

AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations

Wen-Kang Shen^{1,†}, Si-Yi Chen^{1,2,†}, Zi-Quan Gan¹, Yu-Zhu Zhang¹, Tao Yue¹, Miao-Miao Chen¹, Yu Xue¹, Hui Hu^{1,3,*} and An-Yuan Guo^{1,3,*}

¹Center for Artificial Intelligence Biology, Hubei Bioinformatics & Molecular Imaging Key Laboratory, Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China, ²Institute of Hematology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China and ³Department of Laboratory Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Received September 06, 2022; Revised September 30, 2022; Editorial Decision October 02, 2022; Accepted October 05, 2022

ABSTRACT

Transcription factors (TFs) are proteins that interact with specific DNA sequences to regulate gene expression and play crucial roles in all kinds of biological processes. To keep up with new data and provide a more comprehensive resource for TF research, we updated the Animal Transcription Factor Database (AnimalTFDB) to version 4.0 (<http://bioinfo.life.hust.edu.cn/AnimalTFDB4/>) with up-to-date data and functions. We refined the TF family rules and prediction pipeline to predict TFs in genome-wide protein sequences from Ensembl. As a result, we predicted 274 633 TF genes and 150 726 transcription cofactor genes in AnimalTFDB 4.0 in 183 animal genomes, which are 86 more species than AnimalTFDB 3.0. Besides double data volume, we also added the following new annotations and functions to the database: (i) variations (including mutations) on TF genes in various human cancers and other diseases; (ii) predicted post-translational modification sites (including phosphorylation, acetylation, methylation and ubiquitination sites) on TFs in 8 species; (iii) TF regulation in autophagy; (iv) comprehensive TF expression annotation for 38 species; (v) exact and batch search functions allow users to search AnimalTFDB flexibly. AnimalTFDB 4.0 is a useful resource for studying TF and transcription regulation, which contains comprehensive annotation and classification of TFs and transcription cofactors.

INTRODUCTION

Transcription factors (TFs) are proteins with DNA-binding domains (DBDs) that recognize specific DNA sequences to regulate gene expression and affect almost all biological processes (1). Accurate identification and comprehensive annotation for TFs are key prerequisites and basis for studying TF functions and gene expression regulation. In response to the demand of systematical identification and annotation of TFs, several dedicated TF databases have been developed. For example, PlantTFDB (2,3) is the most comprehensive plant TF database, which has identified and well-annotated TFs for 165 plant species. For animal TF databases, The Human Transcription Factors database (1) and REGULATOR (4) contain TF information for a single genome and 82 metazoan species, respectively. There are also some disease-related TFs databases, including DBTFLC (5) and BC-TFdb (6), that identified TFs associated with lung cancer and breast cancer, respectively. Among all the TF databases, Animal Transcription Factor Database (AnimalTFDB) is the most comprehensive animal TF database including classification and annotation of genome-wide TFs and transcription cofactors (TcoFs). We constructed the AnimalTFDB in 2011 (7), and updated it in 2015 (8) and 2019 (9) with more species, annotations and functions. AnimalTFDB has become an essential resource for studying animal TFs and regulations, accessed by millions and cited over 650 times.

As one of the essential regulator types in various biological processes, TFs have been studied in many areas, including gene expression regulation (10), conservation or evolution (11), genetics or diseases (12), TF regulatory networks (13) and TF target prediction (14). Recently, the

*To whom correspondence should be addressed. Tel: +86 27 8779 3177; Fax: +86 27 8779 3177; Email: guoay@hust.edu.cn
Correspondence may also be addressed to Hui Hu. Email: huhui@hust.edu.cn

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

functions of TF in autophagy, post-translational modifications (PTMs) and variations in diseases have been determined, enhancing our understanding of the biological process influenced by TFs. Variations associated with TFs could frequently result in impaired transcriptional activation of TFs (15). ClinVar (16) and COSMIC (17) databases identified and integrated human variations in cancers and other diseases, which are vital resources for studying TF-related variations. In the past four years, the number of species in Ensembl (18) database has approximately doubled. Thus, we upgraded AnimalTFDB to version 4.0 with the latest genomes, annotation data and new functions. Compared with previous versions, AnimalTFDB 4.0 covers more species, more TFs and TcoFs with updated annotation data. In addition, we not only integrated the TF-related variations (mutations), gene expression, PTMs and autophagy information, but also provided the exact and batch search functions. The new AnimalTFDB 4.0 will be a helpful resource for transcriptional regulation and comparative genomics research.

