



Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

AASIA: A comprehensive protein structural interactome database for agricultural animals

Jiajun Li^{a,1}, Mengdi Yuan^{a,1}, Linyang Jiang^a, Dianke Li^a, Zhongtao Yin^b, Feng Zhu^b, Wenyu Shi^{a,*}, Zhuocheng Hou^{b,*}, Ziding Zhang^{a,*}

^aState Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing 100193, China

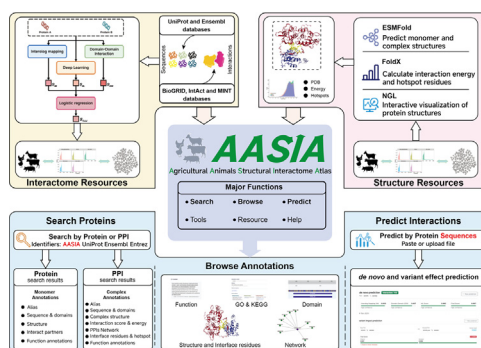
^bNational Engineering Laboratory for Animal Breeding and Key Laboratory of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

HIGHLIGHTS

- AASIA is the first proteome-scale, structure-informed interactome resource for five agricultural animals.
- AASIA constructs interactomes by integrating interolog mapping, domain-domain interaction inference, and deep learning.
- AASIA annotates each interaction with 3D complex structures from ESMFold, interaction sites, and hotspot residues.
- AASIA provides a user-friendly platform with Search, Browse, Predict, and Tools modules for interactome exploration.
- AASIA facilitates deciphering the genome–phenome relationships and structure-aware molecular breeding.

GRAPHICAL ABSTRACT

Overview of AASIA.



ARTICLE INFO

Article history:

Received 22 September 2025

Revised 12 September 2026

Accepted 18 September 2026

Available online xxxxx

Keywords:

Agricultural animals

Database

Deep learning

Protein–protein interactions

Prediction

Structural interactome

ABSTRACT

Introduction: Agricultural animals are essential for global food production and sustainable agriculture. Improving disease resistance and economically important traits is critical for ensuring food security. Proteome-wide protein–protein interactions (PPIs), termed as interactomes, underlie key biological processes in cells, making them critical for deciphering the genome–phenome relationship. However, species-specific PPI resources remain limited for most agricultural animals.

Objectives: Motivated by the rapid development of deep learning and breakthroughs in AI-driven protein structure prediction, we attempted to develop an integrated structural interactome resource for agricultural animals.

Methods: We predicted species-specific interactomes through an integrative prediction pipeline that combines interolog mapping, domain–domain interaction inference, and deep learning. The 3D complex structures of all predicted PPIs were further generated using ESMFold.

Results: We established AASIA (Agricultural Animals Structural Interactome Atlas; <https://aasia.zzd-lab.com>), a user-friendly database comprising 410,421 high-confidence PPIs and corresponding complex

* Corresponding authors.

E-mail addresses: shiwuyang@cau.edu.cn (W. Shi), zchou@cau.edu.cn (Z. Hou), zidingzhang@cau.edu.cn (Z. Zhang).

¹ Jiajun Li and Mengdi Yuan contributed equally to this work.

<https://doi.org/10.1016/j.jare.2026.09.015>

2090-1232/© 2026 Published by Elsevier B.V. on behalf of Cairo University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

structures across five key species, including *Anas platyrhynchos*, *Bos taurus*, *Gallus gallus*, *Ovis aries*, and *Sus scrofa*.

Conclusions: AASIA provides the first proteome-scale structural interactome resource for five key agricultural animals. By integrating interaction prediction with structural modeling, it enables mechanistic insight into agriculturally relevant traits and supports applications in variant interpretation, multi-omics integration, and molecular breeding.

© 2026 Published by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Agricultural animals, such as chickens, ducks, pigs, cattle, and sheep, play a critical role in global food production systems [1–4]. As the demand for animal-derived products continues to rise, improving economically important traits, including disease resistance, growth rate, fertility, and product quality, has become increasingly critical for ensuring food security and sustainable farming development [5–7]. Beyond their agricultural value, these species also provide indispensable systems for studying high-impact infectious diseases, exemplified by Avian influenza and African swine fever, underscoring their broader significance to public and animal health [4,8–11].

Understanding the molecular basis of such complex traits requires a systems-level view of protein–protein interactions (PPIs). Proteins rarely act in isolation but operate within intricate interaction networks that govern biological processes ranging from immune response to reproduction. Mapping interactomes thus provides essential insights into regulatory pathways and facilitates the identification of key molecular targets for genetic improvement and disease intervention [4,10,12,13]. While reference organisms such as *Homo sapiens* [14,15], *Mus musculus* [16], and *Saccharomyces cerevisiae* [17] benefit from well-established interactome resources, agricultural animals remain underrepresented in existing interaction resources. Databases such as STRING [18], BioGRID [19], IntAct [20], and MINT [21] provide limited interaction data for these species and typically lack structural annotations, hindering fine-grained functional analysis. Incorporating structural information into interactome construction enables deeper insights into molecular interactions. Structural models allow for residue-level mapping of binding interfaces, identification of functional hotspots, and prediction of mutational effects, which are critical for linking genotype to phenotype and informing molecular breeding strategies [13,22]. Yet, structure-informed interactomes remain rare, especially for agricultural animals.

Recent advances in high-throughput PPI discovery have accelerated interactome research. Experimental techniques such as yeast two-hybrid (Y2H) and affinity purification–mass spectrometry (AP–MS) have enabled large-scale mapping in reference organisms like *Homo sapiens* [15], *Caenorhabditis elegans* [23], and *Saccharomyces cerevisiae* [17]. Considering that experimental methods are generally cost-intensive, computational approaches are playing increasingly important roles. For instance, conventional homology-based methods [e.g., interolog mapping (IM) and domain–domain interaction (DDI) inference] leverage evolutionary conservation to infer PPIs in less-studied species [24–26]. More recently, deep learning models, such as D-SCRIPT [27], PPITrans [28], and SGPPi [29], have enabled accurate PPI predictions from sequence or structure features, often with cross-species generalization. Concurrently, breakthroughs in structure prediction, particularly AlphaFold2/3 [30–32] and ESMFold [33], have made proteome-scale modeling of both monomers and complexes feasible. These developments have enabled recent efforts to construct structure-informed interactomes in humans [22,34] and plants [35]. However, no comparable resource currently exists for agricultural animals.

To address this gap, we present the Agricultural Animals Structural Interactome Atlas (AASIA, <https://aasia.zzzdlab.com>), the first comprehensive database of structure-informed protein interactions for agricultural animals. AASIA covers five representative animals: *Anas platyrhynchos* (duck), *Bos taurus* (cattle), *Gallus gallus* (chicken), *Ovis aries* (sheep), and *Sus scrofa* (pig), and it contains a total of 410,421 high-confidence PPIs with predicted complex structures generated by ESMFold. These interactomes were constructed through an integrative pipeline that combines IM, DDI, and transformer-based deep learning, followed by large-scale structure prediction. The resulting species-specific networks include 43,886 interactions in duck, 173,331 in cattle, 61,593 in chicken, 40,624 in sheep, and 90,987 in pig, each annotated with structural and binding interface information. In addition to data exploration tools for querying, browsing, and visualization, the platform features a sequence-based prediction module that supports both *de novo* inference of interactions between user-submitted protein pairs and quantitative evaluation of sequence variant effects on interaction score. The current functionalities of AASIA will facilitate community-wide applications such as genotype–phenotype association, mutation impact prioritization, and structure-guided breeding strategies.

Materials and methods

Multi-species data collection and preprocessing

To construct a cross-species PPI prediction pipeline, we systematically collected PPI data from multiple sources. First, we retrieved experimentally supported PPI datasets for seven model organisms (i.e., *H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, *E. coli*, and *A. thaliana*) from three well-established public databases, including BioGRID [19], IntAct [20], and MINT [21]. The corresponding proteome sequences were obtained from the UniProt [36] database. Additionally, for the target species under investigation (i.e., *A. platyrhynchos*, *B. taurus*, *G. gallus*, *O. aries*, and *S. scrofa*), we collected proteome sequences also from the UniProt database. To ensure consistency and interoperability, protein identifiers were mapped to UniProt IDs, Entrez Gene IDs, Ensembl Gene IDs, gene names, and protein names, thereby facilitating seamless integration within AASIA.

To obtain high-quality PPIs from these seven model organisms, we first removed duplicate records and retained only those whose interaction-type annotations were consistent with physical interactions. Genetic interactions and records with non-physical or unspecified interaction types were excluded. The numbers of unique PPIs in each species before and after this filtering step are provided in Table S1. We then employed the HIPPIE [37] scoring scheme to assess the reliability of PPIs. Briefly, the scoring scheme computes a confidence score (ranging from 0 to 1) for each experimental PPI, which is based on the type of experimental evidence, the number of supporting publications, and the diversity of species in which the interaction has been observed. Note that only high-confidence interactions (HIPPIE score ≥ 0.73 [37]) were retained in model training. The breakdown of experimental evidence for

the source PPI datasets is summarized in Fig. S1, and the distribution of HIPPIE scores across the seven reference organisms is shown in Fig. S2. As a result, we compiled a non-redundant, high-confidence PPI dataset, which includes 84,622 PPIs for *H. sapiens*, 7,892 for *M. musculus*, 537 for *C. elegans*, 1,234 for *D. melanogaster*, 591 for *E. coli*, 18,395 for *S. cerevisiae*, and 4,371 for *A. thaliana*. Each PPI was assigned a confidence score as described above, providing a quantitative measure of interaction reliability.

Prediction of protein–protein interactions in agricultural animals

To predict protein interactions in agricultural animals, we employed a predictive framework integrating three individual prediction methods (Fig. 1). This ensemble model improves the accuracy of cross-species PPI inference. The methodological details are elaborated as follows.

