

TcoFBase: a comprehensive database for decoding the regulatory transcription co-factors in human and mouse

Yuexin Zhang^{1,2,†}, Chao Song^{1,2,†}, Yimeng Zhang^{2,3,†}, Yuezhu Wang^{1,2,†}, Chenchen Feng², Jiaxin Chen², Ling Wei², Qi Pan², Desi Shang^{1,4,5,6}, Yanbing Zhu⁷, Jiang Zhu², Shuangfang Fang⁸, Jun Zhao², Yongsan Yang², Xilong Zhao², Xiaozheng Xu², Qiuyu Wang^{1,2,4,5,6,*}, Jincheng Guo^{8,*} and Chunquan Li^{1,2,4,5,6,*}

¹The First Affiliated Hospital, Institute of Cardiovascular Disease, Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China, ²School of Medical Informatics, Daqing Campus, Harbin Medical University, Daqing 163319, China, ³The Key Laboratory of Molecular Biology for High Cancer Incidence Coastal Chaoshan Area, Shantou University Medical College, Shantou 515041, China, ⁴School of Computer, University of South China, Hengyang, Hunan 421001, China, ⁵The First Affiliated Hospital, Cardiovascular Lab of Big Data and Imaging Artificial Intelligence, Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China, ⁶Hunan Provincial Base for Scientific and Technological Innovation Cooperation, University of South China, Hengyang, Hunan 421001, China, ⁷Experimental and Translational Research Center, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China and ⁸Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China

Received August 14, 2021; Revised September 21, 2021; Editorial Decision September 30, 2021; Accepted October 04, 2021

ABSTRACT

Transcription co-factors (TcoFs) play crucial roles in gene expression regulation by communicating regulatory cues from enhancers to promoters. With the rapid accumulation of TcoF associated chromatin immunoprecipitation sequencing (ChIP-seq) data, the comprehensive collection and integrative analyses of these data are urgently required. Here, we developed the TcoFBase database (<http://tcof.liclab.net/TcoFbase>), which aimed to document a large number of available resources for mammalian TcoFs and provided annotations and enrichment analyses of TcoFs. TcoFBase curated 2322 TcoFs and 6759 TcoFs associated ChIP-seq data from over 500 tissues/cell types in human and mouse. Importantly, TcoFBase provided detailed and abundant (epi) genetic annotations of ChIP-seq based TcoF binding regions. Furthermore, TcoFBase supported regulatory annotation information and various functional annotations for TcoFs. Meanwhile, TcoFBase embedded five types of TcoF regulatory analyses for users, including TcoF gene set enrichment, TcoF binding ge-

omic region annotation, TcoF regulatory network analysis, TcoF-TF co-occupancy analysis and TcoF regulatory axis analysis. TcoFBase was designed to be a useful resource that will help reveal the potential biological effects of TcoFs and elucidate TcoF-related regulatory mechanisms.

INTRODUCTION

The transcriptional regulation program that establishes and maintains specific cell states in mammals is extremely complicated (1). Studies have revealed that major regulators of the transcriptional regulation program are transcription factors (TFs), transcription co-factors (TcoFs) and chromatin regulators. TFs typically bind in a cooperative fashion to distal DNA elements to regulate gene expression by recruiting interactive TcoFs and forming physical contact loops (2). Recent studies have also provided new insights into TcoFs, which are crucial for DNA loop maintenance and proper gene control. TcoFs play a decisive role in gene expression through interaction with TFs, although they do not have DNA-binding properties of their own. Increasing evidence has demonstrated that dysfunction of TcoFs can lead to aberrant gene expression and cause a broad range of

*To whom correspondence should be addressed. Tel: +86 15004591078; Fax: +86 0734 8279018; Email: lcqbio@163.com
Correspondence may also be addressed to Qiuyu Wang. Tel: +86 13351294769; Fax: +86 0734 8279018; Email: wangqiuyu900490@163.com
Correspondence may also be addressed to Jincheng Guo. Tel: +86 1062600822; Fax: +86 1062601356; Email: obgene@qq.com

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

diseases, suggesting that TcoFs may serve as potential therapeutic targets (3–8). Thus, it is urgent to investigate and unveil TcoF-mediated regulatory mechanisms. Importantly, as a type of transcriptional regulator, protein activities of TcoFs can affect the expression of downstream target genes by indirectly occupying DNA regulatory elements, such as promoter regions, enhancer regions and super-enhancer regions. These specific occupancies have an important function in maintaining cell identity. For example, the TcoF Bromodomain Containing 4 (BRD4) recruits many enhancers and super-enhancers, while its inhibition can suppress the transcription level of super-enhancers and enhancers associated genes (9,10). In addition, some studies have also found that upstream signaling perturbations further influence TcoFs and alter the expression levels of downstream target genes. These studies indicated the importance and widespread utility of TcoF-mediated regulatory axes for addressing key issues associated with physiological and pathological processes.

Previous studies have developed several TcoF-related databases to investigate the biological content of TcoFs, which have provided valuable information for decoding the biological functions of TcoFs. For example, TcoFDB (11) stores a list of TcoFs in mammals and supplies information needed to explore the roles of TcoFs. Some TF-related databases also provide useful data for TcoFs, such as AnimalTFDB (12), and ApicoTFdb (13), which offer basic annotation information for TcoFs. Meanwhile, hTF-target (14) is a comprehensive resource which focused on predicting TcoF–target interactions from a huge set of TcoF ChIP-seq data. However, until now, a comprehensive database to depict comprehensive regulatory elements of TcoFs has not been available. Given the development of high-throughput techniques, chromatin immunoprecipitation sequencing (ChIP-seq) has become an important strategy for identifying the target genes and functions of TcoFs. With the rapid accumulation of TcoF-associated ChIP-seq data, the comprehensive collection and integrative analysis of these data has become an urgent need. More importantly, integrative annotations and analyses will be very useful for elucidating the mechanisms of TcoFs which underlie transcriptional regulation. Therefore, it is highly desirable to construct a comprehensive resource of TcoFs, which provides the extensive annotations of TcoFs and associated regulatory analyses.

To address this needs, here we developed the TcoFBase database (<http://tcof.liclab.net/TcoFbase>), which aimed to document a large number of available resources for mammals TcoFs, and provided extensive annotations and analyses of TcoFs. The current version of TcoFBase curated a total of 2322 TcoFs and 6759 TcoFs associated ChIP-seq data from over 500 tissues and cell types in humans and mice. Emphatically, TcoFBase provided the detailed and abundant (epi) genetic annotations in ChIP-seq-based TcoF binding regions by processing and integrating large scale high throughput experimental data and low throughput experimental data, including super-enhancers, enhancers, transcription factor binding sites (TFBS), methylation sites, common single nucleotide polymorphisms (SNPs), risk SNPs, expression quantitative trait locus (eQTL) and 3D chromatin interactions. TcoFBase

also provided TcoF downstream target genes by mapping binding regions onto genomes using five methods. Furthermore, TcoFBase supported upstream regulatory information for TcoFs and various functional annotations, including pathways, gene ontology (GO) terms, cancer hallmarks, protein-protein interactions (PPIs), survival, expression and disease information. In particular, TcoFBase embedded five types of TcoF regulatory analyses for users, including TcoF gene set enrichment, TcoF binding genomic region annotation, TcoF regulatory network analysis, TcoF-TF co-occupancy analysis and TcoF regulatory axis analysis. TcoFBase was also a user friendly database to query, browse, analyze and visualize information associated with TcoFs (Figure 1). We believe that TcoFBase may become a useful and effective platform for exploring potential functions and regulation of TcoFs in diseases and biological processes.