DATA SOURCE

In AnimalTFDB 4.0, we downloaded all the protein sequences of 183 animal genomes from Ensembl (version 105). These 183 species are classified into 12 catalogs according to their taxonomy, namely ‘Afrotheria’, ‘Amphibians’, ‘Birds&Reptiles’, ‘Fishes’, ‘Laurasiatheria’, ‘Other Chordates’, ‘Other Eukaryotes’, ‘Other Mammals’, ‘Other Vertebrates’, ‘Primates’, ‘Rodents’ and ‘Xenarthra’ (Supplementary Table S1). Compared with AnimalTFDB 3.0, catalogs with a large increase in the number of species are ‘Fishes’ (from 11 to 59 species) and ‘Birds&Reptiles’ (from 7 to 24 species).

We collected a large quantity of annotations from the NCBI Entrez Gene (19) and Ensembl (18) databases, including basic gene information, homologous genes, gene phenotypes and Gene Ontology (GO). We acquired genome-wide association studies (GWAS) phenotypes from the latest GWAS Catalog (20) and dbSNP (release 155) (19). Protein-protein interaction (PPI) data were gathered from BioGRID (version 4.4) (21) and HPRD (22), and the protein functional domains were predicted by PfamScan for all protein domain models in the Pfam (version 35.0) database (23). We obtained the signaling pathway information from BioCarta and KEGG (24) databases. Furthermore, transcription factor binding site (TFBS) information was extracted from HOCOMOCO (25), TRANSFAC (26), JASPAR (27) and CIS-BP (28) databases.

Next, we integrated 190 627 and 8 294 851 variation records for TFs and TcoFs from ClinVar (16) and COSMIC (v96) (17), respectively. PTM information of TFs and TcoFs for eight species was obtained from CPLM (29) and EPSD (30), containing 131 378 phosphorylation sites and 38 943 lysine modification sites (including acetylation, methylation and ubiquitination). In addition, we accessed information from THANATOS (31) on whether a TF or TcoF is involved in regulating autophagy-related processes (autophagy, apoptosis, and necrosis). Moreover, TF expression data for 38 species were collected from TCGA (32), EMBL-

Table 1. The comparisons on data volume, annotation and tool between version 3.0 and 4.0 of AnimalTFDB database

AnimalTFDB		Version 3.0	Version 4.0
Data volume	Species	97	183
	TF families	73	73
	TF genes	125 135	274 633
	Cofactor genes	80 060	150 726
	Cofactor families	83	82
Annotations	Species with expression data	22	38
	Phenotype	Yes	Yes
	DBDs WebLogo	Yes	Yes
	TFBS	Yes	Yes
	GWAS	Yes	Yes
	Variations	No	Yes
	Autophagy	No	Yes
	PTM	No	Yes
Tools	TF prediction server	Yes	Yes
	BLAST search	Yes	Yes
	PPI network	Yes	Yes
	TFBS prediction server	Yes	Yes
	Exact and batch search	No	Yes

EBI Expression Atlas (33), GTEx (34), Bgee (Version 15.0) (35), FANTOM5 (36), the Human Protein Map (37), the Human Protein Atlas (38) and some articles with large-scale expression data (39–41). In AnimalTFDB 4.0, the number and types of data are more comprehensive than in previous version (Table 1).

IMPROVED TF/COFACTOR FAMILY RULES AND PREDICTION PIPELINE

The classification and assignment rules for animal TF families

TFs are usually classified into different families according to their conserved DBDs. In AnimalTFDB 4.0, we classified TFs into 73 families and six categories, which are consistent with AnimalTFDB 3.0. We optimized the TFs family assignment rules in AnimalTFDB 3.0 by manually checking the results in human and mouse genomes. We made several rules to classify a TF into its correct family. First, we classified the TFs based on the family-specific domain when a superfamily has several families. For example, for the zf-C2H2 superfamily, we classified proteins with both zf.C2H2 and ZBTB domains into the ZBTB family and proteins containing only zf.C2H2 domain into the zf.C2H2 family. The second rule is that if a TF has several different DBDs, it is assigned to the family with the smallest E value of DBD. The third rule is removing enzyme proteins since they were annotated as enzymes although they contain some DBDs.