Interolog mapping method

The core principle of IM is that if two proteins (A and B) have homologs (A' and B') that were reported to interact, then A and B are also likely to interact. To construct the interolog template library, we utilized known intra-species PPIs from the aforementioned protein interaction datasets collected in the preceding steps. Given an input protein pair within agricultural animals, BLASTP was used to identify homologous proteins by aligning sequences against the template library. Homologs were considered if they met the following criteria: sequence identity ≥ 0.3 , alignment coverage of the query protein ≥ 0.4 , and an E-value $\leq 1e-10$ [38–40]. For each identified interolog template, we assigned a quality score ($S_{im_template}$) based on the corresponding HIPPIE score [37]. The final interaction probability (S_{IM}) of a protein pair was computed by integrating the quality scores of all matched interolog templates using Bayesian integration [41],

$$S_{IM} = 1 - \prod_{i=1}^n (1 - S_{im_template,i}) \quad (1)$$

where $S_{im_template,i}$ is the quality score of the i -th matched interolog template, and n is the total number of involved interolog templates of the query protein pair.

Domain–domain interaction inference

The DDI inference method estimates the interaction probability of a given protein pair by detecting interacting domain pairs between them. Domain annotations for all proteins were obtained using Hmmscan [42] (E-value $\leq 1e-10$) based on the Pfam database [43]. To construct a comprehensive DDI template library, we first collected known DDIs from the 3did database [44]. In addition, we applied the expectation–maximization (EM) algorithm [45] to further expand it by identifying domain pairs that frequently co-occur in the curated PPIs in this study. A final DDI library was compiled by combining DDIs from 3did with those inferred through the EM algorithm. With the principle that DDIs collected from 3did should be more reliable, the confidence score ($S_{ddi_template}$) for each DDI in the library was assigned based on the following formula,

$$S_{ddi_template} = \frac{1}{2} \times (S_{DDI_known} + S_{DDI_EM}) \quad (2)$$

where S_{DDI_known} takes 1 or 0 respectively to represent whether the DDI is known to be from 3did or not, and S_{DDI_EM} is the score of the DDI from the EM algorithm, ranging from 0 to 1. For a given query protein pair, all matching domain pairs were retrieved from the DDI library. The final interaction probability score of a protein pair (S_{DDI}) was computed by integrating the confidence scores of all identified DDIs through Bayesian integration,

$$S_{DDI} = 1 - \prod_{i=1}^n (1 - S_{ddi_template,i}) \quad (3)$$

where $S_{ddi_template,i}$ is the confidence score of the i -th identified DDI, and n represents the number of DDIs associated with the query protein pair.

Deep learning method

The deep learning strategy for cross-species PPI prediction was implemented using PPITrans proposed by Wang et al. [28]. We retrained the predictive model on a curated dataset specific to our study. The model architecture consists of three key modules, including a pre-trained embedder, a protein encoder, and a PPI classifier [28]. Specifically, ProtT5 [46] was employed as the pre-trained language model to generate contextualized embeddings for each protein sequence. These embeddings were then processed by a transformer to capture long-range dependencies, followed by a pooling operation to obtain protein-level representations. The resulting representations were subsequently passed through a multi-layer perceptron to output an interaction probability score for the input pair.

To enhance biological relevance, we assembled positive samples from the compiled high-confidence human PPI set. From this set, we removed PPIs involving very short proteins (shorter than 50 amino acids) and, due to GPU memory constraints, also excluded proteins longer than 800 amino acids. This procedure yielded a final set of 44,035 high-confidence human PPIs as positive samples. Considering the inherent class imbalance in real-world PPI datasets, we adopted a positives-to-negatives ratio of 1:10 to better reflect biological realism and ensure training stability. Accordingly, 430,370 negative samples were generated by randomly pairing human proteins while excluding self-interactions and known positives. The dataset was split into training, validation, and testing sets at a ratio of 8:1:1. Model training was conducted using PyTorch (v2.0.0) and fairseq (v0.13.0). Training was performed on an NVIDIA A30 GPU with 24 GB of memory. Optimization was conducted using the AdamW optimizer with a learning rate of $3e-5$ and a batch size of 48, for a maximum of 20 epochs. Early stopping was applied based on the validation loss. After training, the model was applied to predict protein interactions within agricultural animals. The resulting probability scores provide a quantitative measure of interaction confidence (S_{DL}) ranging from 0 to 1.

Model integration by logistic regression

To maximize prediction performance, we integrated these individual scores (i.e., S_{IM} , S_{DDI} , and S_{DL}) in a feature vector and trained a logistic regression (LR) model to derive the overall interaction probability (S_{Total}) for each protein pair in agricultural animals. Specifically, a separate dataset was constructed using *M. musculus* PPIs for cross-species evaluation. The 1:10 positive-to-negative samples were constructed following the preprocessing procedure used to prepare the human PPI dataset for the deep learning method, yielding 5,992 positive pairs and 59,920 randomly generated negative pairs. The dataset was randomly split into 80% for training and 20% for independent testing. Model training was performed using scikit-learn (v1.3.2). The trained LR model was subsequently applied to predict PPIs in agricultural animals. To identify high-confidence interactions, we applied a stringent false positive rate (FPR) threshold of 0.1%.

Structural prediction of monomeric and complex proteins

To obtain 3D structures for both monomeric proteins and high-confidence interacting proteins, we employed ESMFold, a state-of-the-art structure prediction framework powered by the ESM-2 protein language model [33]. In contrast to traditional approaches that

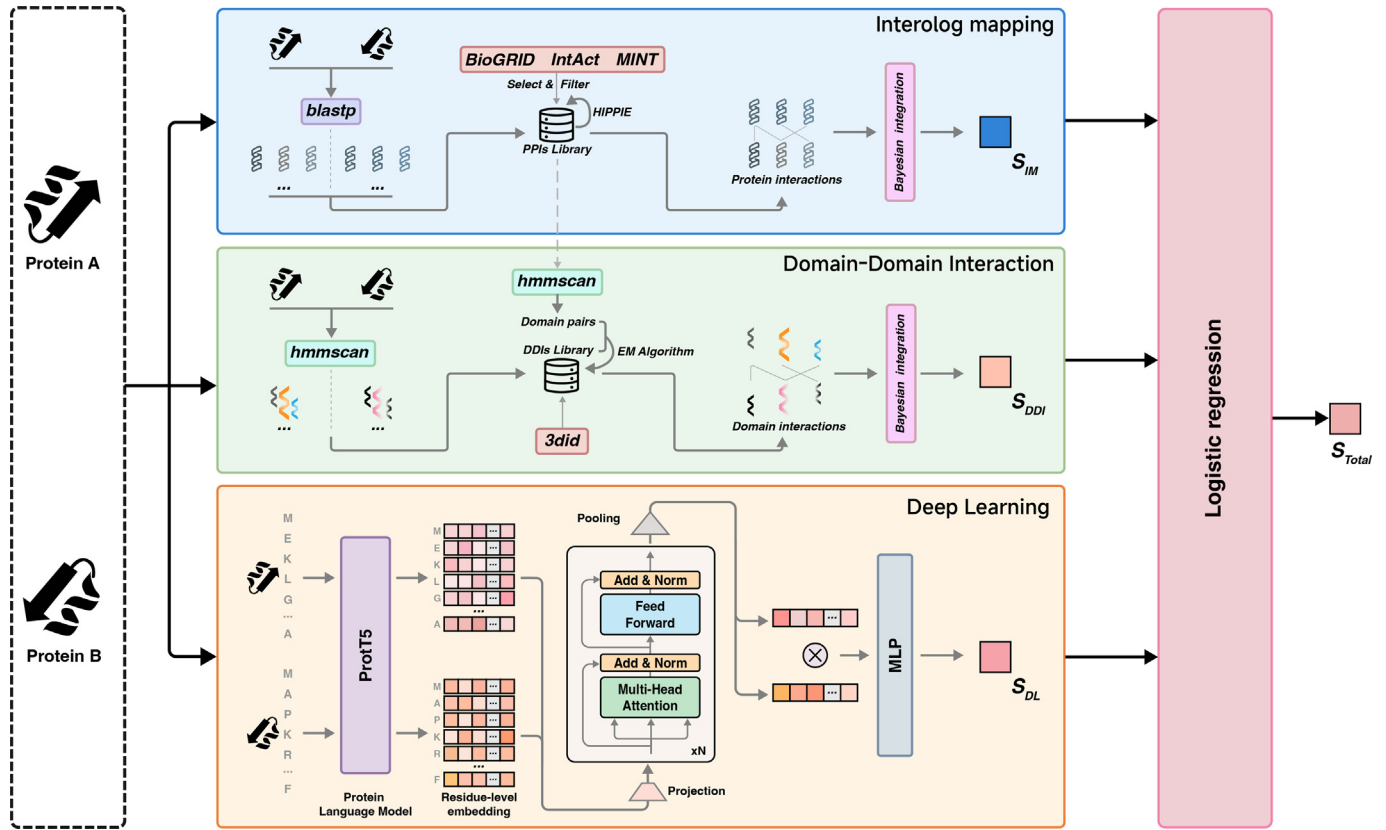


Fig. 1. The architecture of the PPI prediction framework for agricultural animals. The model integrates three complementary methods to predict PPIs, including (i) Interolog mapping, which transfers known interactions from model organisms based on sequence similarity; (ii) Domain-domain interaction, which infers potential interactions through conserved domain pairings; and (iii) Deep learning, which leverages a protein language model and transformer-based architecture to learn sequence-derived interaction patterns. Each method independently produces a confidence score: S_{IM} , S_{DDI} , and S_{DL} . These scores are then combined using a logistic regression (LR) model to output a final interaction probability S_{Total} ranging from 0 to 1.

rely on multiple sequence alignments or structural templates, ESMFold performs end-to-end structure prediction directly from amino acid sequences, substantially accelerating the process while maintaining high predictive accuracy. We utilized the official GPU-accelerated Linux implementation of ESMFold, following the standard pipeline provided in its GitHub repository. For monomer prediction, each protein sequence was individually submitted in FASTA format. For complex structure prediction, we followed ESMFold's multimer input format and encoded each interacting protein pair as a single FASTA entry, with the two chains separated by a ":" character. Given GPU memory limits, multimer modeling was restricted to pairs with a combined length $\leq 2,000$ residues.

