

# TFBSbank: a platform to dissect the big data of protein–DNA interaction in human and model species

Dongsheng Chen<sup>1,\*</sup>, Sanjie Jiang<sup>2</sup>, Xiaoyan Ma<sup>3</sup> and Fang Li<sup>4,\*</sup>

<sup>1</sup>Fitzwilliam College, University of Cambridge, Storey's Way, Cambridge CB3 0DG, UK, <sup>2</sup>St Edmund's College, University of Cambridge, Mount Pleasant, Cambridge CB3 0BN, UK, <sup>3</sup>Downing College, University of Cambridge, Regent St, Cambridge CB2 1DQ, UK and <sup>4</sup>Department of Gynaecology, Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, Shanghai 200040, China

Received June 16, 2016; Revised October 16, 2016; Editorial Decision October 18, 2016; Accepted October 23, 2016

## ABSTRACT

**Genome-wide transcription factors (TFs) binding data has been extensively generated in the past few years, which poses a great challenge to data interpretation. Therefore, comprehensive and dedicated functional annotation databases for TF–DNA interaction are in great demands to manage, explore and utilize those invaluable data resources. Here, we constructed a platform ‘TFBSbank’ which houses the annotation of 1870 chromatin immunoprecipitation (ChIP) datasets of 585 TFs in five species (human, mouse, fly, worm and yeast). There are mainly five functional modules in TFBSbank aimed at characterizing ChIP peaks, identifying putative targets, predicting TF responsive enhancers, revealing potential cofactors/collaborators and discovering enriched TF motifs. TFBSbank has two distinctive features compared to the existing databases. Firstly, we provided putative cofactors/collaborators analysis (for *Drosophila melanogaster*), as they are crucial for the *in vivo* functions of TFs. Additionally, this database predicted the enrichment of both known and *de novo* motifs based on ChIP data. TFBSbank is freely accessible at <http://tfbsbank.co.uk>**

## INTRODUCTION

The regulation of gene expression is the cornerstone of organism development. Gene expression can be regulated at the epigenetic level by DNA modification and histone modification (1, 2), at the transcriptional level by transcription factors (TFs) regulation (3) and at the post-transcription level by small RNAs mediated silencing such as miRNAs, siRNA and piwiRNA (4–6). Although the importance of TFs has been well recognized for decades, the mechanisms through which TFs select their targets and how they interact with cofactors and collaborators remain largely un-

known. To understand the interactions between TFs and DNA, international cooperation projects such as modENCODE (7) and ENCODE (8) have been launched for species including *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (worm), *Mus musculus* (mouse) and *Homo sapiens* (human), using methods such as ChIP-chip or ChIP-Seq. To date, there are just a few repositories available for chromatin immunoprecipitation (ChIP) data. Specifically, GEO (9), Arrayexpress (10) and SRA (11) are data warehouses mainly for the storage of ChIP data. hmChIP (12) provides the querying and retrieving of protein–DNA interaction data for mouse and human. ChIPBase (13) focuses on TF–lncRNA and TF–miRNA interactions. In this study, we developed a specialized and powerful platform to dissect the TF genome binding profiles in five species (fly, worm, yeast, mouse and human). Briefly, we performed functional annotation of ChIP data. Subsequently, all the results were stored in a MySQL database and presented in integrated web interfaces, allowing users to search and browse the functional annotation of TFs ChIP data.

## MATERIALS AND METHODS

### Data collection

ChIP-chip and ChIP-Seq data were downloaded from ENCODE (human and mouse), modENCODE (fly and worm) and UCSC (yeast) respectively. In case of peak files annotated with different reference genome versions, liftOver from UCSC was used to convert genome positions (dm3 for fly, ce6 for worm, hg19 for human, mm9 for mouse and sacCer3 for yeast).

### Basic analysis

Brief descriptions of the TF (species name, gene ID, gene name, GO terms) were retrieved from Ensemble BioMart database. The length distribution of peaks and their distance to TSSs were calculated using a custom R script.