Optimized TF and cofactors prediction pipelines

We built the TF prediction pipeline based on the TF family and classification rules. Firstly, we downloaded the Hidden Markov Model (HMM) profiles for DBDs of 58 TF families from the latest Pfam database (version 35.0) (23). Secondly, we reconstructed the remaining 14 TF families based on the DBD sequences from classical species (human, mouse, zebrafish, and fly) with HMMER (v3.1b2) (42). Thirdly,

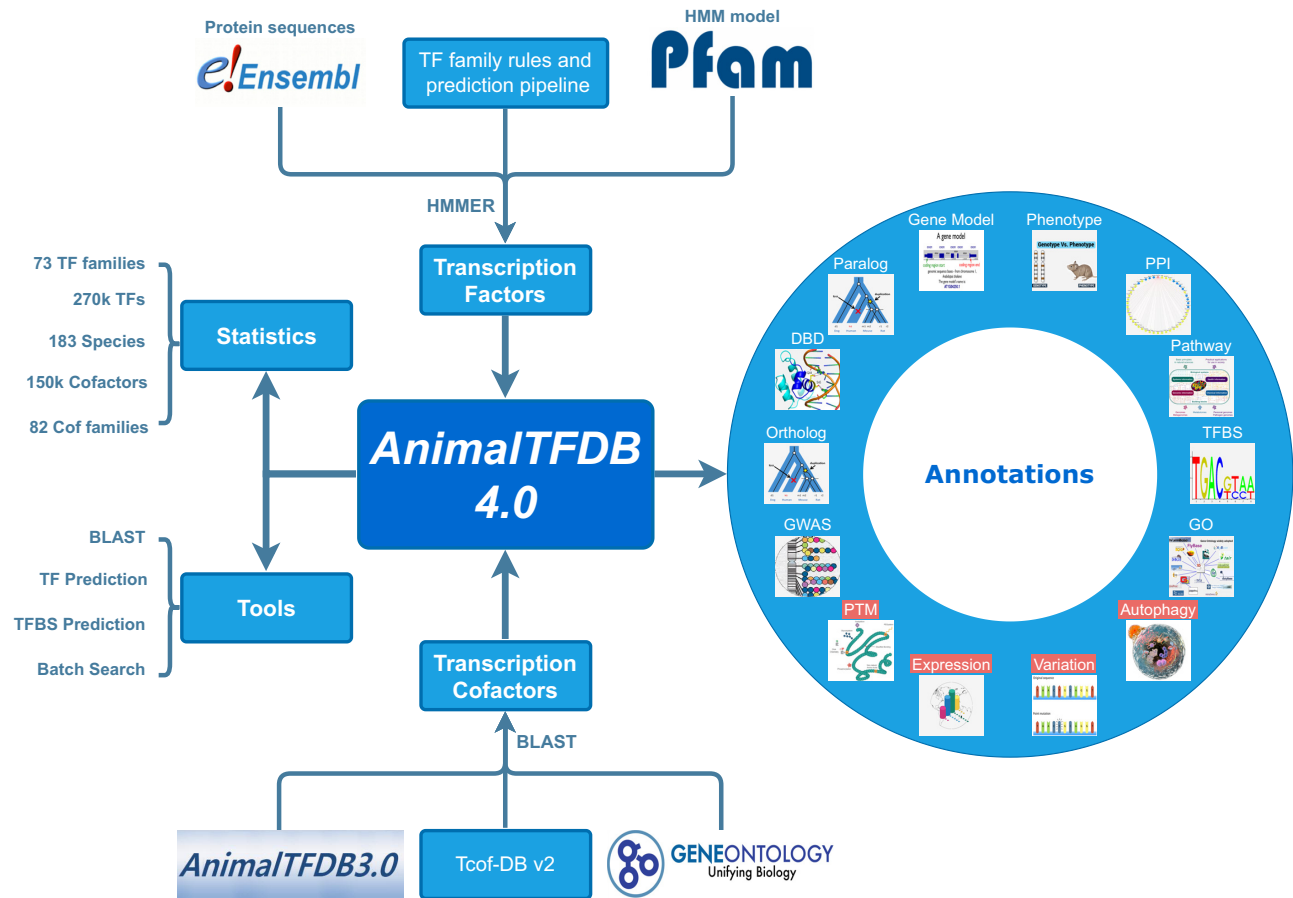


Figure 1. The workflow, data and annotation summaries of AnimalTFDB 4.0. The top of the figure illustrates the TFs prediction workflow, and the bottom of the figure describes the TcoFs prediction workflow. The left of the figure shows the data statistics and tools in AnimalTFDB 4.0. The right of the figure shows the multiple types of annotations in AnimalTFDB 4.0, of which variations, gene expression, PTMs and autophagy are newly added.

we used the *hmmsearch* program in the HMMER package to search all protein sequences of each species against the DBD HMM profiles to predict TFs (Figure 1). The E-value thresholds for each family were shown in Supplementary Table S2. Since the DBD HMM profiles and protein sequences were updated, we adjusted the E-value thresholds used in AnimalTFDB 3.0 for some families by manually checking human prediction results. For instance, the previous E-value threshold of Fork head domain was $1e-4$ and we adjusted it to $1e-3$ because FOXO3B and FOXD3 can only be predicted using E-value $1e-3$. Similarly, we adjusted the E-value threshold of Homeobox domain to $1e-2$. Furthermore, we examined the 'Others' family and removed some proteins that are not TFs, such as centromere proteins (CENPA, CENPS, CENPT, CENPX).