MATERIALS AND METHODS

TcoF datasets

A list of TcoFs in mammals was collected from TcoF-DB v2 (11) and AnimalTFDB 3.0 (12) (Figure 1 middle-top panel). Meanwhile, we collected 6759 TcoF-related ChIP-seq datasets in different human and mouse cell and tissue types for human and mouse from ReMap (15), ENCODE (16), Cistrome (17), and ChIP-Atlas (18) (Figure 1, middle-top panel). To standardize the format and genome version, genomic peaks in all ChIP-seq data were converted to the hg19 (human) and mm10 (mouse) genomic versions using the liftOver tool of UCSC (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) (19). Additionally, we also provided online liftOver tools to convert genome coordinates for users. Users can substitute hg38 assembly information for hg19 genomic regions and mm10 assembly information for mm39 genomic regions.

Annotations of TcoF regulatory regions

Super-enhancers and enhancers. To annotate the potential roles of TcoF-associated regions, we obtained enhancer data from EnhancerAtlas (20), FAMTOM5 (21), ENCODE (16), HANCER (22), DENdb (23) and ENdb (24), including 14 797 266 human enhancers and 439 092 mouse enhancers (Figure 1 upper right panel and Supplementary Table S1). Some databases identified enhancers by high-throughput experimental data, such as FAMTOM5 (21) and ENCODE (16). Moreover, experimentally validated enhancers were downloaded from ENdb, which were manually curated based on low-throughput experimental data. Moreover, we collected H3K27ac ChIP-seq data from Roadmap, Genomics of Gene Regulation Project, NCBI GEO/SRA and ENCODE (25), which contained both H3K27ac ChIP-seq and the corresponding input controls. We ran Bowtie (26) for each H3K27ac ChIP-seq dataset and all peaks were called using MACS (27). ROSE (10) was used to identify super-enhancer regions. In addition, we also downloaded human and mouse super-enhancers from SEA (28) and dbSuper (29) (Figure 1, upper right panel and Supplementary Table S1). In total, TcoFBase curated

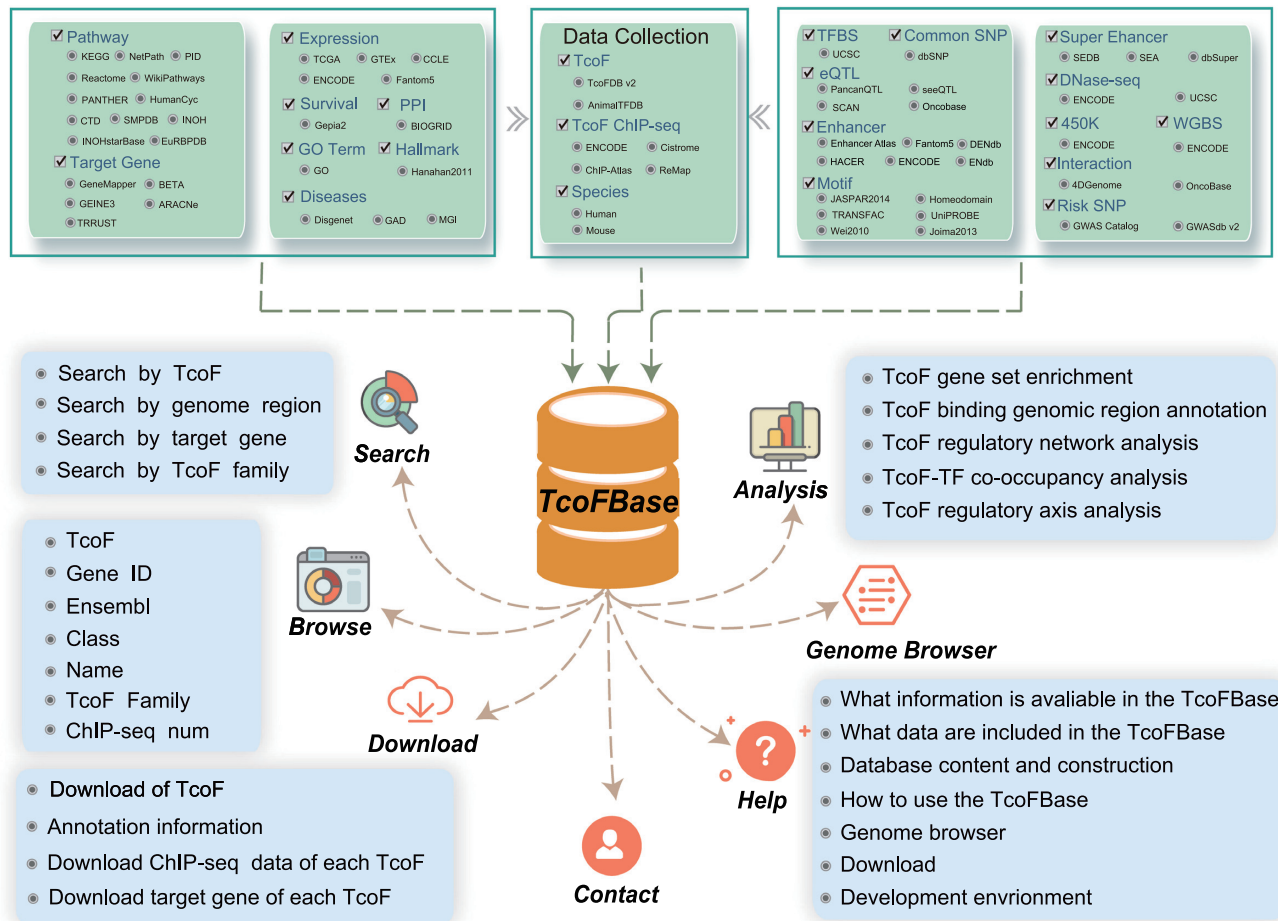


Figure 1. Database content and construction. TcoFBase has a large number of TcoF-related ChIP-seq datasets and abundant curated multi-omics annotations for TcoFs. TcoFBase also contains analytical tools and multiple functions to browse, search, download and visualize TcoFs.

2 678 273 human super-enhancers and 11,609 mouse super-enhancers.