Energetic evaluation of protein complexes

To quantitatively assess the energetic profiles of predicted protein complexes, we performed comprehensive binding energy calculations using FoldX 5.0, a robust force field-based platform for protein structure analysis [47]. Briefly, our computational workflow comprised the following three steps. First, structural energy optimization was conducted using the "RepairPDB" module to rectify steric clashes and geometric inaccuracies inherent in predicted models. Second, interaction energy was calculated by the "AnalyseComplex" protocol, which decomposes interaction energies into van der Waals forces, electrostatic complementarity, and solvation effects. Third, the "AlaScan" program was used to perform alanine scanning mutagenesis across all residues, and positions with $\Delta\Delta G \geq 0.5 \text{ kcal} \cdot \text{mol}^{-1}$ were designated as hotspots. Specifically,

interface residues were defined as those with inter-chain atomic distance $\leq 4\text{\AA}$ in the complex structures.

Database development

The AASIA platform adopts a modular three-tier architecture, with Vue.js for the frontend, Express.js as the backend framework, and MySQL for metadata management. The frontend is built with Vue 3.0 (<https://vuejs.org>), Vue Router (<https://router.vuejs.org>), and Vite (<https://vite.dev>) for efficient routing and build optimization, with Bootstrap (<https://getbootstrap.com/>) ensuring responsive design. Interactive data visualization is enabled through Cytoscape.js (<https://js.cytoscape.org/>) for PPI networks [48], Mol* (<https://molstar.org/>) [49] and NGL Viewer (<https://nglviewer.org/>) for 3D structural models [50], and ECharts.js (<https://echarts.apache.org/>) for statistical plots. The backend uses Express.js (<https://expressjs.com/>) to implement RESTful APIs, enabling structured access to interaction data, structures, and annotations. The entire system is deployed by Apache HTTP Server (<https://www.apache.org/>), ensuring robust web service delivery.

Results and discussion

The overall architecture of AASIA

As a meticulously curated, user-friendly, and interactive database, AASIA integrates multiple layers of biological information (Fig. 2). It encompasses a comprehensive collection of high-

confidence PPIs inferred through an integrated predictive framework that combines homology-based methods and deep learning, which illuminate the functional protein networks essential for understanding physiological traits and disease resistance mechanisms, thereby facilitating molecular breeding and health management in animals. In addition, the database offers monomeric and complex structural models predicted by ESMFold, enriched with annotations of interface residues, hotspot residues, and energy information.

Regarding its architectural design, AASIA incorporates several interactive modules to present information with clarity and accessibility, such as Search, Browse, and Predict (Fig. 2). The Search module allows retrieval of protein-level and interaction-level information using a variety of identifiers. The Browse module enables systematic navigation through protein interaction networks, structural features, and functional annotations. The Predict module accepts user-submitted protein sequences, performs de novo inference of potential interactions, and can also evaluate the effects of sequence mutations on interactions.

High-confidence prediction of species-specific protein interactomes

To systematically delineate the protein interaction landscapes of agricultural animals, we employed an integrative framework that combines three complementary predictive strategies: IM, DDI, and deep learning. These predictions were integrated using an LR model trained on a curated benchmark dataset, yielding a robust and generalizable interaction probability score for each protein pair. To ensure stringency, high-confidence interactions were defined at a 0.1% false positive rate (FPR) control (interaction score > 0.899), calibrated against an independent test set.

To evaluate the predictive performance of the proposed models, we plot precision-recall (PR) and receiver operating characteristic (ROC) curves and quantify the performance using the areas under the curves (AUPRC and AUROC). Given the class imbalance, we emphasize AUPRC as the more informative metric [51]. Using a curated *M. musculus* dataset (the 5,992 positive pairs and 59,920 negative pairs; see Materials and Methods), we benchmarked each component and the integrated model. As shown in Fig. 3, the LR model outperformed all individual predictors, achieving AUROC and AUPRC of 0.914 and 0.679, respectively (Fig. 3a, b). Leveraging this predictive pipeline, we then generated proteome-wide PPI networks for five agricultural animals. We characterized the distributions of predicted PPI scores across species (Fig. 3c) and summarized the size and proteome coverage of each predicted network in Table 1. The resulting networks include 43,886 PPIs for *A. platyrhynchos* (23.1% coverage), 173,331 PPIs for *B. taurus* (34.0% coverage), 61,593 PPIs for *G. gallus* (27.6% coverage), 40,624 PPIs for *O. aries* (19.5% coverage), and 90,987 PPIs for *S. scrofa* (28.1% coverage). As the foundational layer of the AASIA platform, these interactomes cover the corresponding proteomes ranging from 19.5% to 34.0%, representing a substantial expansion of interaction data in these non-model organisms. To investigate the quality of the predicted interactomes, we compared them with the corresponding PPI data from the published literature or public databases, although these data resources are very limited. First, we compared the AASIA *B. taurus* PPIs with previously published *B. taurus* PPIs predicted from D-SCRIPT [27], which contains 476,399 PPIs covering 17,811 proteins. Interestingly, these two PPI data sets share 8,127 PPIs involving 1,860 proteins, which are significantly overlapped (Fisher's exact test, $P < 2.2 \times 10^{-16}$; Fig. S3a). We further examined the overlap between our data and the *B. taurus* PPIs deposited in Interactome3D (2024_12 version) [38], a database that collects experimentally solved protein complex structures. Of the 632 non-redundant *B. taurus* interactions

in Interactome3D, 207 (32.75%) were also present in our AASIA, and the overlap was highly significant (Fisher's exact test, $P < 2.2 \times 10^{-16}$; Fig. S3b). Collectively, the above two comparative analyses provide additional evidence for the overall reliability of the predicted interactomes in AASIA.

Comprehensive structural modeling and annotation of protein monomers and complexes

Building upon the predicted species-specific interactomes, we next sought to construct 3D structural models to enable residue-level interpretation of protein interactions and functions across agricultural animals. To this end, we employed ESMFold [33], a state-of-the-art structure prediction model that generates 3D structures directly from sequences, without requiring multiple sequence alignments (MSA) or structural templates, thus enabling highly efficient large-scale modeling.

Structural modeling was performed for both individual proteins and interacting protein pairs. For monomers, ESMFold was applied across the entire proteomes of the five species, irrespective of interaction status. The majority of predicted structures exhibited high confidence, with pLDDT (predicted Local Distance Difference Test) scores predominantly above 70, indicating reliable fold-level accuracy and broad structural coverage (Fig. 3d). For interaction modeling, we used the multimer mode of ESMFold to generate binary complex structures for all high-confidence PPIs. The resulting models showed more variable pLDDT scores compared to monomers (Fig. 3e), yet a substantial proportion retained well-folded and plausible architectures, supporting their potential biological relevance.

To enhance structural interpretability, we systematically annotated each complex model with interfacial residues and hotspot sites. Interfacial residues were defined based on geometric contact criteria, while hotspot residues were identified through in silico alanine scanning. These structural annotations are fully integrated into the AASIA platform and accessible via the PPI-level search interface, facilitating downstream analyses such as binding site identification, variant effect prediction, and structure-function relationship exploration.

Overall, this structural modeling framework adds a critical mechanistic dimension to the AASIA interactomes, enabling the exploration of interaction specificity, structural constraints, and potentially trait-associated variations in a spatial context. These structure-based annotations establish a valuable foundation for future studies on protein function, interaction disruption, and rational design in agricultural animal systems.

Utility of the AASIA platform

AASIA features a unified and flexible search portal that supports both protein-level and interaction-level queries (Fig. 4a). Users can search using various identifiers, including AASIA IDs (our database-specific identifiers), UniProt, Ensembl, Entrez IDs, gene names, or general keywords, with support for both fuzzy search and exact matching. Search results are paginated, efficiently rendered, and linked to interactive views that enable intuitive data exploration.

Furthermore, the Browse module is organized around two complementary views. In the protein view (Fig. 4b), entries include basic identifiers and aliases, sequence and domain features, monomer structures predicted by ESMFold with interactive 3D visualization, and a network visualization of interacting partners labeled with gene symbols and accompanied by filters for confidence and partner count. Functional context is further provided through homologous proteins, Gene Ontology, and KEGG terms. This view allows users to move from primary sequence to structure and then

