulated valuable omics data [49]. Compared to earlier released databases for pigs, such as IAnimal [20], ISwine [21], and AAOD [22], PIGOME contains the most data types and the most up-to-date multi-omics datasets covering comprehensive meta-information (Table 1). Moreover, PIGOME provides a user-friendly interface for browsing and analyzing omics data via interactive webpages, powerful search engines, and advanced tools. The integrated genomic, transcriptomic, and epigenomic data provide an efficient approach for discovering target genes and loci associated with economic traits and human-related diseases.

With the continuous development and innovation of high-throughput sequencing methods, more technologies have been developed, such as single-cell RNA-seq (scRNA-seq), spatial transcriptomics, and 3D genome [50,51]. The amount of omics data in public databases is also increasing rapidly. PIGOME will continue to update new omics types and expand its data volume. In the near future, other types of variations in the pig genome, such as structure variations (SVs), copy number variations (CNVs), and presence/absence variations (PAVs) will be incorporated into PIGOME. We aim to focus on cutting-edge single-cell sequencing and enhance the display of related data, such as scRNA-seq, single-cell ATAC-seq (scATAC-seq), and spatial transcriptomics data. In addition, PIGOME will update the latest gene annotation, genome-wide association studies (GWAS), epigenome-wide association studies (EWAS), and transcriptome-wide association studies (TWAS), and incorporate multi-types of QTL information, such as expression quantitative trait loci (eQTLs) and splicing quantitative trait loci (sQTLs), to help users better understand gene functions. Furthermore, we will strengthen the connections among various data in the database and develop more comprehensive online tools. Finally, we aim to establish PIGOME as a key resource for exploring pig functional genomics, which we believe will be of great value to the broad scientific community in the fields of animal genetics, breeding, and biomedical research.

Data availability

PIGOME is available at https://pigome.com. It has also been submitted to Database Commons [52] at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB), which is publicly accessible at https://ngdc.cncb.ac.cn/databasecommons/database/id/9718.

CRediT author statement

Guohao Han: Methodology, Software, Visualization, Writing – original draft. **Peng Yang:** Methodology, Software, Data curation, Formal analysis. **Yongjin Zhang:** Data curation, Formal analysis. **Qiaowei Li:** Formal analysis. **Xinhao Fan:** Formal analysis. **Ruipu Chen:** Formal analysis. **Chao Yan:** Formal analysis. **Mu Zeng:** Formal analysis. **Yalan Yang:** Conceptualization, Project administration, Funding acquisition, Writing – review & editing. **Zhonglin Tang:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

We gratefully acknowledge all researchers who generated the omics data used in this study. This work was supported by the National Key Scientific Research Project (Grant No. 2023YFF1001100), the Shenzhen Innovation and Entrepreneurship Plan — a major special project of science and technology (Grant No. KJZD20230923115003006), the National Natural Science Foundation of China (Grant Nos. U23A20229 and 32172697), and the Agricultural Science and Technology Innovation Program (Grant Nos. CAAS-ZDRW202406 and CAAS-CSAB-202402), China.

ORCID

0000-0001-8648-7347 (Guohao Han)
0000-0002-0941-669X (Peng Yang)
0009-0003-3962-7505 (Yongjin Zhang)
0000-0002-1878-7863 (Qiaowei Li)
0009-0004-2208-0133 (Xinhao Fan)
0009-0005-7501-580X (Ruipu Chen)
0000-0003-3066-8258 (Chao Yan)
0009-0003-2375-8666 (Mu Zeng)
0000-0003-4912-1869 (Yalan Yang)
0000-0002-4538-4349 (Zhonglin Tang)

References

[1] Zhang S, Wu X, Han D, Hou Y, Tan J, Kim SW, et al. Pork production systems in China: a review of their development, challenges and prospects in green production. Front Agr Sci Eng 2021;8:15.
[2] Wu Y, Zhao J, Xu C, Ma N, He T, Zhao J, et al. Progress towards pig nutrition in the last 27 years. J Sci Food Agric 2020;100:5102–10.
[3] Lunney JK, Van Goor A, Walker KE, Hailstock T, Franklin J, Dai C. Importance of the pig as a human biomedical model. Sci Transl Med 2021;13:eabd5758.
[4] Tu CF, Chuang CK, Yang TS. The application of new breeding technology based on gene editing in pig industry – a review. Anim Biosci 2022;35:791–803.
[5] Yang H, Wu Z. Genome editing of pigs for agriculture and biomedicine. Front Genet 2018;9:360.
[6] Yang Y, Yan J, Fan X, Chen J, Wang Z, Liu X, et al. The genome variation and developmental transcriptome maps reveal genetic differentiation of skeletal muscle in pigs. PLoS Genet 2021;17:e1009910.
[7] Kyselova J, Tichy L, Jochova K. The role of molecular genetics in animal breeding: a minireview. Czech J Anim Sci 2021;66:107–11.
[8] Kasper C, Ribeiro D, de Almeida AM, Larzul C, Liaubet L, Murani E. Omics application in animal science–a special emphasis on stress response and damaging behaviour in pigs. Genes (Basel) 2020;11:920.
[9] Long JA. The "omits" revolution: use of genomic, transcriptomic, proteomic and metabolomic tools to predict male reproductive traits that impact fertility in livestock and poultry. Anim Reprod Sci 2020;220:106354.
[10] Yang Y, Zhou R, Li K. Future livestock breeding: precision breeding based on multi-omics information and population personalization. J Integr Agric 2017;16:2784–91.
[11] Koltes JE, Cole JB, Clemmens R, Dilger RN, Kramer LM, Lunney JK, et al. A vision for development and utilization of high-throughput phenotyping and big data analytics in livestock. Front Genet 2019;10:1197.
[12] Yang Y, Fan X, Yan J, Chen M, Zhu M, Tang Y, et al. A comprehensive epigenome atlas reveals DNA methylation regulating skeletal muscle development. Nucleic Acids Res 2021;49:1313–29.
[13] Kim JM, Park JE, Yoo I, Han J, Kim N, Lim WJ, et al. Integrated transcriptomes throughout swine oestrous cycle reveal dynamic changes in reproductive tissues interacting networks. Sci Rep 2018;8:5436.
[14] Jin L, Tang Q, Hu S, Chen Z, Zhou X, Zeng B, et al. A pig BodyMap transcriptome reveals diverse tissue physiologies and