TFBS. Studies have revealed that TcoFs contribute to gene transcription by interacting with TFs rather than by directly binding DNA elements. To investigate the cooperative relationships between TFs and TcoFs, we first used Find Individual Motif Occurrences (FIMO) software (30) to identify the motif occurrences in TcoF binding regions. Over 3000 DNA binding motifs for ~700 mammalian TFs were collected from Jolma2013 (31), JASPAR CORE 2014 vertebrates (32), Homeodomains (33), UniPROBE (34) and Wei2010 (35) (Figure 1 upper right panel and Supplementary Table S1). Motif-based TF binding sites within TcoF binding regions were identified with a P -value threshold of $1e-6$. Second, we obtained 5 547 656 human TFBS and 2 858 356 mouse TFBS from UCSC (19) and used BEDTools (36) to identify the intersected TF peaks within TcoFs peaks as TFBS (Figure 1 upper right panel and Supplementary Table S1).

SNP/Linkage disequilibrium (LD) SNPs/risk SNP/eQTL. To annotate the effects of SNPs in TcoF regulatory regions, we downloaded 38 063 729 human common SNPs from db-

SNP (37) and used VCFTools v0.1.13 (38) to calculate SNPs with a minimum allelic frequency (MAF) > 0.05 (Figure 1 upper right panel and Supplementary Table S1). Plink v1.9 (39) was used to identify the LD SNPs ($r^2 = 0.8$) in five super-populations (East Asian, African, European, Ad Mixed American and South Asian). A total of 264 514 human risk SNPs were downloaded from the GWAS Catalog (40) and GWASdb v2 (41) and 2 886 133 human eQTLs were collected from PanCanQTL (42), seeQTL (43), SCAN (44) and Oncobase (45) (Figure 1, upper right panel and Supplementary Table S1).

Chromatin interaction/ DNase I hypersensitive signals (DHSs). Chromatin interaction data (e.g. ChIA-PET 3C, 4C, 5C and Hi-C) are useful for understanding the precise regulatory mechanism between TcoFs and downstream genes. We obtained these data from 4DGenome (46) and Oncobase (45) (Figure 1 upper right panel and Supplementary Table S1). DHS annotation data of TcoF regulatory regions were collected from UCSC (19) and ENCODE (16) for 69 860 705 human DHSs from 293 samples and 9 802 229 mouse DHSs from 56 samples (Figure 1, upper right panel and Supplementary Table S1).

Methylation. Methylation data such as 450K arrays and whole-genome shotgun bisulfite sequencing (WGBS) are useful to understand the regulatory relationships between the regulatory regions and TcoFs. We downloaded 30 392 523 methylation sites from 450k arrays and 166 855 665 WGBS datasets from ENCODE (16) (Figure 1, upper right panel and Supplementary Table S1).

TcoFs target genes and TcoF–target network

We provided five different methods to identify the downstream target genes of TcoFs. First, we used a python script from ROSE (ROSE geneMapper.py) (10) to annotate TcoF downstream target genes in each human and mouse TcoF ChIP-seq dataset (Figure 1, upper left panel). Notably, target genes obtained using three strategies of ROSE (10) (overlap, proximal and closest) were combined and collected. Second, binding and expression target analysis (BETA) (47) was used to identify TcoF downstream genes (Figure 1, upper left panel). BETA is a software package that integrates ChIP-seq data of chromatin regulators to infer direct target genes. BETA minus (47) was used here to identify target genes by inputting bed format files of human and mouse TcoF ChIP-seq datasets. Third, we performed GENIE3 (48) to identify human TcoF target genes by inputting The Cancer Genome Atlas (TCGA) cancer expression matrices (mouse TcoFs target genes were identified by inputting FANTOM5 mouse cell expression matrices) (Figure 1, upper left panel). GENIE3 (48) is a random forest-based approach to predict the strength of putative regulatory links between target genes and their putative regulators. Fourth, we utilized the mutual information-based software ARACNe (49) to identify human TcoF target genes via TCGA expression data (Figure 1, upper left panel). Finally, we identified mammalian TcoF target genes by collecting curated TcoF–target gene pairs from TRRUST v2 (50) (Figure 1, upper left panel). Importantly, TRRUST v2 provided literature-curated TcoF–target pairs in humans and mice. All of the TcoF target genes were embedded in the current version of TcoFBase. We also defined the target gene weight for each TcoF based on calculating the sum of prediction methods. Users can perform hypergeometric test between genes of interest and TcoF target genes to identify the upstream TcoFs regulating the genes. The enrichment significance *P*-value for the particular TcoF was calculated as:

$$P = 1 - \sum_{i=0}^{x-1} \frac{\binom{k}{i} \binom{n-k}{s-i}}{\binom{n}{s}}$$

where *n* is the number of TcoF target genes and *k* is the number of input genes, of which *x* genes are involved in the TcoF containing *s* target genes.

Importantly, we constructed a human TcoF–target network by merging the TcoF–target genes from the above five methods. The mouse TcoF–target network was constructed by merging the TcoF–target genes from the above four methods except for ARACNe. To ensure the high credibility of the network, we reserved TcoF–target pairs that were supported by more than half of the methods. For human, we reserved the TcoF–gene pairs that were identified

in more than three methods. For mice, TcoF–gene pairs that were identified in more than two methods were reserved. Based on this network, users can locate the sub-network of interest to find the TcoF mediated regulatory axes.

Functional annotations of TcoFs

To better illustrate the biological functions of TcoFs, we collected more annotation information of TcoFs, including TcoF associated pathways, gene ontologies, hallmarks, survival, expression and diseases from multiple sources. Briefly, we collected expression matrices with the FPKM value of human TcoFs from TCGA (51), GTEx (52), CCLE (53) and ENCODE (16) (Figure 1, upper left panel). Moreover, TcoFBase provides the TPM value of mouse TcoFs from FANTOM5 (54) and the FPKM value of mouse TcoFs from ENCODE (Figure 1, upper left panel). The experimentally supported mammalian (human and mouse) TcoF–disease associations were collected from DisGeNET (55), GAD (56) and MGI (57) (Figure 1, upper left panel). Meanwhile, we obtained PPIs from BIOGRID (58). The detailed information for 2,169 human and mouse pathways were integrated from a database of ten pathways, including KEGG, Reactome, PANTHER, SMPDB, NetPath, PID, HumanCyc, CTD, WikiPathways, INOHstarBase v2.0 and EuRBPDB (59–61) (Figure 1, upper left panel). Users can select at least one pathway database when using the online analysis tool of ‘TcoF regulatory axis analysis’. In addition, we obtained ten cancer hallmarks from Hanahan and Weinberg (62) and 31 GO terms from GO (63) (Figure 1, upper left panel). We used a python package from GEPIA2 to obtain TCGA cancer survival maps of TcoFs (Figure 1, upper left panel).

DATABASE USE AND ACCESS

A search interface for retrieving TcoFs

TcoFBase is a powerful platform that enables users to search, browse, analyze, visualize and download TcoFs of interest (Figure 1, bottom panel). On the ‘Search’ page, TcoFBase provides four query methods for searching TcoF information, including ‘Search by TcoF’ (input a TcoF), ‘Search by genomic region’ (input a genomic region), ‘Search by target gene’ (input a gene official symbol) and ‘Search by TcoF family’ (input a TcoF family name; Figure 2A and B). In the TcoF query, users can input a TcoF, and TcoFBase will return the detailed page of the input TcoF. As a result, users can obtain the details of searched TcoFs, including TcoF Overview, TcoF target gene network, ChIP-seq based regulatory details, downstream target genes, upstream regulatory details, function annotation, PPI, survival, expression and disease information (Figure 2C). In detail, in the table of downstream target genes, TcoFBase provides the downstream target genes of TcoF through five different prediction methods (ROSE, BETA, GENIE3, ARACNe and TRRUST). For upstream regulatory details, the table describes TcoF upstream regulatory regions, including promoter region, enhancer region and super-enhancer region, and the detailed (epi) genetic annotations within regulatory regions are also listed. Users can click ‘Sample ID’ for details about each ChIP-seq sample.