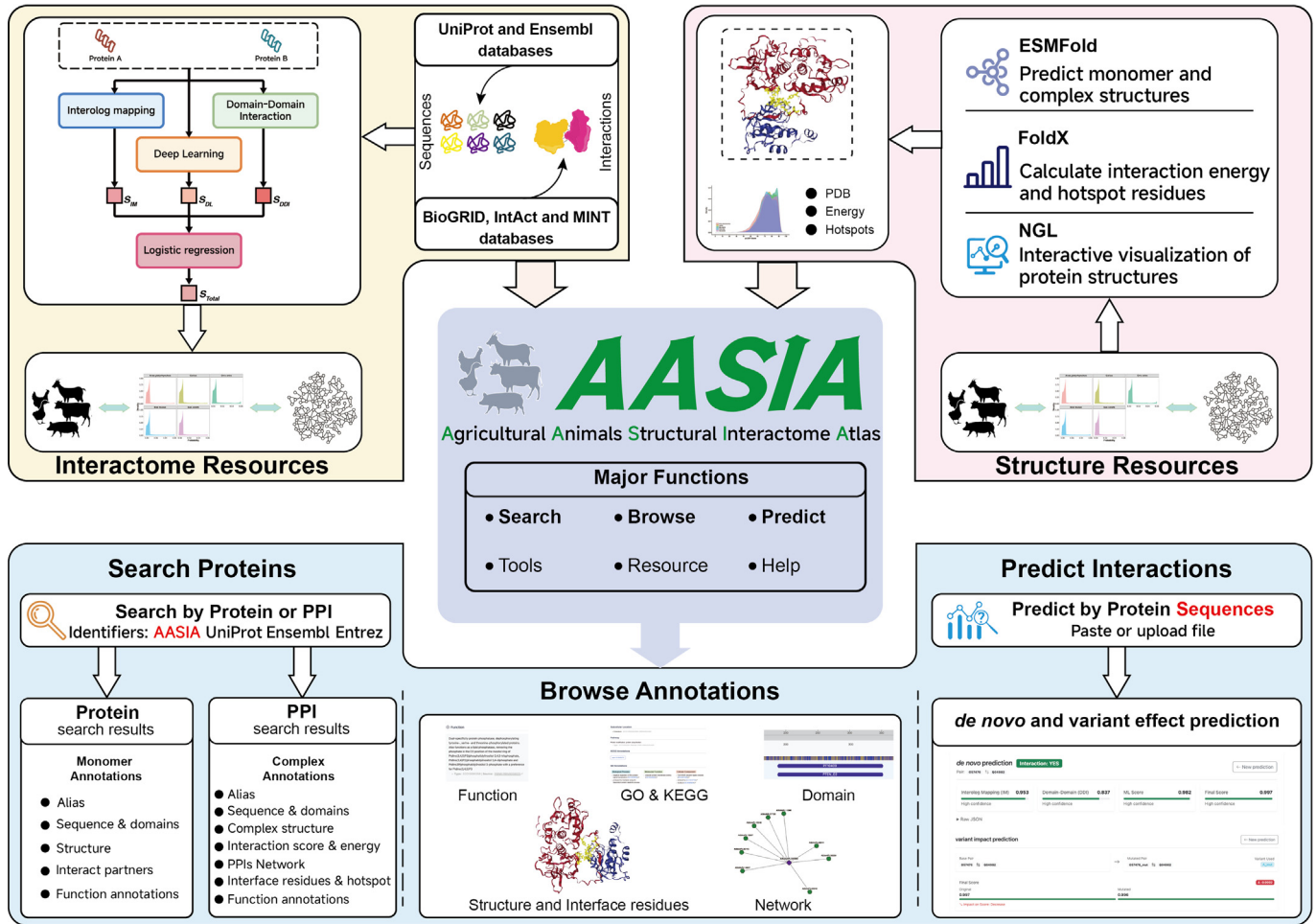


Fig. 2. Overview of AASIA. AASIA is a comprehensive database designed to compile and explore structural interactomes for agricultural animals. The interactomes are predicted by integrating three complementary methods, while structural annotations are derived from ESMFold and FoldX. The platform offers multiple functional modules. *Search* enables querying by protein or PPI identifiers. *Browse* provides access to annotations including protein function, structure, and network context. *Predict* allows users to submit sequences for interaction or variant effect prediction.

to the immediate network neighborhood in a single page. In the PPI view (Fig. 4c), each protein pair is presented with detailed structural and functional information, including a predicted complex structure, an interaction confidence score, and residue-level annotations. Interfacial residues and hotspot residues can be highlighted directly on the 3D model and are summarized in aligned tables for inspection. Energy-related information, including binding energies, is provided to aid interpretation of interaction affinity. Navigation between monomer and complex views is seamless, enabling residue-level analysis within biological and structural contexts.

Beyond data retrieval, AASIA incorporates a prediction module for *de novo* interaction inference and variant-effect evaluation. This module enables rapid and interpretable inference of candidate PPIs, making it suitable for uncharacterized or user-specified proteins of interest. Users can submit a pair of protein sequences from the same species through text input or file upload. Upon submission, the system applies the integrated predictive framework to perform interaction prediction for the query pair (Fig. 4d). To evaluate the effect of sequence variants, users provide amino acid substitutions in a compact notation to the sequence input. The system then reports the predicted interaction score for the wild-type pair, the score for the variant pair, and the difference between them. This functionality is particularly useful for assessing the potential effects of mutations, characterizing previously unannotated pro-

teins, and generating testable hypotheses in the absence of experimental data.

Biological insights from the *G. gallus* structural interactome

The predicted *G. gallus* structural interactome was exemplified to further interrogate the overall qualities of the predicted PPIs and 3D structures of interacting proteins in the current AASIA database. In addition to conducting global network analyses and functional analyses of predicted interactions, we provide rich case studies to demonstrate the reliable structure prediction of interacting proteins and highlight their potential biological significance.

Network topology analysis

We examined its global topological features of the *G. gallus* interactome (Fig. 5). The PPI network is densely connected, with an average of 24 and a median of 6 interactions per protein (Fig. 5a). In general, it follows similar network topology properties as previous PPI networks [17,52]. First, the network degree conforms to a scale-free power law distribution (Fig. 5b). Scale-free networks typically have many nodes with few edges, and only a few highly connected ones. The probability $P(k)$ of nodes having k edges decays as a power law $P(k) \propto k^{-\gamma}$ in scale-free networks. As shown in Fig. 5b, the *G. gallus* interactome can be characterized by a power law distribution, where $P(k) \propto k^{-1.156}$. Moreover, the *G.*

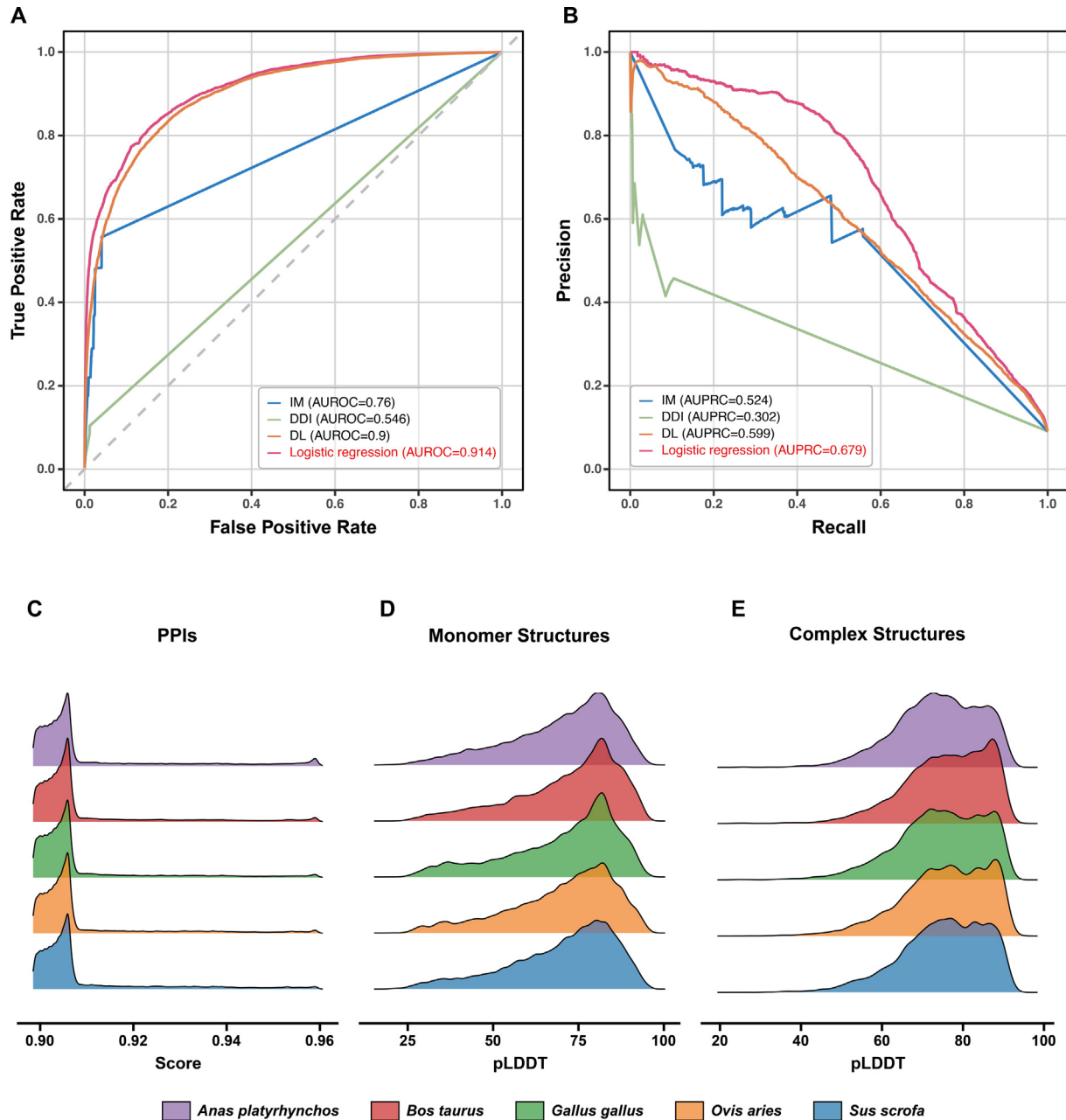


Fig. 3. Performance evaluation and prediction score distributions in AASIA. (a) Receiver operating characteristic (ROC) curves of four PPI prediction models on the *M. musculus* dataset. The logistic regression model achieved the highest AUROC (0.914), outperforming DL, IM, and DDI. (b) Precision–recall (PR) curves, with logistic regression showing the best overall balance ($AUPRC = 0.679$). (c) Distribution of predicted PPI scores across five agricultural animals, highlighting a consistent skew toward high-confidence predictions. (d) Distribution of pLDDT scores for ESMFold-predicted monomer structures, with the majority of models exceeding the high-confidence threshold ($pLDDT > 70$). (e) pLDDT score distributions of predicted complex structures.

Table 1

The number of predicted high-confidence PPIs in the five agricultural animals.

Species	Proteome ID	No. of proteins ^a	Coverage	No. of high-confidence PPIs	No. of shared PPIs ^b
<i>Anas platyrhynchos</i>	UP000296049	16,574	3,828/16,574 = 23.1%	43,886	0
<i>Bos taurus</i>	UP000009136	23,836	8,092/23,836 = 34.0%	173,331	86
<i>Gallus gallus</i>	UP000000539	18,372	5,077/18,372 = 27.6%	61,593	50
<i>Ovis aries</i>	UP000002356	21,218	4,144/21,218 = 19.5%	40,624	0
<i>Sus scrofa</i>	UP000008227	22,800	6,407/22,800 = 28.1%	90,987	2

^a Proteins are from the UniProt one protein sequence per gene dataset (i.e., one entry per protein-coding gene); alternative isoforms were not taken into account.