\*To whom correspondence should be addressed. Email: dc571@cam.ac.uk or dschen2016@gmail.com  
Correspondence may also be addressed to Fang Li. Tel: +86 21 5403 0319; Tax: +86 21 5073 0190; Email: fang.li@tongji.edu.cn

**Table 1.** Statistics of the TFBSbank

Species	Scientific name	Number of TFs	Number of datasets	Data source	Reference genome version genome	Available analysis
Fruit Fly	<i>Drosophila melanogaster</i>	77	183	modENCODE	dm3	1,2,3,4,5
Worm	<i>Caenorhabditis elegans</i>	106	214	modENCODE	ce6	1,2,5
Human	<i>Homo sapiens</i>	151	962	ENCODE	hg19	1,2,4,5
Mouse	<i>Mus musculus</i>	50	168	ENCODE	mm9	1,2,4,5
Yeast	<i>Saccharomyces cerevisiae</i>	201	343	UCSC	sacCer3	1,2,5

1, basic analysis; 2, target analysis; 3, co-binding analysis; 4, known motif analysis; 5, *de novo* motif analysis.

### Putative target genes assignment

The ChIP peaks were compared to transcript annotation database using a custom Perl script. ‘TxDb.Dmelanogaster.UCSC.dm3.ensGene’, ‘TxDb.Celegans.UCSC.ce6.ensGene’, ‘TxDb.Mmusculus.UCSC.mm9.knownGene’, ‘TxDb.Hsapiens.UCSC.hg19.knownGene’ and ‘TxDb.Scerevisiae.UCSC.sacCer3.sgdGene’ were used for the gene annotation of fly, worm, mouse, human and yeast respectively. To determine the genomic features of the ChIP peaks, locations of peaks were compared with the locations of untranslated region (UTR), intron, exon, promoter and intergenic regions. In total, five methods were used to assign target genes: ‘physical overlap’, ‘nearest gene’ and ‘neighbor overlap’ (1, 10, 100 kb). ‘Physic overlap’ reports all genes directly overlapping with ChIP peaks. ‘Nearest gene’ reports the transcription starting site (TSS) nearest to ChIP peak center. ‘Neighbor overlap’ reports all genes overlapping with the extensions of peaks center by certain distances (1, 10, 100 kb). For instance, if a ChIP peak named ‘ChIP-peak-001’ is located at the position ‘chr2L: 300 000-301 000’, then all the genes directly overlapping with this region will be considered as targets of ‘ChIP-peak-001’ according to the ‘physical overlap’ method. Genes with TSS closest to the peak center ‘chr2L: 300 500’ will be considered as target of ‘ChIP-peak-001’ according to the ‘nearest target’ method. Regarding the ‘neighbor overlap method’, we will extend the peak center by 1, 10 and 100 kb to each direction, and all the genes overlapping with the extensions will be considered as putative targets. Therefore, five sets of target genes were predicted based on those criteria, and they were called ‘overlap\_target’, ‘nearest\_target’, ‘neighbor\_1kb\_target’, ‘neighbor\_10kb\_target’ and ‘neighbor\_100kb\_target’ respectively.

### Hypergeometric tests for GO term and KEGG pathway

Standard Hypergeometric tests were conducted to identify enriched gene ontology (GO) terms in putative targets. Specifically, the ‘GOHyperGParams’ method from the R package GStats (14) was employed to identify over-represented GO terms and ‘KEGGHyperGParams’ method was utilized to reveal enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. After calculating *P*-values through Hypergeometric tests for each term, we performed multiple test correction using Benjamini-Hochberg (BH) method to control the rate of type I errors. All the GO and KEGG terms with the adjusted *P* value less than 0.01 were considered as significantly enriched.