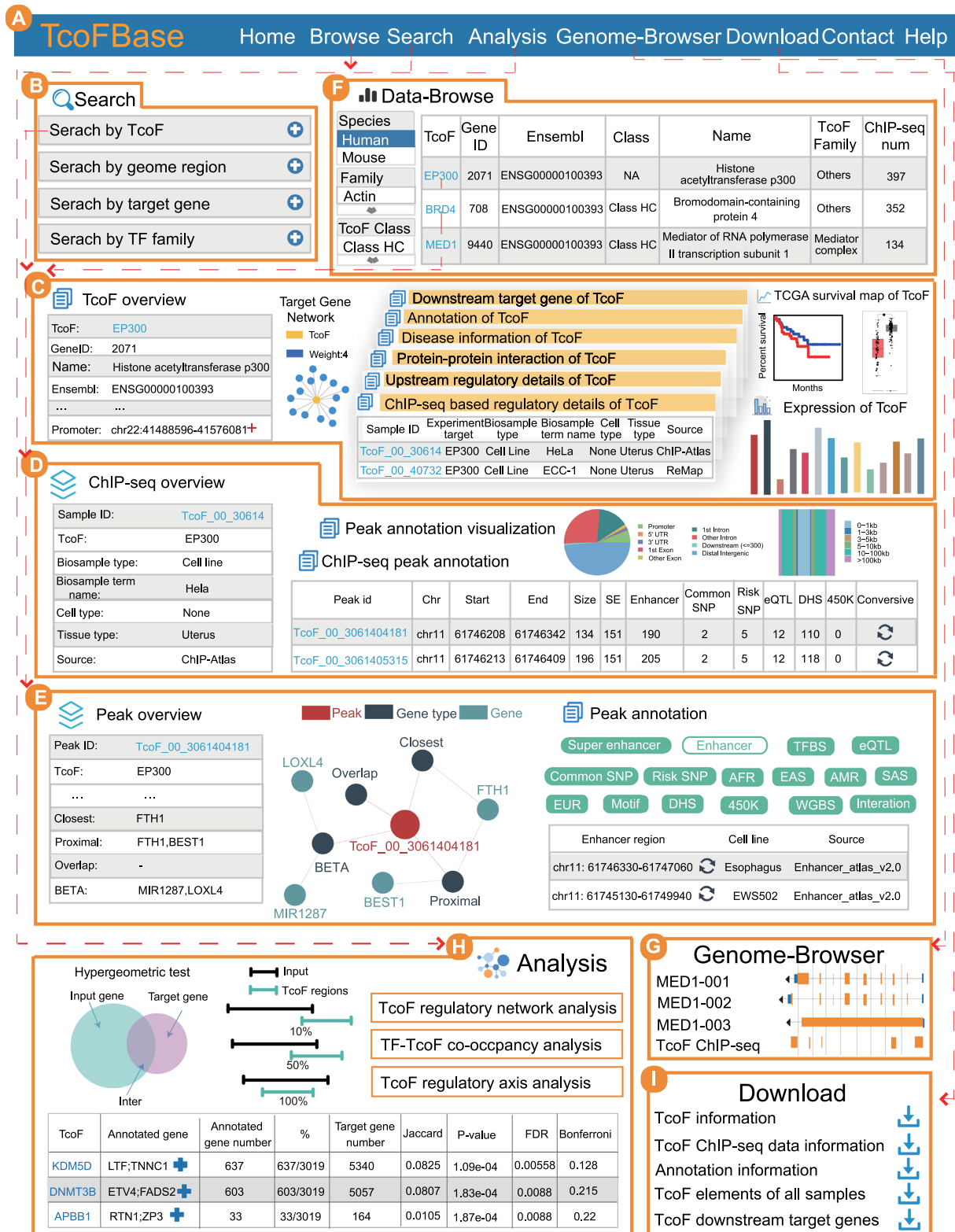


Figure 2. Main functions and usage of TcoFBase. (A) The navigation bar of TcoFBase. (B) Four inquiry modes are available. (C) Search results including TcoF overview, detailed interactive tables of ChIP-seq information, upstream regulatory details, PPIs, disease information, annotation, expression, downstream target genes and TCGA cancer survival maps. (D) Interactive table with detailed information about ChIP-seq samples of interest and visualization of peak annotation. (E) Detailed interactive tables of annotation information and ChIP-seq peak associated genes are identified through four strategies. Network diagram about these regions is displayed. (F) Browsing TcoFs. (G) Genome browser. (H) An online analysis tools for TcoF. (I) Data download.

TcoFBase displays ChIP-seq overview, ChIP-seq peak annotation and peak annotation visualization. For ChIP-seq peak annotation, the interactive table describes the Peak ID, genome location, size and detailed (epi) genetic information in TcoF ChIP-seq regions (Figure 2D). TcoFBase provides genomic annotations and the distribution of TcoF-binding regions relative to TSS to visualize peaks using ChIPseeker. Moreover, users can click ‘Peak ID’ to obtain detailed information about TcoF ChIP-seq regions. On the peak details page, TcoFBase displays a peak overview, statistics charts and the peak-gene network. The peak-gene network is provided through BETA and ROSE. In ‘Peak annotation’, TcoFBase lists detailed (epi) genetic annotation information, including super-enhancers, enhancers, common SNPs, risk SNPs, eQTLs, DHSs, LD SNPs, TFs predicted by motifs, TFBS, 3D chromatin interactions, methylation sites of the 450k array and WGBS. Meanwhile, we provided the hg38 and mm39 assembly version of each genomic region. By clicking the conversion button, users can find the corresponding hg38 or mm39 versions of these regions. In addition, TcoFBase provides Pearson’s correlation coefficients between TcoFs and their target genes in different TCGA cancers (Figure 2E).

In the ‘Search by genomic region’ query, with the input of a genomic region TcoFBase will identify TcoF ChIP-seq peak regions that overlap the submitted region and output a results table. Users can click ‘Peak ID’ to obtain the detailed information about the peak. In the ‘Search by target gene’ query, users can submit a gene of interest and select the target gene prediction methods (ROSE, BETA, GENIE3, ARACNe and TRRUST); TcoFBase will extract its upstream TcoFs in the result table. Users can also select a TcoF family in the ‘Search by TcoF family’ query. TcoFBase can display TcoFs of this family on the results page.