^b Shared PPIs represent overlaps between the predicted PPIs and experimental PPIs reported in BioGRID, IntAct, and MINT.

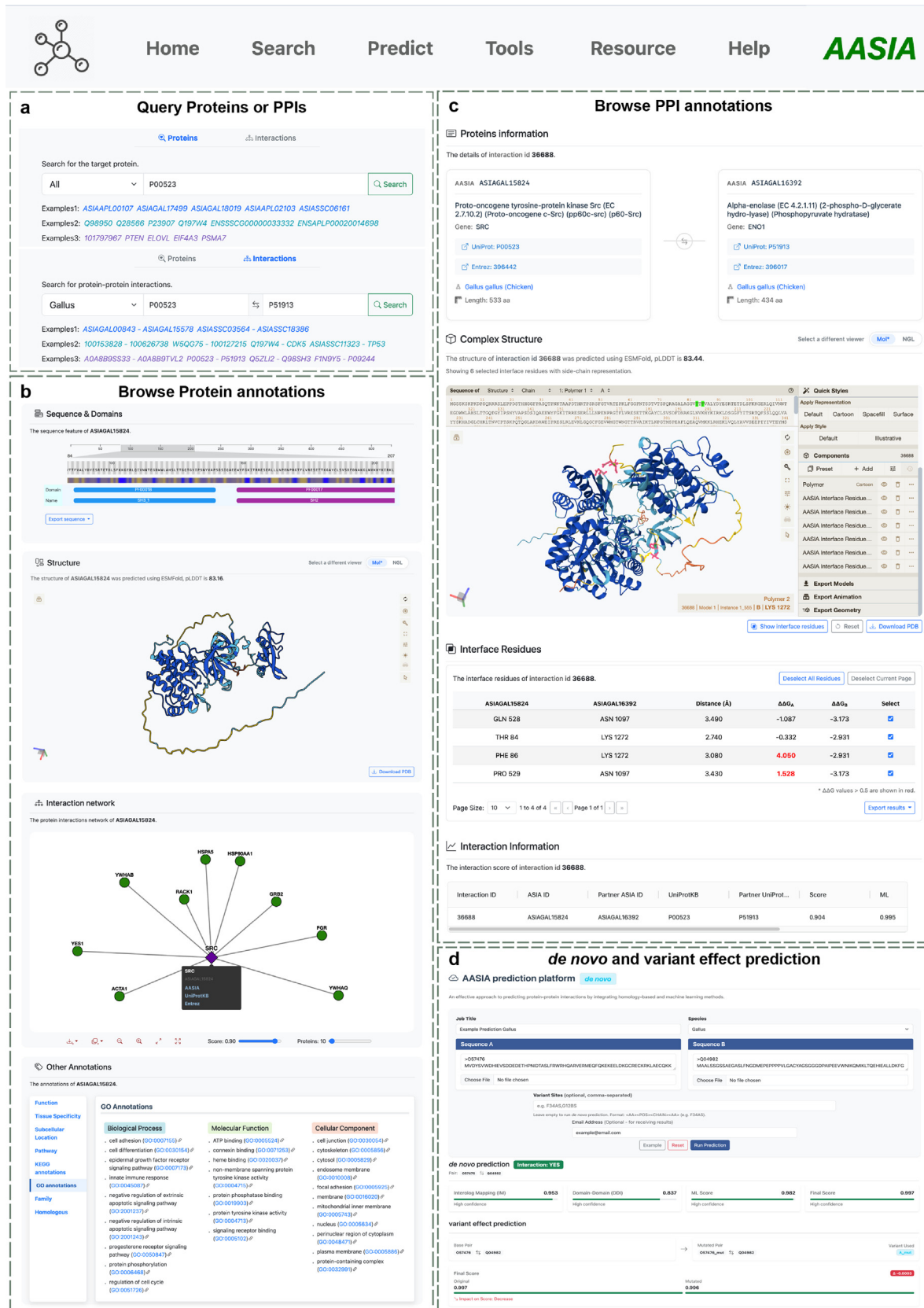


Fig. 4. Functional modules of AASIA. (a) Search interface supporting both protein- and PPI-level queries. Users can input multiple types of identifiers and select the desired search mode. (b) Protein annotation page, displaying sequence and domain features, monomer structure, interaction partners, and associated GO and KEGG functional terms. Interaction networks are shown with nodes labeled by gene symbols. (c) PPI annotation page, showing interaction-specific information for both proteins, including predicted complex structure, interface residues, and interaction scores. (d) PPI prediction module for sequence-based interaction inference and variant effect assessment. Users can input two protein sequences and specify the target species. For variant prediction, users enter mutations in the variant field. The module returns the overall interaction probability for the query pair and, when mutations are provided, the change in interaction score. Results are downloadable and can optionally be sent via email.

Gallus interactome exhibits small-world properties, characterized by an average shortest path length (L) of 3.3 (Fig. 5c) and a clustering coefficient (CC) of 0.63. Compared to randomized networks of the same size and degree distribution ($L_{\text{random}} = 3.64$, $C_{\text{random}} = 0.01$), the observed network meets the classical small-world criteria (i.e., a much higher clustering coefficient than randomized networks and a comparable short path length).

Comparison with experimentally supported interactions

To assess the reliability of our predicted PPIs, we first compared them against experimentally supported *G. gallus* interactions curated from public databases [19–21]. We observed 50 overlapping interaction pairs between our predictions and the experimentally supported PPI set (Fig. 5d; Table 1). This overlap represented significant enrichment (Fisher's exact test, $P = 2.8 \times 10^{-5}$). Moreover, we also compared the predicted *G. gallus* interactions against 1,000 degree-preserving randomized networks, which retained the original proteins, number of PPIs, and degree of every protein while randomizing interaction partners. The randomized networks yielded a mean overlap of 15.72 pairs (standard deviation = 2.92; maximum = 24), and none reached the observed value of 50 (one-sided empirical $P < 0.001$). Gene Ontology (GO) and KEGG enrichment analysis of overlapping interactions highlighted biologically relevant terms (e.g., protein folding, FoxO signaling pathway) (Fig. S4). Moreover, we also compared the GO term semantic similarity across predicted, experimentally supported, and randomly paired interactions within the AASIA *G. gallus* network. The predicted PPIs exhibited significantly higher GO semantic similarity in both biological process (BP) and cellular component (CC) compared to random pairs, and performed comparably to experimentally supported interactions (Fig. 5e, f).

Structural characterization of the actin-capping protein assembly

To further evaluate the reliability and structural plausibility of the predicted interactome, we focused on a selection of PPIs with high-confidence prediction scores and well-defined structural characteristics. As a representative case, we examined the capping protein heterodimer (CAPZA1 and CAPZB) and their interactions with ACTIN (Fig. 6). ACTIN monomers polymerize into polarized filaments with barbed and pointed ends. CAPZA1 and CAPZB assemble into a stable heterodimeric capping protein, CapZ, that exhibits pseudo twofold symmetry and features flexible C-terminal extensions. The symmetric core of CapZ positions the subunits at the filament terminus, while the flexible extensions engage grooves on the two terminal ACTIN monomers. These features enable CapZ to bind the barbed ends with high affinity and 1:1 stoichiometry, thereby preventing both polymerization and depolymerization of actin filaments [53–55].

Within the *G. gallus* interactome, AASIA recovered the experimentally supported binary interactions among CAPZA1, CAPZB, and ACTIN. IM templates for CAPZA1–CAPZB and CAPZB–ACTIN were obtained from *H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*, and *S. cerevisiae*, whereas those for CAPZA1–ACTIN were obtained from the same set except *C. elegans*. Structural modeling of the CAPZA1 and CAPZB heterodimer showed strong agreement with the crystal structure (PDB: 1IZN), with an RMSD of 1.6 Å, indicating near-native accuracy achieved by ESMFold (Fig. 6a). Although no experimental complex structures are available for the subunit (CAPZA1/CAPZB) and ACTIN pairs, the predicted binary complexes showed high confidence. The CAPZA1–ACTIN pair has an interaction score of 0.91 with pLDDT 79.57, and the CAPZB–ACTIN pair achieves an interaction score of 0.91 with pLDDT 77.11 (Fig. 6b). Notably, the observed binding regions of CapZ aligned with previously reported actin-interacting motifs. Residues 259 to 286 in CAPZA1 and residues 266 to 277 in CAPZB were both

accurately positioned at the predicted CapZ–ACTIN interface, in agreement with mutagenesis data identifying these segments as critical for binding [53]. This concordance indicates that the predicted complexes are generally reliable and makes them suitable for downstream functional analyses.

Building on prior evidence that capping proteins can engage two actin monomers simultaneously at the filament ends, we constructed a ternary complex model by ESMFold that integrates CAPZA1, CAPZB, and two ACTIN monomers, mimicking their engagement at filament barbed ends. The resulting structure exhibited strong structural coherence and consistent interaction features (Fig. 6c), further reinforcing the reliability of the interaction and the structural robustness of our prediction framework.

Structural analysis of a trait-associated hotspot residue at the protein interface

Beyond recapitulating known structural assemblies, the predicted interactome in AASIA also serves as a valuable resource for investigating functional changes caused by genomic variants. It has been common knowledge that perturbations of PPIs may contribute to a wide range of economically important traits of agricultural animals, such as growth performance, reproductive traits, and immune responses [56–58]. By mapping genomic variant information onto the structural models of interacting proteins, AASIA enables structure-based exploration of how such variants perturb molecular interactions to drive phenotypic diversity.