As a result, in AnimalTFDB 4.0, we predicted a total of 1659 human TFs, which is consistent with 99.33% TFs in the previous version. The deleted TFs were SMARCA1, CCDC88A, ZBED5 and centromere protein genes, which have no relevant evidence to prove that they are TFs. The added TFs are NFILZ, FOXO3B, ZNF738 etc., all of which have DNA-binding TF activity in publications or GeneCards database annotation (43). We also compared the human TFs in AnimalTFDB 4.0 with those in The Hu-

man Transcription Factors Database (1) and found that among the 1639 TFs in The Human Transcription Factors Database, 1556 (94.93%) of them (1499 TFs and 57 TcoFs) are in our AnimalTFDB 4.0. The remaining 83 genes (5.06%) were commented on their website as 'Likely to be sequence specific TF' or without literature evidence. However, the majority of the unique 160 TFs in AnimalTFDB 4.0 were explicit TFs, including transcriptional activators (HSFX3, HSFX4, SMAD2, SMAD6, SMAD7, UBTF, TCF19, TCF25 etc.) and repressors (LRRFIP1, LRRFIP2, MIER1, MIER3, ID1/2/3/4 etc.). These comparisons provide good evidence for the high accuracy of our TF prediction results.

For TcoFs, we collected 1024 human TcoFs from AnimalTFDB 3.0, Tcof-DB v2 database (44) and GO database based on the related GO terms ('transcription coactivator activity', 'transcription corepressor activity', 'transcription cofactor activity', 'regulation of transcription', 'chromatin remodeling', 'chromatin-mediated maintenance of transcription', 'histone *ylation', 'histone *ylase activity' and 'histone *transferase activity'). TcoFs in the other 182 species were identified by performing reciprocal best-hit BLAST between each of them and human with E-value $\leq 1e-4$, identity $\geq 30\%$ and coverage $\geq 50\%$ (Figure 1).

DATA SUMMARY AND OVERALL FUNCTIONS

In AnimalTFDB 4.0, we identified 274 633 TFs and 150 726 TcoFs in 183 animal species (Figure 1). The numbers of TFs and TcoFs for each species are shown in Supplementary Table S3, and there are 1659 TFs (8.30% in protein-coding genes) and 1024 TcoFs (5.12%) in human. The data demonstrate that the number of TF family ranged between 58 and 73, with most species (95.08%) having at least 70 TF families. Meanwhile, the number of TcoF families ranged from 56 to 82, and 181 species have at least 74 TcoF families. To provide comprehensive information for TFs and TcoFs, we collected various types of annotations, including ‘Ortholog’, ‘Paralog’, ‘DBD’, ‘GWAS’, ‘Gene Model’, ‘Gene Phenotype’, ‘PPI’, ‘Pathway’, ‘TFBS’ and ‘GO’. In AnimalTFDB 4.0, some new annotations were added, which include variations, gene expression, post-translational modifications and autophagy regulation information (Figure 1).

Users can browse AnimalTFDB 4.0 by species to obtain the list of TFs and TcoFs for individual species. Users can also browse by family to acquire the distribution of gene counts across species for a single family. Besides the data browsing function, we provided several functional tools in AnimalTFDB, including TF prediction, TFBS prediction, BLAST and batch search. The TF prediction function allows users to identify potential TFs in their uploaded protein sequences using the prediction pipelines in AnimalTFDB 4.0 (Figure 1). The TFBS prediction function can predict potential TFBS in user uploaded DNA sequences using the TFBS MEME files from HOCOMOCO (25), TRANSFAC (26), JASPAR (27), CIS-BP (28) and hTFtarget (14) (Figure 1). In BLAST function, TF protein sequences of all species or a specific species could be selected to perform BLAST search (Figure 1). For user's convenience, we added the batch search function allowing users to search AnimalTFDB 4.0 more flexibly (Figure 1).