User-friendly browsing and genome browser

The ‘Browse’ page is organized as an interactive table that allows users to quickly search for TcoFs. Users can use the sorting function and customize filters (‘Species’, ‘Family’ and ‘TcoF Class’) to quickly search TcoFs. To further view the detailed information for TcoFs, users can click the ‘TcoF’ button (Figure 2F). TcoFBase uses the JBrowse genome browser to visualize the regulatory information of human TcoFs by customizing tracks of interest, such as TcoF ChIP-seq regions, gene, enhancer, super-enhancer and DHS (Figure 2G).

Online analysis tools

We designed five types of analyses to elucidate the mechanistic studies of TcoFs underlying transcriptional regulation (Figure 2H): (I) *TcoF gene set enrichment*. Users can input a gene set of interest, choose the species (human or mouse) and set the *P*-value/FDR to perform TcoF gene set enrichment. TcoFBase will identify upstream TcoF regulators that are significantly enriched by hypergeometric test in the highly credible TcoF–target network. (II) *TcoF binding genomic region annotation*. In this analysis, input of genomic region(s) in bed format will identify the TcoFs that can bind to input genomic region(s). (III) *TcoF regulatory*

network analysis. Users can submit a gene list to extract a TcoF-mediated regulatory network from the highly credible TcoF–target network. The TcoF-mediated regulatory network consists of submitted genes and their regulated TcoFs. (IV) *TcoF–TF co-occupancy analysis*. Users can submit a TcoF–TF pair or individual TcoF/TF(s) of interest and choose the species (human or mouse) to identify the TcoF–TF pair(s) with the same DNA binding loci. Users can set different statistical thresholds to control for false positivity by the ‘FIMO’ option. (V) *TcoF regulatory axis analysis*. Users can input gene(s) of interest and users select at least one pathway database. TcoFBase will then identify significantly enriched pathways in which terminal downstream genes are TcoFs.

Data download

TcoFBase supported download of TcoF-associated files, including ‘TcoF information’, ‘TcoF ChIP-seq data information’, ‘TcoF ChIP-seq samples of different genome versions’ and ‘TcoF downstream target genes’ (Figure 2I). In addition, the database supported the packaged downloads of all TcoF ChIP-seq samples and TcoF downstream target genes. Meanwhile, we also provided download of annotation information, such as PPI, pathway, disease, super-enhancer, enhancer, common SNP, risk SNP, eQTL and DHS. TcoFBase supported export for all data tables in TcoF detail pages, search results and analysis results.

Case study

Case study of BRD4. To illustrate how to use TcoFBase, we used TcoF ‘BRD4’ as input for ‘Search by TcoF’. BRD4 has been demonstrated to play a crucial role in transcriptional regulation during tumorigenesis and embryogenesis. On the results page, users first obtain the overview information for BRD4, including basic gene annotations, upstream regulatory elements (promoter, enhancer and super-enhancer) and the target gene network (Figure 3A). Here, we also provide the detailed weights of the target genes in the table of ‘Downstream target genes’. Intriguingly, some known target genes of BRD4 such as NANOG, MYOD1 and POU5F1 were identified in this search as highly convincing targets with high weights (64–66), suggesting the potential of using TcoFBase for the identification of TcoF target genes (Figure 3B).

Second, TcoFBase provides upstream regulatory information for BRD4, annotation of BRD4 and disease information for BRD4. In the upstream regulatory aspect, TcoFBase displays abundant (epi) genetics annotations of TcoF upstream regulatory regions. For instance, the TFBS of two TFs, GATA1 and SRF, were found to bind the upstream super-enhancer regions of BRD4 and regulate the expression of BRD4. GATA1 and SRF were previously validated as upstream regulators of BRD4 (Figure 3C upper left table) (67). Importantly, TcoFBase has the strong ability to uncover downstream regulatory clues for BRD4 via processing of BRD4-related ChIP-seq data. Users can click ‘Sample ID’ to obtain the detailed annotations of BRD4 ChIP-seq data (Figure 3D). Notably, as a result of the peak annotation visualization of BRD4 ChIP-seq data, we found