evolutionary dynamics of transcription. Nat Commun 2021; 12:3715.
[15] Liu Y, Liu Y, Ma T, Long H, Niu L, Zhang X, et al. A splicing mutation in *PHKG1* decreased its expression in skeletal muscle and caused PSE meat in Duroc × Luchuan crossbred pigs. Anim Genet 2019;50:395–8.
[16] Zhao Y, Hou Y, Xu Y, Luan Y, Zhou H, Qi X, et al. A compendium and comparative epigenomics analysis of *cis*-regulatory elements in the pig genome. Nat Commun 2021;12:2217.
[17] Kogelman LJA, Cirera S, Zhernakova DV, Fredholm M, Franke L, Kadarmideen HN. Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Med Genomics 2014;7:57.
[18] Pezoulas VC, Hazapis O, Lagopati N, Exarchos TP, Goules A, Tzioufas AG, et al. Machine learning approaches on high throughput NGS data to unveil mechanisms of function in biology and disease. Cancer Genomics Proteomics 2021; 18:605–26.
[19] Ghosh M, Sharma N, Singh AK, Gera M, Pulicherla KK, Jeong DK. Transformation of animal genomics by next-generation sequencing technologies: a decade of challenges and their impact on genetic architecture. Crit Rev Biotechnol 2018; 38:1157–75.
[20] Fu Y, Liu H, Dou J, Wang Y, Liao Y, Huang X, et al. IAnimal: a cross-species omics knowledgebase for animals. Nucleic Acids Res 2023;51:D1312–24.
[21] Fu Y, Xu J, Tang Z, Wang L, Yin D, Fan Y, et al. A gene prioritization method based on a swine multi-omics knowledgebase and a deep learning model. Commun Biol 2020;3:502.
[22] Tian X, Li R, Fu W, Li Y, Wang X, Li M, et al. Building a sequence map of the pig pan-genome from multiple *de novo* assemblies and Hi-C data. Sci China Life Sci 2020;63:750–63.
[23] Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, et al. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 2016;17:66.
[24] Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178–92.
[25] Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: a modern graphical user interface for custom BLAST databases. Mol Biol Evol 2019;36:2922–4.
[26] Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res 2022; 50:D988–95.
[27] Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res 2019; 47:D155–62.
[28] Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 2021;38:5825–9.
[29] Hu ZL, Park CA, Reecy JM. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res 2022;50:D956–61.
[30] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.
[31] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37:907–15.
[32] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33:290–5.
[33] Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. Proc Natl Acad Sci U S A 2014; 111:E5593–601.
[34] Wang W, Fan X, Liu W, Huang Y, Zhao S, Yang Y, et al. The spatial-temporal alternative splicing profile reveals the functional diversity of FXR1 isoforms in myogenesis. Adv Sci (Weinh) 2024; 11:e2405157.
[35] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 2011;17:10.
[36] Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 2012; 40:37–52.
[37] Zhang J, Chen S, Yang J, Zhao F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. Nat Commun 2020;11:90.
[38] Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 2013;495:333–8.
[39] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9.
[40] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso*-2; *iso*-3. Fly (Austin) 2012;6:80–92.
[41] Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 2011; 27:1571–2.
[42] Meng J, Lu Z, Liu H, Zhang L, Zhang S, Chen Y, et al. A protocol for RNA methylation differential analysis with MeRIP-seq data and exomePeak R/Bioconductor package. Methods 2014; 69:274–81.
[43] Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187–91.
[44] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013;1303.3997.
[45] Jain A, Tuteja G. TissueEnrich: tissue-specific gene enrichment analysis. Bioinformatics 2019;35:1966–7.
[46] Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res 2006;34:W451–4.
[47] Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in *Drosophila*. Genome Biol 2003;5:R1.
[48] Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 2007;35:W71–4.
[49] Davoudi P, Do DN, Colombo SM, Rathgeber B, Miar Y. Application of genetic, genomic and biological pathways in improvement of swine feed efficiency. Front Genet 2022;13:903733.
[50] Kumar KR, Cowley MJ, Davis RL. Next-generation sequencing and emerging technologies. Semin Thromb Hemost 2019;45:661–73.
[51] Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: an overview. Hum Immunol 2021;82:801–11.
[52] Ma L, Zou D, Liu L, Shireen H, Abbasi AA, Bateman A, et al. Database Commons: a catalog of worldwide biological databases. Genomics Proteomics Bioinformatics 2023;21:1054–8.

Genomics, Proteomics & Bioinformatics, 2025, 23, 1–10
https://doi.org/10.1093/gpbjnl/qzaf016
Database