NEW ANNOTATIONS FOR TFS AND COFACTORS

Variations

TFs act as essential regulators in the transcription process and their variations may change their protein sequences and functions, leading to transcriptional dysregulation. We collected the variation information from the latest ClinVar (16) and COSMIC (v96) (17) databases, which aggregated human health and disease related genomic variations. There are 898 (54.12%) human TFs and 706 (68.94%) TcoFs having a total of 190 627 variations in ClinVar, with an average of 118 records per gene. For each ClinVar variation record in TF or TcoF, the variation ID, variation type, position and clinical significance were shown in Figure 2A. In addition, almost all human TFs (98.25%) and TcoFs (99.21%) could find variants records in COSMIC, with a total of 8 294 851 COSMIC variants, all of which are somatic mutations in 119 human cancers. Among these data, about half mutations occurred in the coding sequence (CDS) and half in the 3'UTR, 5'UTR or intron regions. Mutations occurring in the CDS region were classified into 14 types, of which the major mutation types are missense (69.47%), coding silent (synonymous) (16.27%), and nonsense (6.36%), all of which belong to substitutions (Supplementary Table S4).

Since there are so many mutations per gene, we displayed the related COSMIC information and provided a filtering function in AnimalTFDB 4.0 (Figure 2B). Users can filter the primary site, primary histology and description according to their needs. Moreover, users can clear all filters by clicking the clear button and export results by clicking the export button (Figure 2B).

Gene expression

In AnimalTFDB 4.0, we provided gene expression information of TFs and TcoFs for 38 species, doubling the data volume from previous version. These expression data were from normal tissues, cell lines, different stages and cancers in human and other species. As a summary, expression data were available for 81.48–100% TFs and 89.34–100% TcoFs in 38 species (Supplementary Table S5). Expression data are classified into mRNA and protein expression, both of them are available for humans, while only mRNA expression data are available for other 37 species. Besides gene expression data in previous version, we collected more large-scale gene expression datasets for human, including gene expression in 54 non-diseased tissue sites from GTEx project (34), 60 tissues from FANTOM5 project (36), and The Human Protein Atlas (38) based on RNA-seq of gene expression in 256 tissues and 69 cell lines (Supplementary Table S6). In total, we collected gene expression data in different tissues, cell lines and cancers from eight datasets for human (Figure 2C). In addition, we collected large-scale gene expression data from some articles for more species, such as gene expression in seven organs across different developmental stages for 7 mammal species (39) and gene expression for nine species in different developmental stages and tissues (40) (Supplementary Table S6).

Autophagy regulation information

Autophagy is a complex and vital process that controls cellular remodeling and quality control (45), and many TFs are involved in regulating autophagy (46). We gathered information about whether a TF or TcoF is involved in regulating autophagy-related processes from the THANATOS (31) database. Autophagy-related processes include autophagy, apoptosis, and necrosis. The regulation modes include positive regulation, negative regulation or both. In total, we collected 1,023 autophagy records from six model organisms (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Caenorhabditis elegans*, *Danio rerio* and *Drosophila melanogaster*) involving 338 TFs and 382 TcoFs. The autophagy regulation information for each TF or TcoF was shown in Figure 2D.

Post-translational modifications

Post-translational modifications influence the functional regulation of TFs and their co-regulators (47). Growing evidence shows that the PTMs of TFs have positive and negative consequences on transcription (48). Here, we parsed 38 943 lysine modification sites (including 14 041 acetylation, 1169 methylation and 23 733 ubiquitination) from CPLM (29) database and 131 378 phosphorylation sites