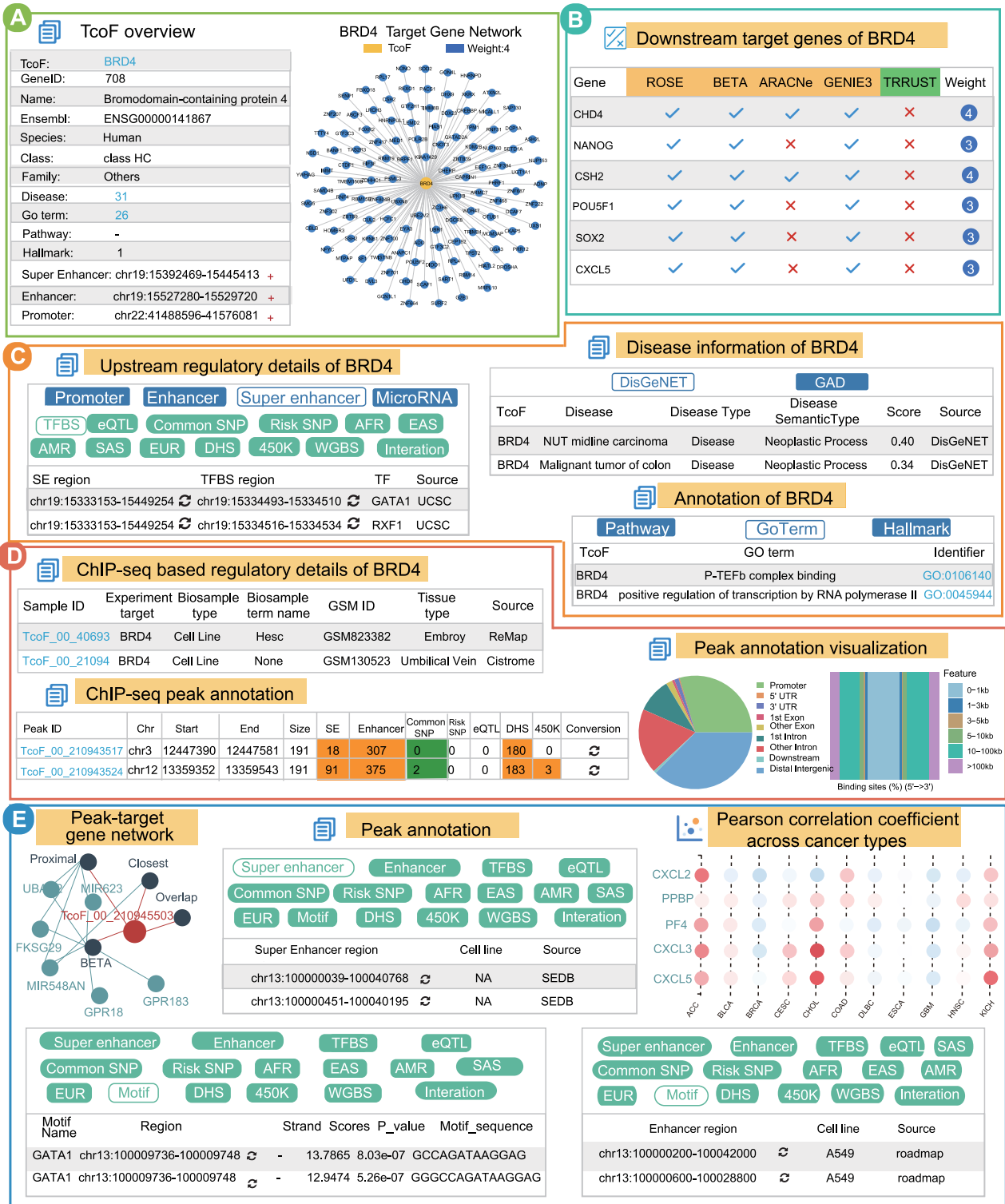


Figure 3. Validation results associated with BRD4. (A) BRD4 overview and BRD4 target network. (B) Downstream target genes of BRD4. (C) Upstream regulatory information, disease and functional annotation of BRD4. (D) ChIP-seq information for BRD4, detailed information and visualization of peak annotation about TcoF_00_21094. (E) Peak-target gene network, detailed interactive table of annotation information (super-enhancer, motif and enhancer) of TcoF_00_210943517 and Pearson's correlation coefficients between BRD4 and downstream target genes.

Table 1 . Comparison of TcoF associated information in TcoFBase with other databases

| Parameter | TcoFBase | TcoFDBv2 | AnimalTFDB3.0 | hTFtarget |
|-------------------------------|-----------------|----------|---------------|----------------|
| TcoF (human and mouse) number | 2322 | 1386 | 1995 | Less than 200 |
| TcoF ChIP-seq sample number | 6759 | – | – | Less than 1000 |
| Region annotation | ✓ | – | – | – |
| Downstream target genes | ✓ | – | – | ✓ |
| Peak annotation visualization | ✓ | – | – | – |
| Genome browser | ✓ | ✓ | – | ✓ |
| Online analysis | ✓ | – | – | – |
| GO Term | ✓ | ✓ | ✓ | – |
| Cancer hallmark | ✓ | – | – | – |
| Survival | ✓ | – | – | – |
| PPI | ✓ | ✓ | ✓ | – |
| Pathway sources | 10 ^a | 2 | 2 | – |
| Expression sources | 5 ^b | 2 | 5 | – |
| Disease information sources | 3 ^c | 3 | – | – |

^aPathway were collected from ten resources: KEGG, Reactome, PANTHER, SMPDB, NetPath, PID, HumanCyc, CTD, WikiPathways, INOHstarBase v2.0 and EuRPDB.

^bExpression matrixes were collected from five resources: ENCODE, GTEX, CCLE, NCBI and FANTOM5.

^cDisease were collected from three resources: DisGeNET, GAD and MGI.

that more than half of the BRD4 binding peaks were located in distal regions (>10 kb), which coincides with the regulatory role of BRD4 in super-enhancer mediated transcriptional regulation (Figure 3D right) (68).

Finally, TcoFBase also displays detailed information for each peak of BRD4, such as downstream target genes, downstream regulatory information and Pearson correlation coefficients between BRD4 and downstream target genes across TCGA cancers. For example, CXCL5 was identified as the target gene of BRD4 peak ‘TcoF_00_210943517’ (Figure 3E, upper left). In the peak annotations of ‘TcoF_00_210943517’, we found that SOX2 was occupied at this BRD4 peak, which implied that BRD4 might regulate CXCL5 expression by cooperating with SOX2 (Figure 3E, bottom left). Actually, previous studies have demonstrated the cooperative fashion in which SOX2 and BRD4 regulated CXCL5 expression in glioblastoma (69). BRD4 was considered to be a key regulator in super-enhancer/enhancer formation. TcoFBase also identified a subset of super-enhancers and enhancers that overlapped with peak ‘TcoF_00_210943517’, suggesting that the regulatory mechanisms between BRD4 and CXCL5 were mediated by distal regulatory elements (Figure 3E, middle and bottom right). These results indicated that TcoFBase could be used to explore the potential regulatory mechanisms for TcoFs.

Case study of online analysis. In addition, we also provided two case studies of online analysis to verify the function of TcoFBase. We utilized differentially expressed genes of breast invasive carcinoma (BRCA) as inputs for ‘TcoF gene set enrichment’ analysis and the output table of significantly enriched TcoFs was showed in Supplementary Table S2. As a result, EGFR was identified as the top enriched TcoF from the enrichment analysis results, which was demonstrated to play key roles in breast carcinogenesis (Supplementary Materials and Supplementary Figure S1). We also performed the same tests for heart failure in mouse and found that Tnni2 was significantly enriched, which is critical for heart failure (Supplementary Figure S2 and Supplementary Table S3). These results demonstrated

the availability and biological value of TcoFBase for TcoF research.

DISCUSSION

TcoFs have been demonstrated to participate in transcriptional regulation by interacting with TFs and maintaining the DNA contact loops in mammals. Thoroughly revealing TcoF-mediated transcription regulation networks is crucial to understand the molecular mechanisms underlying physiological and pathological processes. Some databases, such as TcoFDB (11), AnimalTFDB (12), hTFtarget (14) and ApicoTFdb (13), have curated the basic knowledge, annotations and predicted target for a large number of TcoFs. However, a comprehensive regulatory database for TcoFs has not been built. Meanwhile, the rapid release of TcoF-associated ChIP-seq data has enabled us to find the target genes of TcoFs and construct a TcoF related global regulatory network. Based on the urgent need and available regulatory data, we developed TcoFBase to document mammalian TcoFs, and provide annotations and enrichment analyses of TcoFs. The current version of TcoFBase has incorporated a total of 2322 TcoFs and 6759 TcoF-associated ChIP-seq data from over 500 tissues and cell types in humans and mice. Table 1 compares TcoFBase with previous TcoF-related databases for TcoF number, TcoF ChIP-seq number and datasets source. As a result, TcoFBase is advanced in multiple aspects, such as TcoF ChIP-seq data scale and abundant (epi) genetic annotations in TcoF binding regions (Table 1 and Supplementary Table S1). Moreover, TcoFBase also supports upstream regulatory information and multiple functional annotations for TcoFs. Overall, TcoFBase is an efficient resource for investigating the regulatory cues of TcoFs.

As far as we know, TcoFBase is the first integrative resource to focus on unveiling the regulatory mechanisms of TcoFs. Researchers can explore regulatory information about TcoFs and perform TcoF-associated regulatory analyses of interest. TcoFBase has multiple advantages, including (i) comprehensive genetic and epigenetic annotations in TcoF-associated ChIP-seq regions and user-friendly dis-

plays on the details page; (ii) identification of TcoF downstream target genes by five methods; (iii) integrative annotation information about TcoF upstream regulatory regions; (iv) useful online analysis tools; (v) four search methods to access TcoFs; (vi) user-friendly browsing and (vii) full-featured detail pages containing disease information, survival and different functional annotations of TcoFs.