### Co-binding analysis

To investigate the potential co-binding of TFs, we extracted experimentally confirmed TF binding sites (TF-

BSs) from Redfly (15) database and overlapped them with ChIP peaks. TFs having TFBSs directly overlapping with the ChIP peaks were considered as ‘overlapping TFs’. In addition, we extracted the protein–protein interaction networks from DroID (16). TFs having protein–protein interactions with investigated TFs were considered as ‘interacting TFs’. The intersection between ‘overlapping TFs’ and ‘interacting TFs’ gave rise to a subset of TFs which not only co-bound with but also directly interacted with the investigated TF, and they were considered as putative cofactors/collaborators.

### Motif enrichment scanning

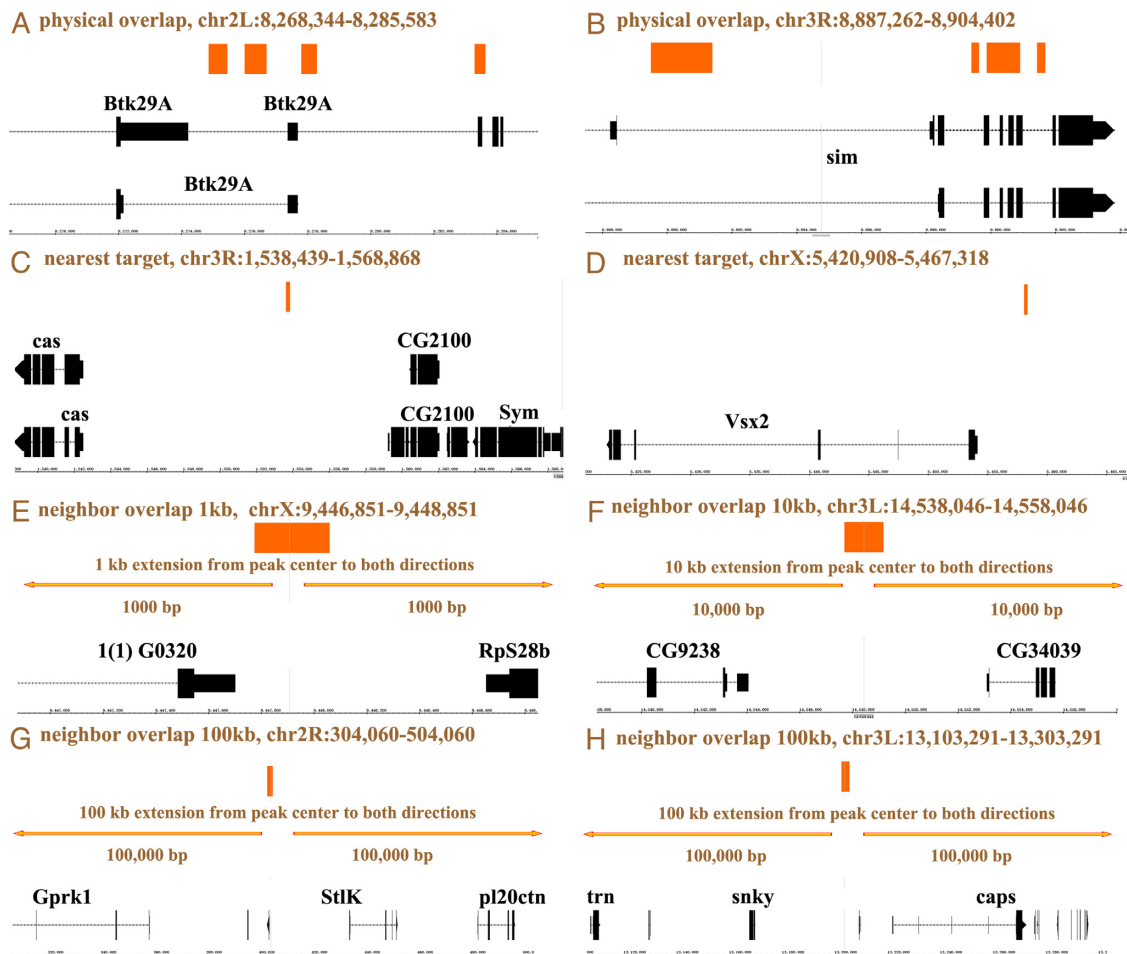
The enrichment analysis of known motifs in TF ChIP peaks was performed using PWMEnrich (17) with default parameters. We were only able to conduct known motif enrichment analysis for human, mouse and fly, as PWMEnrich only supports the analysis for those three species. The *de novo* motif scanning was performed using findMotifsGenome.pl from HOMER (18). Specifically, the DNA sequences within 100 bp of the peak centers were extracted from the genome and compared to the genome background sequences to reveal potential enriched *de novo* motifs.