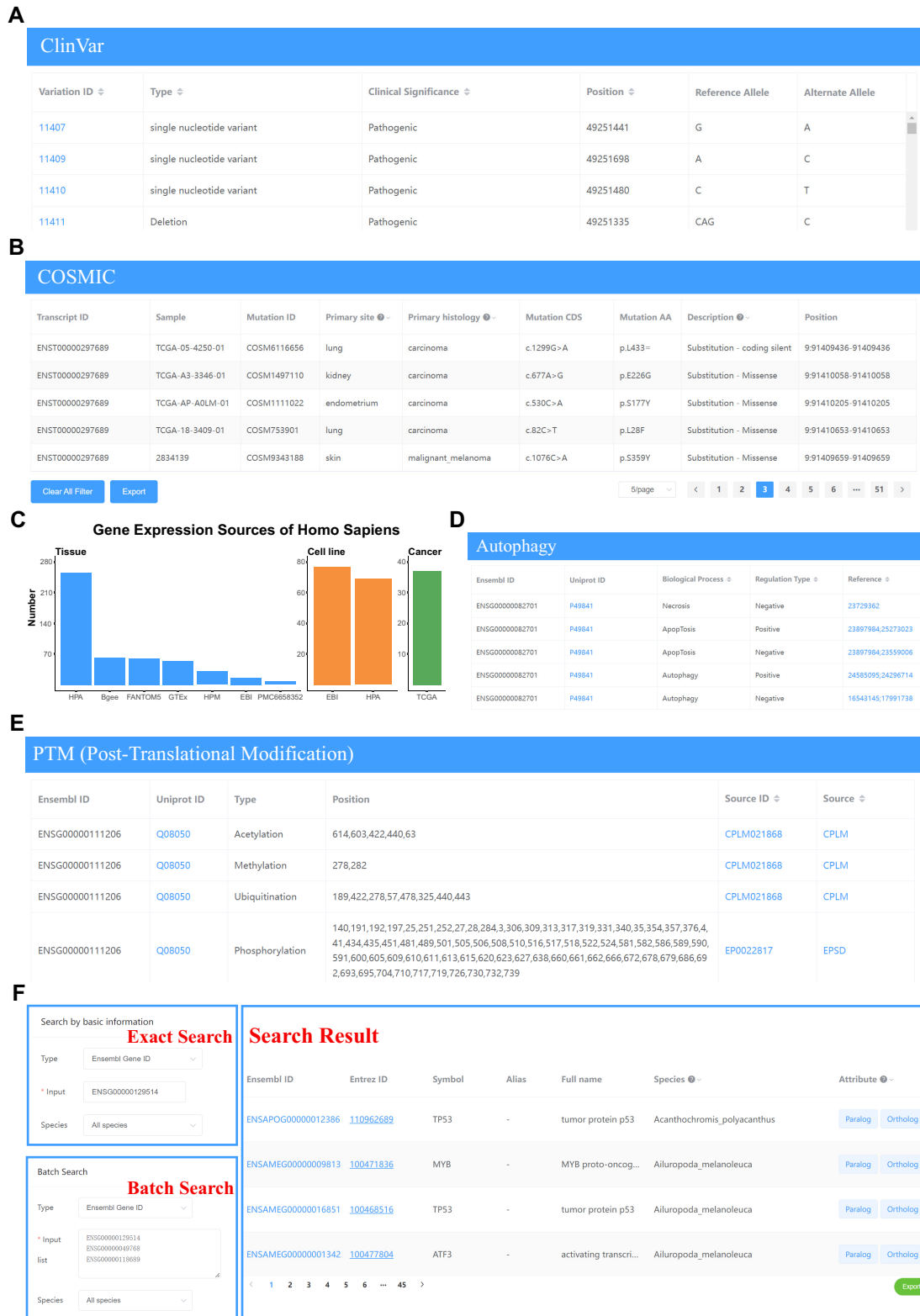


Figure 2. New features of AnimalTFDB 4.0. (A) The ClinVar variations of human TFs and TcoFs. (B) The COSMIC mutations of human TFs and TcoFs. (C) The gene expression sources of *Homo Sapiens*. The X-axis is the gene expression datasets and the Y-axis is the number of tissues, cell lines and cancer types. (D) The autophagy related information. (E) The post-translational modification information. (F) The exact and batch search function.

from EPSD (30) database in eight model species (*H. sapiens*, *M. musculus*, *R. norvegicus*, *C. elegans*, *B. taurus*, *Cavia porcellus*, *Gallus* and *D. melanogaster*). There are 2941 TFs and 2343 TcoFs with PTM information containing 1588 human TFs, 1013 human TcoFs, 980 mouse TFs and 835 mouse TcoFs, as well as 373 TFs and 494 TcoFs in the remaining six species (Supplementary Table S7). For each TF or TcoF, the positions of PTMs were shown in Figure 2E.

Exact and batch search function

Besides the TF/TFBS prediction and BLAST functions on AnimalTFDB, for user's convenience, we added the exact search and batch search function allowing users to search the database flexibly. Users can search by entering the gene ID, transcript ID, protein ID or gene name or gene alias to find matched results in database (Figure 2F). Users can also make a batch search by inputting a list of genes (gene ID, transcript ID, protein ID and gene name are accepted) (Figure 2F). The batch search can guide downstream analysis in some situations. For example, after identifying the differentially expressed genes (DEGs) between two groups, users can put all DEGs as the input of 'Batch search' to determine which of them are TFs and explore their functions. By exact search or batch search, users can obtain gene ID, gene name, species and attributes (types of gene annotation information). Users can filter species and attributes and export results by clicking the export button (Figure 2F).

SUMMARY AND FUTURE PERSPECTIVES

With the increasing sequenced and well annotated animal genomes, we updated AnimalTFDB to version 4.0 with new features. AnimalTFDB 4.0 provides 274 633 TFs and 150 726 TcoFs from 183 animal genomes. In addition, we added several new annotations including TF-related variations, gene expression, PTMs and autophagy information, as well as new functions such as the exact and batch search. Variation information of human TFs from ClinVar and COSMIC will provide useful resources for researchers to further explore the TF mutations and their associated diseases. More comprehensive TF expression information will help users better understand the relationship between TF and development as well as the diversity of TF expression. The PTM site and autophagy regulation information will be essential resources for studying the function and regulation of TF. The exact search and batch search functions will allow users to search AnimalTFDB flexibly. In summary, we believe these improvements will make AnimalTFDB more comprehensive and useful. Undoubtedly, the genomic data for various species will continue to grow. We will continue regularly updating the AnimalTFDB database to make it as a core resource for TF regulation.