The current version of TcoFBase processed a large number of TcoF-related ChIP-seq datasets and curated abundant multi-omics annotations of TcoFs. We compared TcoFBase with other TcoF-associated databases for information and functions, which showed the advantages of TcoFBase. However, we did not use the same methods to detect peaks of TcoFs and we also found that the regulatory information for some TcoFs was not supported by ChIP-seq due to lack of data. In future updates, we will collect ChIP-seq samples of TcoFs to identify TcoF regions. Furthermore, we will extend the scale of the regulatory datasets and add more annotation information. We will also embed more precise methods to infer TcoF-mediated regulatory networks and explore the functions and regulation of TcoFs. In summary, TcoFBase is a user-friendly database to query, browse, analyze and visualize information associated with TcoFs. We believe that TcoFBase will become a useful and effective platform for exploring potential functions and regulation of TcoFs in diseases and biological processes.

DATA AVAILABILITY

TcoFBase is a comprehensive database which provided annotations and enrichment analyses of TcoFs (<http://tcof.liclab.net/TcoFbase>). Detailed information is available under the Download page of TcoF.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

This work was supported by National Natural Science Foundation of China [Grant No. 62171166, 81572341, 61601150, 32000439 and 62001145]; Natural Science Foundation for Distinguished Young Scholars of Heilongjiang Province of China [Grant No. JQ2020C004]; Research Foundation of the First Affiliated Hospital of University of South China for Advanced Talents; Wu Liande Youth Science Research Fund of Harbin Medical University (Daqing) [Grant No. DQWLD201703]; China Postdoctoral Science Foundation [2019M661311]; Heilongjiang Provincial Postdoctoral Science Foundation [LBH-Z19075]; Natural Science Foundation of Heilongjiang Province [LH2021F044].

Conflict of interest statement. None declared.

REFERENCES

- Lee, T.I. and Young, R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
- Zabidi, M.A. and Stark, A. (2016) Regulatory enhancer-core-promoter communication via transcription factors and cofactors. *Trends Genet.: TIG*, **32**, 801–814.
- Tao, C.C., Hsu, W.L., Ma, Y.L., Cheng, S.J. and Lee, E.H. (2017) Epigenetic regulation of HDAC1 SUMOylation as an endogenous neuroprotection against Abeta toxicity in a mouse model of Alzheimer's disease. *Cell Death Differ.*, **24**, 597–614.
- Kim, S.Y., Zhang, X., Schiattarella, G.G., Altamirano, F., Ramos, T.A.R., French, K.M., Jiang, N., Szweda, P.A., Evers, B.M., May, H.I. *et al.* (2020) Epigenetic reader BRD4 (Bromodomain-Containing Protein 4) governs nucleus-encoded mitochondrial transcriptome to regulate cardiac function. *Circulation*, **142**, 2356–2370.
- Umbarino, D. (2017) Heart failure: BRD4 inhibition slows HF progression. *Nat. Rev. Cardiol.*, **14**, 382–383.
- Meyer, S.N., Scuppo, C., Vlasevska, S., Bal, E., Holmes, A.B., Holloman, M., Garcia-Ibanez, L., Nataraj, S., Duval, R., Vantrimont, T. *et al.* (2019) Unique and shared epigenetic programs of the CREBBP and EP300 acetyltransferases in germinal center B cells reveal targetable dependencies in lymphoma. *Immunity*, **51**, 535–547.
- Russo, J.W., Nouri, M. and Balk, S.P. (2019) androgen receptor interaction with mediator complex is enhanced in castration-resistant prostate cancer by CDK7 phosphorylation of MED1. *Cancer Discov.*, **9**, 1490–1492.
- Bai, L., Jia, Y., Viswakarma, N., Huang, J., Vluggens, A., Wolins, N.E., Jafari, N., Rao, M.S., Borensztajn, J., Yang, G. *et al.* (2011) Transcription coactivator mediator subunit MED1 is required for the development of fatty liver in the mouse. *Hepatology*, **53**, 1164–1174.
- Tasdemir, N., Banito, A., Roe, J.S., Alonso-Curbelo, D., Camiolo, M., Tschaharganeh, D.F., Huang, C.H., Aksoy, O., Bolden, J.E., Chen, C.C. *et al.* (2016) BRD4 connects enhancer remodeling to senescence immune surveillance. *Cancer Discov.*, **6**, 612–629.
- Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I. and Young, R.A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320–334.
- 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.
- 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.
- Sardar, R., Kaushik, A., Pandey, R., Mohammed, A., Ali, S. and Gupta, D. (2019) ApicoTFdb: the comprehensive web repository of apicomplexan transcription factors and transcription-associated co-factors. *Database*, **2019**, baz094.
- 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.
- Cheneby, J., Menetrier, Z., Mestdagh, M., Rosnet, T., Douida, A., Rhalloussi, W., Bergon, A., Lopez, F. and Ballester, B. (2020) ReMap 2020: a database of regulatory regions from an integrative analysis of Human and Arabidopsis DNA-binding sequencing experiments. *Nucleic Acids Res.*, **48**, D180–D188.
- Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
- Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.H., Brown, M., Zhang, X., Meyer, C.A. *et al.* (2019) Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.*, **47**, D729–D735.
- Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J. and Meno, C. (2018) ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.*, **19**, e46255.
- Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N. *et al.* (2019) The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.*, **47**, D853–D858.
- Gao, T. and Qian, J. (2020) EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.*, **48**, D58–D64.