To explore this application, we investigated a predicted high-confidence *G. gallus* protein interaction between XRCC5, a DNA repair factor, and H2AC39, a histone variant. The predicted interaction was simultaneously supported by IM, DDI inference, and deep learning, with individual scores of 0.99, 0.57, and 0.98, respectively. Human PPIs provided the templates for the IM method, and integration by logistic regression yielded a final interaction score of 0.94. Although this interaction currently lacks direct experimental evidence, the predicted complex structure suggested a plausible and specific interface (pLDDT = 61.84; Fig. 6d). Notably, this interface region overlaps with a known quantitative trait locus (QTL 125459) associated with eggshell strength in *G. gallus* [56,59]. This QTL, identified through association analysis (peak SNP: rs15864122, $P = 0.021$), suggests that sequence variation at this site may influence eggshell mechanical properties. Within our model, the QTL-linked residue R10 on XRCC5 was located directly at the interaction interface and was identified computationally as a putative hotspot residue. Computational alanine scanning using FoldX revealed that R10A mutation may weaken the interaction ($\Delta\Delta G = 1.09 \text{ kcal} \cdot \text{mol}^{-1}$; Fig. 6d). Interestingly, our variant effect prediction module in AASIA also suggested a reduction in the interaction score for the R10A mutation ($\Delta_{\text{interaction}} = -0.03$). Together, these observations provide a structure-based hypothesis that genetic variation at this position may affect XRCC5–H2AC39 binding, potentially influencing chromatin dynamics and gene regulation during eggshell mineralization. Thus, this case study informs how AASIA can facilitate the generation of testable genotype–phenotype hypotheses through structural analysis of predicted interactions, providing new hints for deciphering agricultural animal functional genomics.

Limitations of the current study

Our current study still suffers from several limitations. Regarding the dataset preparation, experimental records labeled as physical interactions in public PPI databases do not uniformly represent direct binary interactions, as large-scale approaches such as AP-MS and Co-IP capture protein associations rather than direct physical binding. As shown in Fig. S1, although applying a HIPPIE confi-

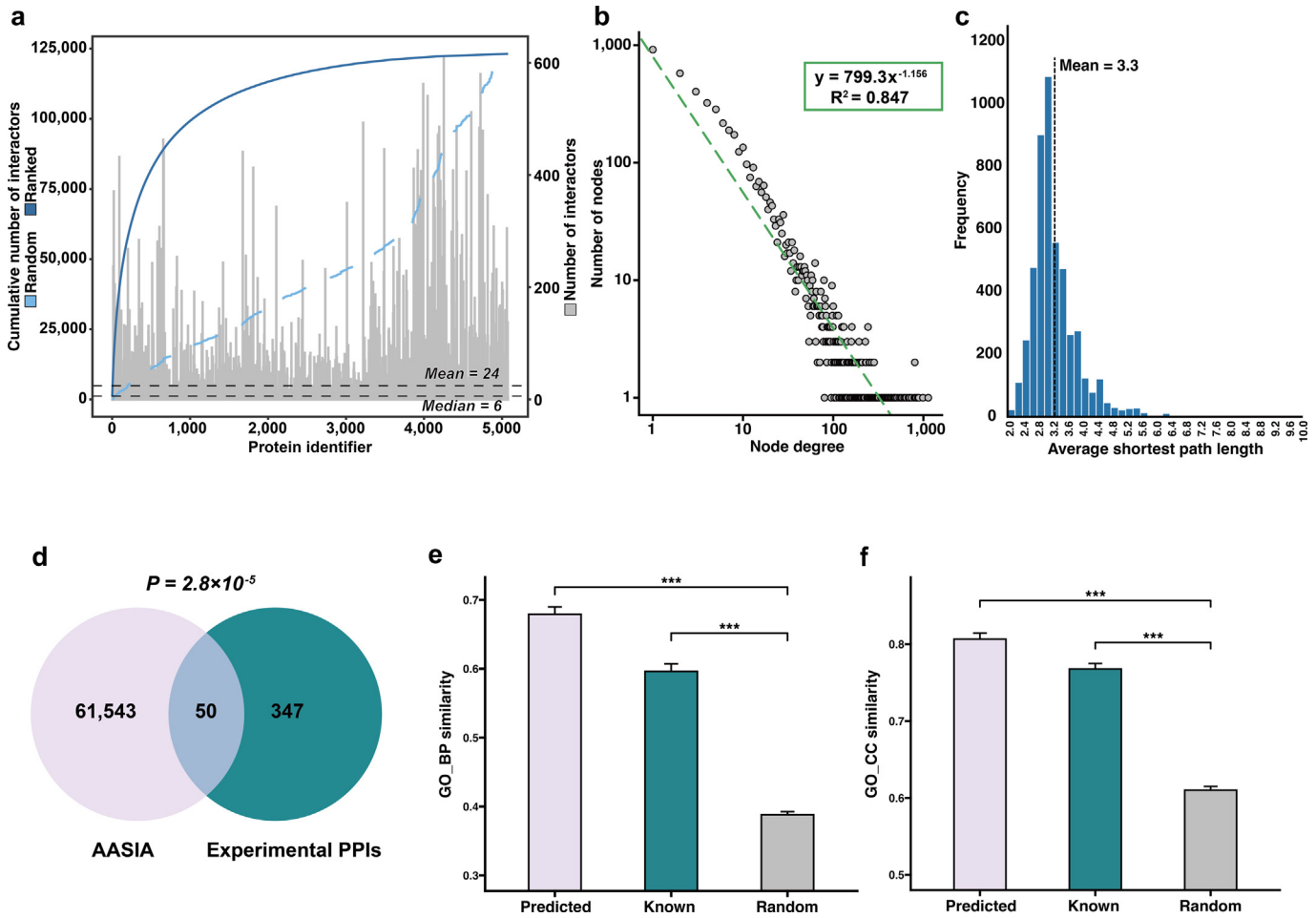


Fig. 5. Network and functional characterization of the predicted *G. gallus* interactome. (a) Distribution of the number of interactors per protein. The sorted cumulative number of interactions reaches saturation after ~3,000 proteins, with a mean of 24 interactors and a median of 6 per protein. (b) Degree distribution of the network follows a power-law behavior, consistent with a scale-free topology. A linear fit on log-log scale is shown in green ($R^2 = 0.847$). (c) Distribution of average shortest path lengths between all connected protein pairs shows a mean path length of 3.3. (d) The venn diagram showing the overlap between predicted PPIs in AASIA and experimentally supported PPIs (Fisher's exact test, $P = 2.8 \times 10^{-5}$). (e-f) Comparison of functional association features among predicted PPIs, known PPIs, and random protein pairs. Bars show mean similarity scores, with 95% confidence intervals (calculated as mean $\pm 1.96 \times$ SEM). One-tailed Mann-Whitney U tests were used to assess random pairs with both known and predicted PPIs (***) $P < 0.001$. (e) GO_BP similarity. (f) GO_CC similarity.

dence threshold of 0.73 reduced the proportion of association-supported interactions, they were not entirely excluded from the compiled dataset. To assess the impact of this data heterogeneity, we evaluated the integrated predictive model on a stricter *M. musculus* test subset retaining only binary-supported PPIs as positive samples, and it achieved an AUROC of 0.928 and an AUPRC of 0.729 (Fig. S5). The overall performance is reasonably better than on the original test set (AUROC = 0.914, AUPRC = 0.679), suggesting that the model retains strong discriminative power for direct PPIs despite training data heterogeneity. Even so, the predicted interactomes may inevitably contain indirect physical interactions. Consequently, the predicted PPIs should be interpreted with awareness that the source evidence used for model construction includes both direct physical and association-supported interactions.

Regarding the methodology for PPI prediction, the well-established IM method was integrated into our final predictive model. To implement the IM method, we used empirical criteria (sequence identity ≥ 0.3 , alignment coverage of the query protein ≥ 0.4 , and an E-value $\leq 1e-10$) for homolog detection [38–40]. To justify the threshold's rationality, we investigated different levels of joint sequence identity (JSI) for homolog detection. Unlike sequence identity alone, JSI incorporates both sequence identity

and alignment coverage for the two proteins in an interolog pair, thereby providing a stricter measure of overall homology. When tested on the *M. musculus* benchmark test set, our integrative model with increasing JSI stringency can yield a slightly higher AUPRC than our current homology detection threshold (Fig. S6a). Under the same prediction threshold in the final integrative model, the precision value can be modestly improved with increasing JSI stringency. However, the final integrative model also results in a monotonic decrease in Recall due to the progressive loss of interolog evidence (Fig. S6a). Regarding the size of predicted interactomes, the retained interaction ratio decreased consistently across all five agricultural species as JSI increased (Fig. S6b). It should be emphasized that the final high-confidence interactions in AASIA are determined by the integrative LR model, which combines evidence from IM, DDI, and DL components, rather than by the IM alone. Therefore, we believe our choice of homology detection is generally reasonable, although it is not optimal.

Regarding the 3D structure prediction of interacting proteins, some important information was not taken into account. For instance, it has been well established that intrinsically disordered regions (IDRs) play important roles in mediating protein interactions [60,61]. The current study uses ESMFold to predict the 3D structures of interacting proteins. To our best knowledge, the IDR