DATA AVAILABILITY

AnimalTFDB 4.0 is publicly accessible for worldwide users without any registration or login. Users can freely access all data in AnimalTFDB 4.0 at <http://bioinfo.life.hust.edu.cn/AnimalTFDB4/>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank colleagues in data production and database construction in groups of Ensembl, dbSNP, TCGA, GTEx, FANTOM5, ClinVar, COSMIC, CPLM, EPSD, THANATOS, TRANSFAC, JASPAR, Cis-BP, HO-COMOCO, HMMER, Pfam and hTFtarget. We are also grateful to our users and all members in our lab for their valuable suggestions and comments.

The computing work in this paper is supported by the Public Service Platform of High Performance Computing by Network and Computing Center of HUST.

FUNDING

National Key R&D Program of China [2021YFF0703704]; Natural Science Foundation for Distinguished Young Scholars of Hubei Province of China [2020CFA070]; National Natural Science Foundation of China [32100527]. Funding for open access charge: National Key R&D Program of China [2021YFF0703704].

Conflict of interest statement. None declared.

REFERENCES

- Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. and Weirauch, M.T. (2018) The human transcription factors. *Cell*, **172**, 650–665.
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J. and Gao, G. (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.*, **45**, D1040–D1045.
- Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J. and Gao, G. (2019) PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.*, **48**, D1104–D1113.
- Wang, K. and Nishida, H. (2015) REGULATOR: a database of metazoan transcription factors and maternal factors for developmental studies. *BMC Bioinf.*, **16**, 114.
- Amalraj, T., Dravid, A.A., Tripathi, R. and Lulu, S.S. (2018) Database of transcription factors in lung cancer (DBTFLC): a novel resource for exploring transcription factors associated with lung cancer. *J. Cell. Biochem.*, **119**, 5253–5261.
- Khan, A., Khan, T., Nasir, S.N., Ali, S.S., Suleman, M., Rizwan, M., Waseem, M., Ali, S., Zhao, X. and Wei, D.-Q. (2021) BC-TFdb: a database of transcription factor drivers in breast cancer. *Database*, **2021**, baab018.
- Zhang, H.-M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H. and Guo, A.-Y. (2012) AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.*, **40**, D144–D149.
- Zhang, H.-M., Liu, T., Liu, C.-J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y. and Guo, A.-Y. (2015) AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.*, **43**, D76–D81.
- Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q. and Guo, A.-Y. (2019) AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.*, **47**, D33–D38.
- Hu, H., Zhang, Q., Hu, F.-F., Liu, C.-J. and Guo, A.-Y. (2021) A comprehensive survey for human transcription factors on expression, regulation, interaction, phenotype and cancer survival. *Brief. Bioinform.*, **22**, bbab002.
- Nitta, K.R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E.E.M. *et al.* (2015) Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife*, **4**, e04837.