21. Consortium, F., the, R.P., Clst, Forrest, A.R., Kawaji,H., Rehli,M., Baillie,J.K., de Hoon,M.J., Haberle,V., Lassmann,T. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462–470.
22. Wang,J., Dai,X., Berry,L.D., Cogan,J.D., Liu,Q. and Shyr,Y. (2019) HACER: an atlas of human active enhancers to interpret regulatory variants. *Nucleic Acids Res.*, **47**, D106–D112.
23. Ashoor,H., Kleftogiannis,D., Radovanovic,A. and Bajic,V.B. (2015) DENdb: database of integrated human enhancers. *Database : the journal of biological databases and curation*, **2015**, bav085.
24. Bai,X., Shi,S., Ai,B., Jiang,Y., Liu,Y., Han,X., Xu,M., Pan,Q., Wang,F., Wang,Q. *et al.* (2020) ENdb: a manually curated database of experimentally supported enhancers for human and mouse. *Nucleic Acids Res.*, **48**, D51–D57.
25. Jiang,Y., Qian,F., Bai,X., Liu,Y., Wang,Q., Ai,B., Han,X., Shi,S., Zhang,J., Li,X. *et al.* (2019) SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res.*, **47**, D235–D243.
26. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
27. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoutte,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
28. Chen,C., Zhou,D., Gu,Y., Wang,C., Zhang,M., Lin,X., Xing,J., Wang,H. and Zhang,Y. (2020) SEA version 3.0: a comprehensive extension and update of the super-enhancer archive. *Nucleic Acids Res.*, **48**, D198–D203.
29. Khan,A. and Zhang,X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, **44**, D164–D171.
30. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
31. Jolma,A., Yan,J., Whittington,T., Toivonen,J., Nitta,K.R., Rastas,P., Morgunova,E., Enge,M., Taipale,M., Wei,G. *et al.* (2013) DNA-binding specificities of human transcription factors. *Cell*, **152**, 327–339.
32. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C.Y., Chou,A., Ienasescu,H. *et al.* (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**, D142–D147.
33. Berger,M.F., Badis,G., Gehrke,A.R., Talukder,S., Philippakis,A.A., Pena-Castillo,L., Alleyne,T.M., Mnaimneh,S., Botvinnik,O.B., Chan,E.T. *et al.* (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, **133**, 1266–1276.
34. Robasky,K. and Bulyk,M.L. (2011) UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.*, **39**, D124–D128.
35. Wei,G.H., Badis,G., Berger,M.F., Kivioja,T., Palin,K., Enge,M., Bonke,M., Jolma,A., Varjosalo,M., Gehrke,A.R. *et al.* (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.*, **29**, 2147–2160.
36. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
37. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
38. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
39. Meeroff,M. (1994) History of the Argentinian Society of Gastroenterology (SAGE). *Acta Gastroenterol. Latinoam.*, **24**, 195–198.
40. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorf,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
41. Li,M.J., Liu,Z., Wang,P., Wong,M.P., Nelson,M.R., Kocher,J.P., Yeager,M., Sham,P.C., Chanock,S.J., Xia,Z. *et al.* (2016) GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **44**, D869–D876.
42. Gong,J., Mei,S., Liu,C., Xiang,Y., Ye,Y., Zhang,Z., Feng,J., Liu,R., Diao,L., Guo,A.Y. *et al.* (2018) PanCanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.*, **46**, D971–D976.
43. Xia,K., Shabalina,A.A., Huang,S., Madar,V., Zhou,Y.H., Wang,W., Zou,F., Sun,W., Sullivan,P.F. and Wright,F.A. (2012) seeQTL: a searchable database for human eQTLs. *Bioinformatics*, **28**, 451–452.
44. Gamazon,E.R., Zhang,W., Konkashbaev,A., Duan,S., Kistner,E.O., Nicolae,D.L., Dolan,M.E. and Cox,N.J. (2010) SCAN: SNP and copy number annotation. *Bioinformatics*, **26**, 259–262.
45. Li,X., Shi,L., Wang,Y., Zhong,J., Zhao,X., Teng,H., Shi,X., Yang,H., Ruan,S., Li,M. *et al.* (2019) OncoBase: a platform for decoding regulatory somatic mutations in human cancers. *Nucleic Acids Res.*, **47**, D1044–D1055.
46. Teng,L., He,B., Wang,J. and Tan,K. (2015) 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics*, **31**, 2560–2564.
47. Wang,S., Sun,H., Ma,J., Zang,C., Wang,C., Wang,J., Tang,Q., Meyer,C.A., Zhang,Y. and Liu,X.S. (2013) Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat. Protoc.*, **8**, 2502–2515.
48. Huynh-Thu,V.A., Irrthum,A., Wehenkel,L. and Geurts,P. (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, **5**, e12776.
49. Lachmann,A., Giorgi,F.M., Lopez,G. and Califano,A. (2016) ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*, **32**, 2233–2235.
50. Han,H., Cho,J.W., Lee,S., Yun,A., Kim,H., Bae,D., Yang,S., Kim,C.Y., Lee,M., Kim,E. *et al.* (2018) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.*, **46**, D380–D386.
51. Corces,M.R., Granja,J.M., Shams,S., Louie,B.H., Seoane,J.A., Zhou,W., Silva,T.C., Groeneveld,C., Wong,C.K., Cho,S.W. *et al.* (2018) The chromatin accessibility landscape of primary human cancers. *Science*, **362**, eaav1898.
52. Consortium, G.T. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
53. Barretina,J., Caponigro,G., Stransky,N., Venkatesan,K., Margolin,A.A., Kim,S., Wilson,C.J., Lehár,J., Kryukov,G.V., Sonkin,D. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
54. de Rie,D., Abugessaisa,I., Alam,T., Arner,E., Arner,P., Ashoor,H., Astrom,G., Babina,M., Bertin,N., Burroughs,A.M. *et al.* (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.*, **35**, 872–878.
55. Pinero,J., Ramirez-Angueta,J.M., Sauch-Pitarch,J., Ronzano,F., Centeno,E., Sanz,F. and Furlong,L.I. (2020) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.*, **48**, D845–D855.
56. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.
57. Law,M. and Shaw,D.R. (2018) Mouse genome informatics (MGI) Is the international resource for information on the laboratory mouse. *Methods Mol. Biol.*, **1757**, 141–161.
58. Oughtred,R., Stark,C., Breitkreutz,B.J., Rust,J., Boucher,L., Chang,C., Kolas,N., O'Donnell,L., Leung,G., McAdam,R. *et al.* (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, D529–D541.
59. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
60. Jassal,B., Matthews,L., Viteri,G., Gong,C., Lorente,P., Fabregat,A., Sidiropoulos,K., Cook,J., Gillespie,M., Haw,R. *et al.* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **48**, D498–D503.
61. Cerami,E.G., Gross,B.E., Demir,E., Rodchenkov,I., Babur,O., Anwar,N., Schultz,N., Bader,G.D. and Sander,C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
62. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.

63. Gene Ontology, C. (2008) The Gene Ontology project in 2008. *Nucleic Acids Res.*, **36**, D440–D444.
64. Liu, W., Stein, P., Cheng, X., Yang, W., Shao, N.Y., Morrisey, E.E., Schultz, R.M. and You, J. (2014) BRD4 regulates Nanog expression in mouse embryonic stem cells and preimplantation embryos. *Cell Death Differ.*, **21**, 1950–1960.
65. Gryder, B.E., Yohe, M.E., Chou, H.C., Zhang, X., Marques, J., Wachtel, M., Schaefer, B., Sen, N., Song, Y., Gualtieri, A. *et al.* (2017) PAX3-FOXO1 establishes myogenic super enhancers and confers BET bromodomain vulnerability. *Cancer Discov.*, **7**, 884–899.
66. Gonzales-Cope, M., Sidoli, S., Bhanu, N.V., Won, K.J. and Garcia, B.A. (2016) Histone H4 acetylation and the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells. *BMC Genomics*, **17**, 95.
67. Wang, W., Tang, Y.A., Xiao, Q., Lee, W.C., Cheng, B., Niu, Z., Oguz, G., Feng, M., Lee, P.L., Li, B. *et al.* (2021) Stromal induction of BRD4 phosphorylation results in chromatin remodeling and BET inhibitor resistance in colorectal cancer. *Nat. Commun.*, **12**, 4441.
68. Donati, B., Lorenzini, E. and Ciarrocchi, A. (2018) BRD4 and cancer: going beyond transcriptional regulation. *Mol. Cancer*, **17**, 164.
69. Ma, T., Hu, C., Lal, B., Zhou, W., Ma, Y., Ying, M., Prinos, P., Quinones-Hinojosa, A., Lim, M., Lattera, J. *et al.* (2021) Reprogramming transcription factors Oct4 and Sox2 Induce a BRD-dependent immunosuppressive transcriptome in GBM-propagating cells. *Cancer Res.*, **81**, 2457–2469.