## RESULTS

### Overview of TFBSbank

TFBSbank currently encompasses the annotation of 1870 TFs ChIP data, including 183, 214, 962, 168 and 343 datasets from fly, worm, human, mouse and yeast respectively (Table 1). The biological samples range from a variety of cell lines or tissues at different development stages. There are five functional modules for the annotation of each TF binding profile: (i) ‘basic analysis’ for characterizing length distribution, distance to TSSs and associated genomic features (promoter, exon, intron and intergenic regions) of ChIP peaks; (ii) ‘target analysis’ for putative target gene assignment and GO term/ KEGG pathway enrichment; (iii) ‘co-binding analysis’ for identifying potential TF responsive cis-regulatory modules (CRMs) and putative cofactor/collaborators; (iv) ‘known motif analysis’ for investigating enriched known motifs in TF ChIP peaks; (v) ‘*de novo* motif analysis’ for scanning enriched *de novo* motifs in TF ChIP peaks.

To demonstrate functions of TFBSbank, we presented here the analytical result of published ChIP-Seq data for Homothorax (*hth*) (7,19), key regulator in organ development and morphology in fruit fly (20–22). The ChIP experiment was conducted using 0–8 h old fly embryos. In total, 6207 peaks have been identified. The GFF format file of

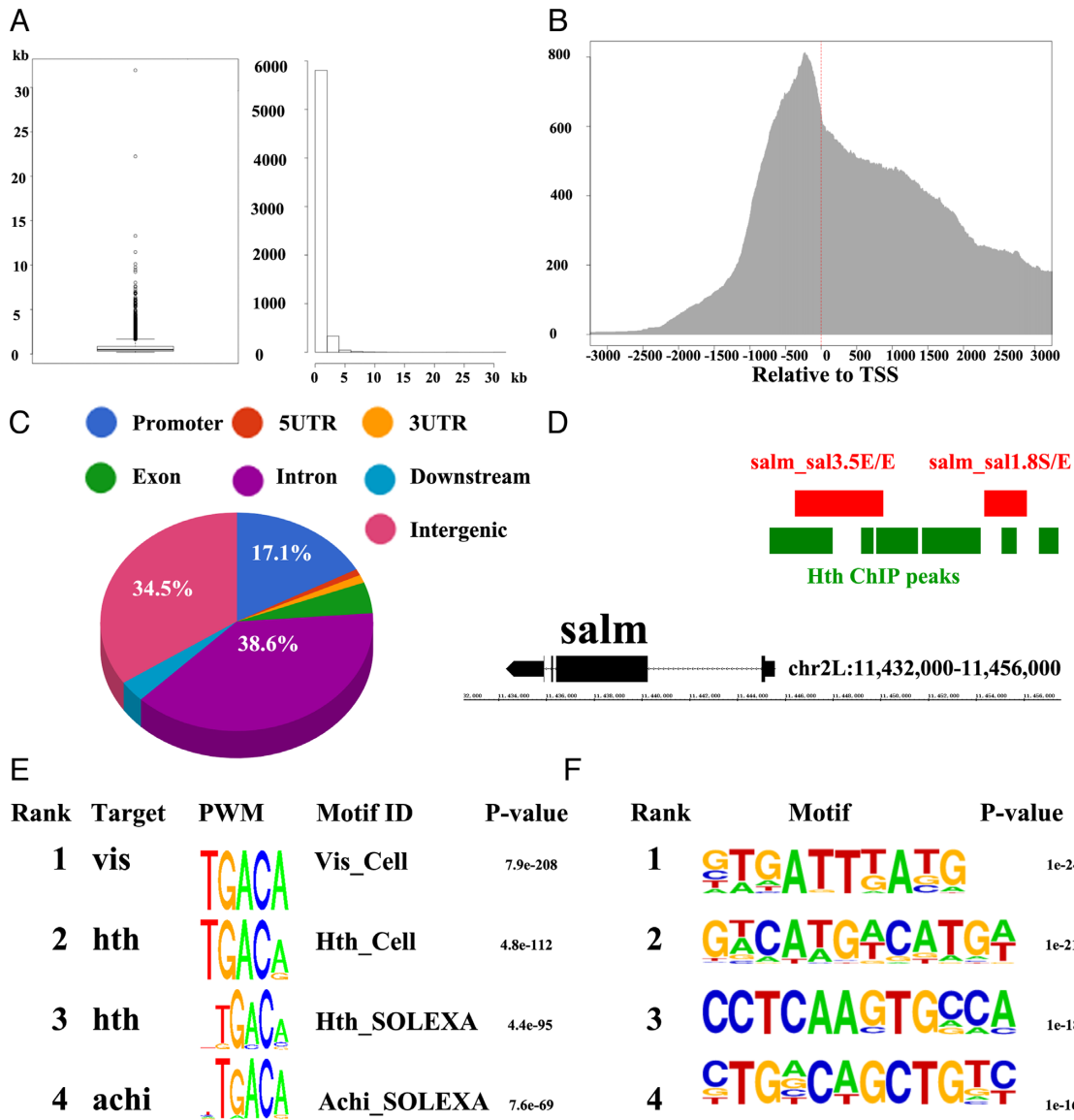


**Figure 1.** Examples of Hth putative targets identified using ‘physical overlap’ method (A and B), ‘nearest target’ method (C and D), ‘neighbor overlap 1 kb’ method (E), ‘neighbor overlap 10 kb’ method (F) and ‘neighbor overlap 100 kb’ method (G and H) respectively. Orange rectangles, Hth ChIP peaks; Orange arrows, extensions from ChIP peak center. Gene models were extracted from ‘TxDb.Dmelanogaster.UCSC.dm3.ensGene’ R package (29) and viewed in IGB (30).

Hth ChIP-Seq data was downloaded from modENCODE database (ID: modENCODE.4070), subsequently analyzed using our standard TFBSbank pipeline. Corresponding results were integrated into TFBSbank with the ID ‘DB0097’, which can be also accessed by the ‘Result demo’ option on the menu located on the top of each webpage.