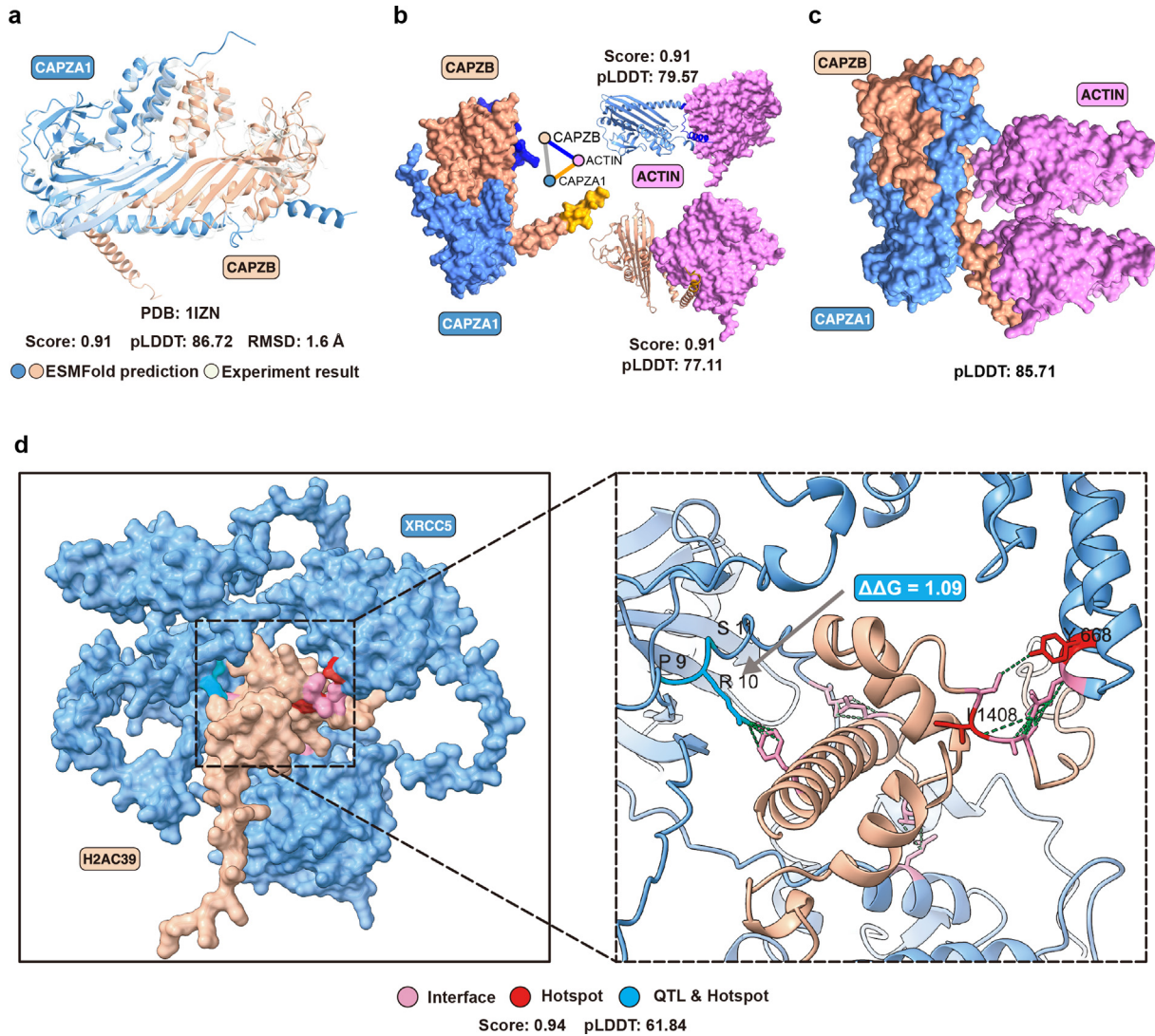


Fig. 6. Structural-based support and functional interpretation of predicted protein-protein interactions. (a–c) Structural modeling of the *G. gallus* actin-capping complex involving CAPZA1, CAPZB, and ACTIN. (a) Predicted complex structure of CAPZA1–CAPZB aligned with the experimentally resolved crystal structure (PDB: 1IZN), showing high structural similarity (RMSD = 1.6 Å) and model confidence (interaction score = 0.91; pLDDT = 86.72). (b) Predicted binary interactions between CAPZA1 or CAPZB and ACTIN, each with high confidence scores and pLDDT values. Predicted binding interfaces closely match experimentally reported actin-binding regions at the C-termini of CapZ subunits. (c) Ternary complex structure of CAPZA1–CAPZB–ACTIN(2), consistent with the known two-two interaction. (d) Predicted structure of the *G. gallus* XRCC5–H2AC39 interaction (interaction score = 0.94; pLDDT = 61.84). The left panel shows the overall complex. Interface residues are colored pink, predicted hotspots are colored red, and the QTL-linked residue R10 (blue outline) is highlighted. The associated QTL (ID: 125459) is linked to eggshell strength in *G. gallus*. The right panel shows a close-up of the interface highlighting R10, which is both a hotspot and a QTL-linked position. FoldX suggested a destabilizing effect upon mutation ($\Delta\Delta G = 1.09 \text{ kcal} \cdot \text{mol}^{-1}$), and our AASIA variant effect prediction module suggested a decrease in interaction score ($\Delta_{\text{interaction}} = -0.03$).

information has not been explicitly used in ESMFold. Indeed, incorporating IDR information to improve PPI predictions has become a new frontier in protein structure prediction. Technically, it remains an unsolved issue, although some promising results have been achieved. In the future AASIA version, we should keep pace with the newest advances in protein structure predictions to generate more accurate structure predictions of protein complexes as well as provide rich structural annotations for the predicted 3D structures.

Conclusion

In this work, we present AASIA (<https://aasia.zzdlab.com>), the first comprehensive resource of high-confidence, structure-informed interactomes for five agricultural animals. By integrating multiple prediction strategies, AASIA encompasses over 410,000 non-redundant high-confidence PPIs, each corresponding with

complex structures and residue-level interface annotations at the proteome scale. The user-friendly platform enables researchers not only to explore large-scale PPI networks at the systems level, but also to investigate fine-grained molecular details such as binding interfaces, hotspot residues, and trait-associated variants in the 3D structural context. Regarding the future development of AASIA, there exist two important directions. First, expanding AASIA to include host-pathogen PPI data could reveal conserved molecular interfaces involved in pathogen infection and host immunity. Structure-informed modeling of such interactions may facilitate the identification of therapeutic targets or vaccine epitopes for key pathogens such as the avian influenza virus and the African swine fever virus. Second, integrating AASIA with complementary omics information (e.g., transcriptomics, epigenomics, and genome-wide variant data) could enable more precise and complete functional annotation. Taken together, it is hoped that AASIA can serve as an important structural interactome database for

agricultural animals, and play a crucial role in accelerating the transformation of omics data into animal biotech breeding innovation.

AI use disclosure statement

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to improve the language and readability of the text. The tool was used solely for language editing and grammatical refinement. The authors reviewed and edited all content and take full responsibility for the accuracy, integrity, and scientific validity of the published article. AI tools were not used for scientific content generation, data interpretation, reference collection, or figure preparation.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

CRediT authorship contribution statement

Jiajun Li: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Mengdi Yuan:** Conceptualization, Methodology, Validation, Investigation. **Linyang Jiang:** Investigation, Data curation. **Dianke Li:** Software, Data curation. **Zhongtao Yin:** Validation. **Feng Zhu:** Data curation. **Wenyu Shi:** Supervision, Project administration. **Zhuocheng Hou:** Conceptualization, Supervision, Project administration, Funding acquisition. **Ziding Zhang:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

The data that support the findings of this study are openly available in AASIA at <https://aasia.zzzdlab.com>.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Drs. Xiaokun Guo and Shiping Yang for the fruitful discussion. This work was supported by the National Key Research and Development Program of China (2022YFF1000104) and the National Natural Science Foundation of China (32270703).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jare.2026.09.015>.

References

- [1] Hellstrand S. Animal production in a sustainable agriculture. *Environ Dev Sustain* 2013;15:999–1036. doi: <https://doi.org/10.1007/s10668-012-9423-z>.
- [2] Davis TC, White RR. Breeding animals to feed people: the many roles of animal reproduction in ensuring global food security. *Theriogenology* 2020;150:27–33. doi: <https://doi.org/10.1016/j.theriogenology.2020.01.041>.
- [3] White RR, Hall MB. Nutritional and greenhouse gas impacts of removing animals from US agriculture. *PNAS* 2017;114:E10301–8. doi: <https://doi.org/10.1073/pnas.1707322114>.
- [4] Pan Z, Yao Y, Yin H, Cai Z, Wang Y, Bai L, et al. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat Commun* 2021;12:5848. doi: <https://doi.org/10.1038/s41467-021-26153-z>.
- [5] White RR, Brady M, Capper JL, McNamara JP, Johnson KA. Cow–calf reproductive, genetic, and nutritional management to improve the sustainability of whole beef production systems. *J Anim Sci* 2015;93:3197–211. doi: <https://doi.org/10.2527/jas.2014-8800>.
- [6] Gomiero T, Pimentel D, Paoletti MG. Is there a need for a more sustainable agriculture? *Crit Rev Plant Sci* 2011;30:6–23. doi: <https://doi.org/10.1080/07352689.2011.553515>.
- [7] Zhao X, Jiang L, Fang X, Guo Z, Wang X, Shi B, et al. Host-microbiota interaction-mediated resistance to inflammatory bowel disease in pigs. *Microbiome* 2022;10:115. doi: <https://doi.org/10.1186/s40168-022-01303-1>.
- [8] Hill EM, Prosser NS, Ferguson E, Kaler J, Green MJ, Keeling MJ, et al. Modelling livestock infectious disease control policy under differing social perspectives on vaccination behaviour. *PLoS Comput Biol* 2022;18:e1010235. doi: <https://doi.org/10.1371/journal.pcbi.1010235>.
- [9] Yang W, Liu X, Wang X. The immune system of chicken and its response to H9N2 avian influenza virus. *Vet Q* 2023;43:1–14. doi: <https://doi.org/10.1080/01652176.2023.2228360>.
- [10] Carrique L, Fan H, Walker AP, Keown JR, Sharps J, Staller E, et al. Host ANP32A mediates the assembly of the influenza virus replicase. *Nature* 2020;587:638–43. doi: <https://doi.org/10.1038/s41586-020-2927-z>.
- [11] Zhu Z, Mao R, Liu B, Liu H, Shi Z, Zhang K, et al. Single-cell profiling of African swine fever virus disease in the pig spleen reveals viral and host dynamics. *PNAS* 2024;121:e2312150121. doi: <https://doi.org/10.1073/pnas.2312150121>.
- [12] Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. *Nature* 2007;450:973–82. doi: <https://doi.org/10.1038/nature06523>.
- [13] Zhou Y, Liu Y, Gupta S, Paramo MI, Hou Y, Mao C, et al. A comprehensive SARS-CoV-2–human protein–protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. *Nat Biotechnol* 2023;41:128–39. doi: <https://doi.org/10.1038/s41587-022-01474-0>.
- [14] Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 2015;163:712–23. doi: <https://doi.org/10.1016/j.cell.2015.09.053>.
- [15] Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. *Nature* 2020;580:402–8. doi: <https://doi.org/10.1038/s41586-020-2188-x>.
- [16] Skinnider MA, Scott NE, Prudova A, Kerr CH, Stoyanov N, Stacey RG, et al. An atlas of protein–protein interactions across mouse tissues. *Cell* 2021;184:4073–4089.e17. doi: <https://doi.org/10.1016/j.cell.2021.06.003>.
- [17] Michaelis AC, Brunner A-D, Zwiebel M, Meier F, Strauss MT, Bludau I, et al. The social and structural architecture of the yeast protein interactome. *Nature* 2023. doi: <https://doi.org/10.1038/s41586-023-06739-5>.
- [18] Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 2023;51:D638–46. doi: <https://doi.org/10.1093/nar/gkac1000>.
- [19] Oughtred R, Rust J, Chang C, Breitkreutz B, Stark C, Willems A, et al. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* 2021;30:187–200. doi: <https://doi.org/10.1002/pro.3978>.
- [20] Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res* 2022;50:D648–53. doi: <https://doi.org/10.1093/nar/gkab1006>.
- [21] Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012;40:D857–61. doi: <https://doi.org/10.1093/nar/gkr930>.
- [22] Xiong D, Qiu Y, Zhao J, Zhou Y, Lee D, Gupta S, et al. A structurally informed human protein–protein interactome reveals proteome-wide perturbations caused by disease mutations. *Nat Biotechnol* 2024. doi: <https://doi.org/10.1038/s41587-024-02428-4>.
- [23] Simonis N, Rual J-F, Carvunis A-R, Tasan M, Lemmens I, Hirozane-Kishikawa T, et al. Empirically controlled mapping of the *Caenorhabditis elegans* protein–protein interactome network. *Nat Methods* 2009;6:47–54. doi: <https://doi.org/10.1038/nmeth.1279>.
- [24] Thanasomboon R, Kalapanulak S, Netphan S, Saithong T. Prediction of cassava protein interactome based on interolog method. *Sci Rep* 2017;7:17206. doi: <https://doi.org/10.1038/s41598-017-17633-2>.
- [25] Cui J, Li P, Li G, Xu F, Zhao C, Li Y, et al. AtPID: *Arabidopsis thaliana* protein interactome database an integrative platform for plant systems biology. *Nucleic Acids Res* 2008;36:D999–D. doi: <https://doi.org/10.1093/nar/gkm844>.
- [26] Yang X, Lian X, Fu C, Wuchty S, Yang S, Zhang Z. HVIDB: a comprehensive database for human–virus protein–protein interactions. *Brief Bioinform* 2021;22:832–44. doi: <https://doi.org/10.1093/bib/bbaa425>.
- [27] Sledzieski S, Singh R, Cowen L, Berger B. D-SCRIPT translates genome to phenotype with sequence-based, structure-aware, genome-scale predictions of protein–protein interactions. *Cell Syst* 2021;12:969–982.e6. doi: <https://doi.org/10.1016/j.cels.2021.08.010>.
- [28] Yang S, Cheng P, Liu Y, Feng D, Wang S. Exploring the knowledge of an outstanding protein to protein interaction transformer. *IEEE/ACM Trans*