12. Barrera, L.A., Vedenko, A., Kurland, J.V., Rogers, J.M., Gisselbrecht, S.S., Rossin, E.J., Woodard, J., Mariani, L., Kock, K.H., Inukai, S. *et al.* (2016) Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science*, **351**, 1450–1454.
13. Xie, G.-Y., Xia, M., Miao, Y.-R., Luo, M., Zhang, Q. and Guo, A.-Y. (2020) FFLtool: a web server for transcription factor and miRNA feed forward loop analysis in human. *Bioinformatics*, **36**, 2605–2607.
14. Zhang, Q., Liu, W., Zhang, H.-M., Xie, G.-Y., Miao, Y.-R., Xia, M. and Guo, A.-Y. (2020) hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. *Genomics Proteomics Bioinformatics*, **18**, 120–128.
15. Hyde, R.K. and Liu, P.P. (2013) Germline PAX5 mutations and b cell leukemia. *Nat. Genet.*, **45**, 1104–1105.
16. Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C. *et al.* (2020) ClinVar: improvements to accessing data. *Nucleic Acids Res.*, **48**, D835–D844.
17. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E. *et al.* (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **47**, D941–D947.
18. Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhairi, J. *et al.* (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.
19. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S. *et al.* (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **50**, D20–D26.
20. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Solis, E. *et al.* (2019) The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
21. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F. *et al.* (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.*, **30**, 187–200.
22. Goel, R., Harsha, H.C., Pandey, A. and Prasad, T.S.K. (2012) Human protein reference database and human proteinpedia as resources for phosphoproteome analysis. *Mol. Biosyst.*, **8**, 453–463.
23. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
24. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. and Tanabe, M. (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.*, **49**, D545–D551.
25. Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A. *et al.* (2018) HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-Seq analysis. *Nucleic Acids Res.*, **46**, D252–D259.
26. Wingender, E. (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.*, **9**, 326–332.
27. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D. *et al.* (2019) JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **48**, D87–D92.
28. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K. *et al.* (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, **158**, 1431–1443.
29. Zhang, W., Tan, X., Lin, S., Gou, Y., Han, C., Zhang, C., Ning, W., Wang, C. and Xue, Y. (2022) CPLM 4.0: an updated database with rich annotations for protein lysine modifications. *Nucleic Acids Res.*, **50**, D451–D459.
30. Lin, S., Wang, C., Zhou, J., Shi, Y., Ruan, C., Tu, Y., Yao, L., Peng, D. and Xue, Y. (2021) EPSD: a well-annotated data resource of protein phosphorylation sites in eukaryotes. *Brief. Bioinform.*, **22**, 298–307.
31. Deng, W., Ma, L., Zhang, Y., Zhou, J., Wang, Y., Liu, Z. and Xue, Y. (2018) THANATOS: an integrative data resource of proteins and post-translational modifications in the regulation of autophagy. *Autophagy*, **14**, 296–310.
32. Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkol.*, **1A**, 68–77.
33. Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A.M.-P., George, N., Fexova, S., Fonseca, N.A., Füllgrabe, A., Green, M., Huang, N. *et al.* (2019) Expression atlas update: from tissues to single cells. *Nucleic Acids Res.*, **48**, D77–D83.
34. THE GTEx CONSORTIUM (2020) The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
35. Bastian, F.B., Roux, J., Niknejad, A., Comte, A., Fonseca Costa, S.S., de Farias, T.M., Moretti, S., Parmentier, G., de Laval, V.R., Rosikiewicz, M. *et al.* (2021) The bgee suite: integrated curated expression atlas and comparative transcriptomics in animals. *Nucleic Acids Res.*, **49**, D831–D847.
36. the FANTOM consortium, Lizio, M., Harshbarger, J., Shiojiri, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F. *et al.* (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.*, **16**, 22.
37. Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S. *et al.* (2014) A draft map of the human proteome. *Nature*, **509**, 575–581.
38. Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F. *et al.* (2020) An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, **367**, eaay5947.
39. Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascensão, K., Rummel, C., Ovchinnikova, S. *et al.* (2019) Gene expression across mammalian organ development. *Nature*, **571**, 505–509.
40. Jin, W., Zhu, Q., Yang, Y., Yang, W., Wang, D., Yang, J., Niu, X., Yu, D. and Gong, J. (2021) Animal-APAdb: a comprehensive animal alternative polyadenylation database. *Nucleic Acids Res.*, **49**, D47–D54.
41. Li, J.J., Huang, H., Bickel, P.J. and Brenner, S.E. (2014) Comparison of *d. melanogaster* and *c. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res.*, **24**, 1086–1101.
42. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R. and Finn, R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.*, **46**, W200–W204.
43. Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T.I., Dahary, D., Fishilevich, S. and Lancet, D. (2021) The genecards suite. In: Abugessaisa, I. and Kasukawa, T. (eds). *Practical Guide to Life Science Databases*. Springer Nature, Singapore, pp. 27–56.
44. Schmeier, S., Alam, T., Essack, M. and Bajic, V.B. (2017) TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. *Nucleic Acids Res.*, **45**, D145–D150.
45. Cheng, Z. (2019) The foxo–autophagy axis in health and disease. *Trends Endocrinol. Metab.*, **30**, 658–671.
46. Zhou, H., Wang, X., Ma, L., Deng, A., Wang, S. and Chen, X. (2019) FoxO3 transcription factor promotes autophagy after transient cerebral ischemia/reperfusion. *Int. J. Neurosci.*, **129**, 738–745.
47. Yu, E.J., Kim, S.-H., Kim, M.J., Seo, W.-Y., Song, K.-A., Kang, M.-S., Yang, C.K., Stallcup, M.R. and Kim, J.H. (2013) SUMOylation of ZFP282 potentiates its positive effect on estrogen signaling in breast tumorigenesis. *Oncogene*, **32**, 4160–4168.
48. He, A., Shen, X., Ma, Q., Cao, J., von Gise, A., Zhou, P., Wang, G., Marquez, V.E., Orkin, S.H. and Pu, W.T. (2012) PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes Dev.*, **26**, 37–42.