- i) ‘Basic analysis’. First, distribution of the length of identified 6207 ChIP peaks is demonstrated in a box-plot (Figure 2A). In addition, a histogram plot was used to summarize the number of ChIP peaks in each length region, indicating that most of the peaks were under 5 kb. Also, we calculated the frequency of distance between Hth ChIP peaks and corresponding TSSs in a window of 500 bp (Figure 2B). Next step, we calculated the proportions of Hth ChIP peaks overlapping with promoter, UTR, exon, intron, downstream and intergenic regions respectively (Figure 2C). The result indicated the majority of Hth peaks were located in introns (38.6%), intergenic regions (34.5%) and promoters (17.1%).
- ii) ‘Target analysis’. Target genes were predicted in five conditions: ‘overlap\_target’, ‘nearest\_target’, ‘neighbor\_1kb\_target’, ‘neighbor\_10kb\_target’ and ‘neighbor\_100kb\_target’ representing genes physically overlapping with ChIP peaks, genes nearest to ChIP peaks and genes overlapping with 1, 10, 100 kb extensions of ChIP peaks respectively (Figure 1). In total, 1476, 2140, 2020, 5515 and 13919 putative target genes were predicted in those five conditions from Hth ChIP data (Supplementary File 1). Consistent to the known roles of Hth during development, GO analysis of predicted target gene datasets from all five conditions showed an significant enrichment of GO terms related to development processes including ‘organ development’, ‘wing disc development’ and ‘appendage morphogenesis’ (Supplementary File 2). Interestingly, we also found that Hth targets were enriched in a variety of signaling pathways including ‘Hedgehog signaling pathways’, ‘Wnt signaling pathway’, ‘Jak-STAT signaling pathway’ and ‘Notch signaling pathway’ (Supplementary File 3), which have been proposed to regulate development processes (23–27). Therefore, our target gene prediction function recapitulates the known function of