- Comput Biol Bioinformatics 2024;21:1287–98. doi: <https://doi.org/10.1109/TCBB.2024.3381825>.
- [29] Huang Y, Wuchty S, Zhou Y, Zhang Z. SGPPi: structure-aware prediction of protein–protein interactions in rigorous conditions with graph convolutional network. *Brief Bioinform* 2023;24:bbad020. doi: <https://doi.org/10.1093/bib/bbad020>.
- [30] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9. doi: <https://doi.org/10.1038/s41586-021-03819-2>.
- [31] Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer 2022:2021.10.04.463034. 10.1101/2021.10.04.463034.
- [32] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;630:493–500. doi: <https://doi.org/10.1038/s41586-024-07487-w>.
- [33] Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;379:1123–30. doi: <https://doi.org/10.1126/science.adc2574>.
- [34] Burke DF, Bryant P, Barrio-Hernandez I, Memon D, Pozzati G, Shenoy A, et al. Towards a structurally resolved human protein interaction network. *Nat Struct Mol Biol* 2023;30:216–25. doi: <https://doi.org/10.1038/s41594-022-00910-8>.
- [35] Zheng S, Blaschek L, Pottier D, Dijkhof LRH, Özmen B, Lim PK, et al. Pupylation-based proximity labeling unravels a comprehensive protein and phosphoprotein interactome of the arabidopsis TOR complex. *Adv Sci* 2025;12:e2414496. doi: <https://doi.org/10.1002/advs.202414496>.
- [36] Bateman A, Martin M-J, Orchard S, Magrane M, Adesina A, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2025. *Nucleic Acids Res* 2025;53:D609–17. doi: <https://doi.org/10.1093/nar/gkaf1010>.
- [37] Schaefer MH, Fontaine J-F, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: integrating protein interaction networks with experiment based quality scores. *PLoS One* 2012;7:e31826. doi: <https://doi.org/10.1371/journal.pone.0031826>.
- [38] Mosca R, Céol A, Aloy P. Interactome3D: adding structural details to protein networks. *Nat Methods* 2013;10:47–53. doi: <https://doi.org/10.1038/nmeth.2289>.
- [39] García-García J, Schleker S, Klein-Seetharaman J, Oliva B. BIPS: BIANA Interolog Prediction Server. A tool for protein–protein interaction inference. *Nucleic Acids Res* 2012;40:W147–51. doi: <https://doi.org/10.1093/nar/gks553>.
- [40] Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han J-D-J, et al. Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res* 2004;14:1107–18. doi: <https://doi.org/10.1101/gr.1774904>.
- [41] Lian X, Yang X, Shao J, Hou F, Yang S, Pan D, et al. Prediction and analysis of human-herpes simplex virus type 1 protein–protein interactions by integrating multiple methods. *Quant Biol* 2020;8:312–24. doi: <https://doi.org/10.1007/s40484-020-0222-5>.
- [42] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011;39:W29–37. doi: <https://doi.org/10.1093/nar/gkr367>.
- [43] El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res* 2019;47:D427–32. doi: <https://doi.org/10.1093/nar/gky995>.
- [44] Mosca R, Céol A, Stein A, Olivella R, Aloy P. 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res* 2014;42:D374–9. doi: <https://doi.org/10.1093/nar/gkt887>.
- [45] Liu X, Huang Y, Liang J, Zhang S, Li Y, Wang J, et al. Computational prediction of protein interactions related to the invasion of erythrocytes by malarial parasites. *BMC Bioinf* 2014;15:393. doi: <https://doi.org/10.1186/s12859-014-0393-z>.
- [46] Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, et al. ProtTrans: toward understanding the language of life through self-supervised learning. *IEEE Trans Pattern Anal Mach Intell* 2022;44:7112–27. doi: <https://doi.org/10.1109/TPAMI.2021.3095381>.
- [47] Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* 2019;35:4168–9. doi: <https://doi.org/10.1093/bioinformatics/btz184>.
- [48] Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* 2016;32:309–11. doi: <https://doi.org/10.1093/bioinformatics/btv557>.
- [49] Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, et al. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* 2021;49:W431–7. doi: <https://doi.org/10.1093/nar/gkab314>.
- [50] Rose AS, Hildebrand PW. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res* 2015;43:W576–9. doi: <https://doi.org/10.1093/nar/gkv402>.
- [51] Richardson E, Trevizani R, Greenbaum JA, Carter H, Nielsen M, Peters B. The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns* 2024;5:100994. doi: <https://doi.org/10.1016/j.patter.2024.100994>.
- [52] Newman MEJ. The structure and function of complex networks. *SIAM Rev* 2003;45:167–256. doi: <https://doi.org/10.1137/S003614450342480>.
- [53] Yamashita A. Crystal structure of CapZ: structural basis for actin filament barbed end capping. *EMBO J* 2003;22:1529–38. doi: <https://doi.org/10.1093/emboj/cdg167>.
- [54] Kovar DR, Wu J-Q, Pollard TD. Profilin-mediated competition between capping protein and formin Cdc12p during cytokinesis in fission yeast. *MBio* 2005;16:2313–24. doi: <https://doi.org/10.1091/mbc.e04-09-0781>.
- [55] Reynolds MJ, Hachicho C, Carl AG, Gong R, Alushin GM. Bending forces and nucleotide state jointly regulate F-actin structure. *Nature* 2022;611:380–6. doi: <https://doi.org/10.1038/s41586-022-05366-w>.
- [56] Hu Z-L, 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. doi: <https://doi.org/10.1093/nar/gkab1116>.
- [57] Honkatukia M, Reese K, Preisinger R, Tuiskulahaavisto M, Weigend S, Roito J, et al. Fishy taint in chicken eggs is associated with a substitution within a conserved motif of the gene. *Genomics* 2005;86:225–32. doi: <https://doi.org/10.1016/j.ygeno.2005.04.005>.
- [58] Youngworth I, Delany ME. A premature stop codon in RAF1 is the priority candidate causative mutation of the inherited chicken wingless-2 developmental syndrome. *Genes* 2019;10:353. doi: <https://doi.org/10.3390/genes10050353>.
- [59] Abasht B, Sandford E, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Extent and consistency of linkage disequilibrium and identification of DNA markers for production and egg quality traits in commercial layer chicken populations. *BMC Genomics* 2009;10:S2. doi: <https://doi.org/10.1186/1471-2164-10-S2-S2>.
- [60] Orand T, Jensen MR. Binding mechanisms of intrinsically disordered proteins: insights from experimental studies and structural predictions. *Curr Opin Struct Biol* 2025;90:102958. doi: <https://doi.org/10.1016/j.sbi.2024.102958>.
- [61] Omid A, Möller MH, Malhis N, Bui JM, Gsponer J. AlphaFold-Multimer accurately captures interactions and dynamics of intrinsically disordered protein regions. *Proc Natl Acad Sci* 2024;121:e2406407121. doi: <https://doi.org/10.1073/pnas.2406407121>.