bor\_1kb\_target’, ‘neighbor\_10kb\_target’ and ‘neighbor\_100kb\_target’ representing genes physically overlapping with ChIP peaks, genes nearest to ChIP peaks and genes overlapping with 1, 10, 100 kb extensions of ChIP peaks respectively (Figure 1). In total, 1476, 2140, 2020, 5515 and 13919 putative target genes were predicted in those five conditions from Hth ChIP data (Supplementary File 1). Consistent to the known roles of Hth during development, GO analysis of predicted target gene datasets from all five conditions showed an significant enrichment of GO terms related to development processes including ‘organ development’, ‘wing disc development’ and ‘appendage morphogenesis’ (Supplementary File 2). Interestingly, we also found that Hth targets were enriched in a variety of signaling pathways including ‘Hedgehog signaling pathways’, ‘Wnt signaling pathway’, ‘Jak-STAT signaling pathway’ and ‘Notch signaling pathway’ (Supplementary File 3), which have been proposed to regulate development processes (23–27). Therefore, our target gene prediction function recapitulates the known function of



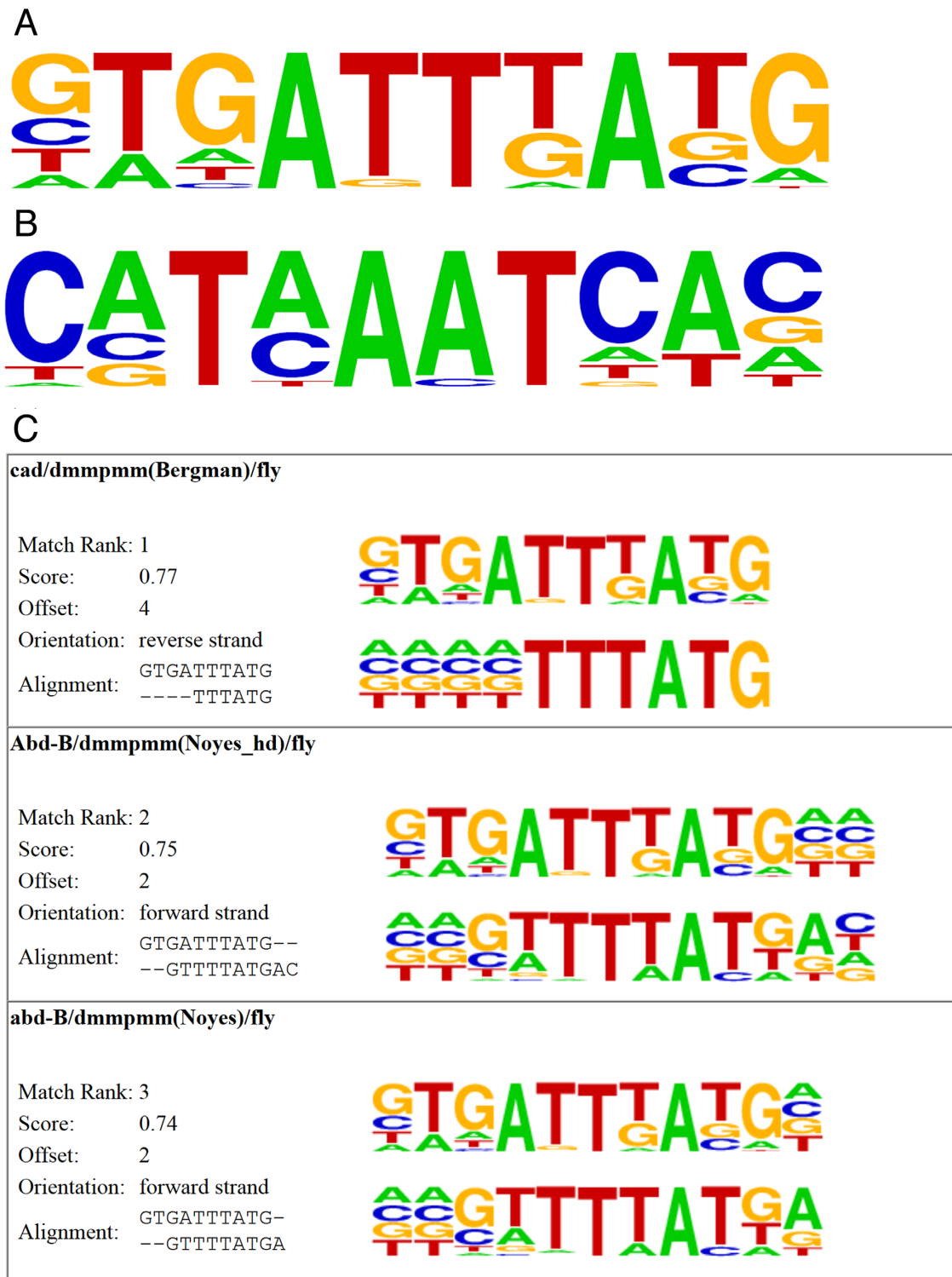
**Figure 2.** The Analysis results of Hth ChIP-Seq data. (A) The length distribution of ChIP peaks. (B) The frequency of distance between Hth peaks and TSSs. (C) The proportions of Hth peaks associated with promoter, 5 UTR, 3 UTR, exon, intron, downstream and intergenic regions. (D) Genome annotation tracks of Hth peaks and experimentally validated enhancers from Redfly database. (E) Enriched known motifs in Hth peaks. (F) Enrichment of *de novo* motifs in Hth peaks.

Hth as key regulator in a variety of signaling pathways to orchestrate organism development.

iii) ‘Co-binding analysis’. Generally, TFs regulate target genes’ expression patterns through binding to their CRMs together with other cofactors or collaborators. To dissect the complicated interaction networks between trans factors and cis elements, co-binding analysis was performed for fruit fly by comparing ChIP peaks with experimentally-verified enhancers and TFBSs retrieved from Redfly database (15). Basically, co-binding analysis identifies: (a) overlapping enhancers and TFBSs which have been verified through experimental methods (Figure 2D); (b) ‘interacting proteins’ having interactions with investigated TF; (c) ‘putative cofactors/collaborators’ which not only inter-

acted with but also co-bound with investigated TF. The co-binding analysis of Hth ChIP-Seq data suggests that 6207 ChIP peaks overlapped with 386 experimentally verified enhancers (Supplementary File 4) and 608 TFBSs (Supplementary File 5) in Redfly database. Among the 62 proteins interacting with Hth (Supplementary File 6), five of them were considered as potential cofactors/collaborators of Hth, being Tailless (tll), Asense (ase), Empty spiracles (ems), Engrailed (en) and Cubitus interruptus (ci).

iv) ‘Known motif analysis’. To understand the cis-regulatory code of TFs, we conducted known motifs enrichment scanning using an R package called PWMEnrich (17). As a result, 188 existing motifs were identified in Hth peaks with *P*-value



**Figure 3.** Information for *denovo* motif 1 indentified by HOMER (18). (A) Forward logo of *denovo* motif 1. (B) Reverse opposite logo of *denovo* motif 1. (C) Matched known motifs of *denovo* motif 1.

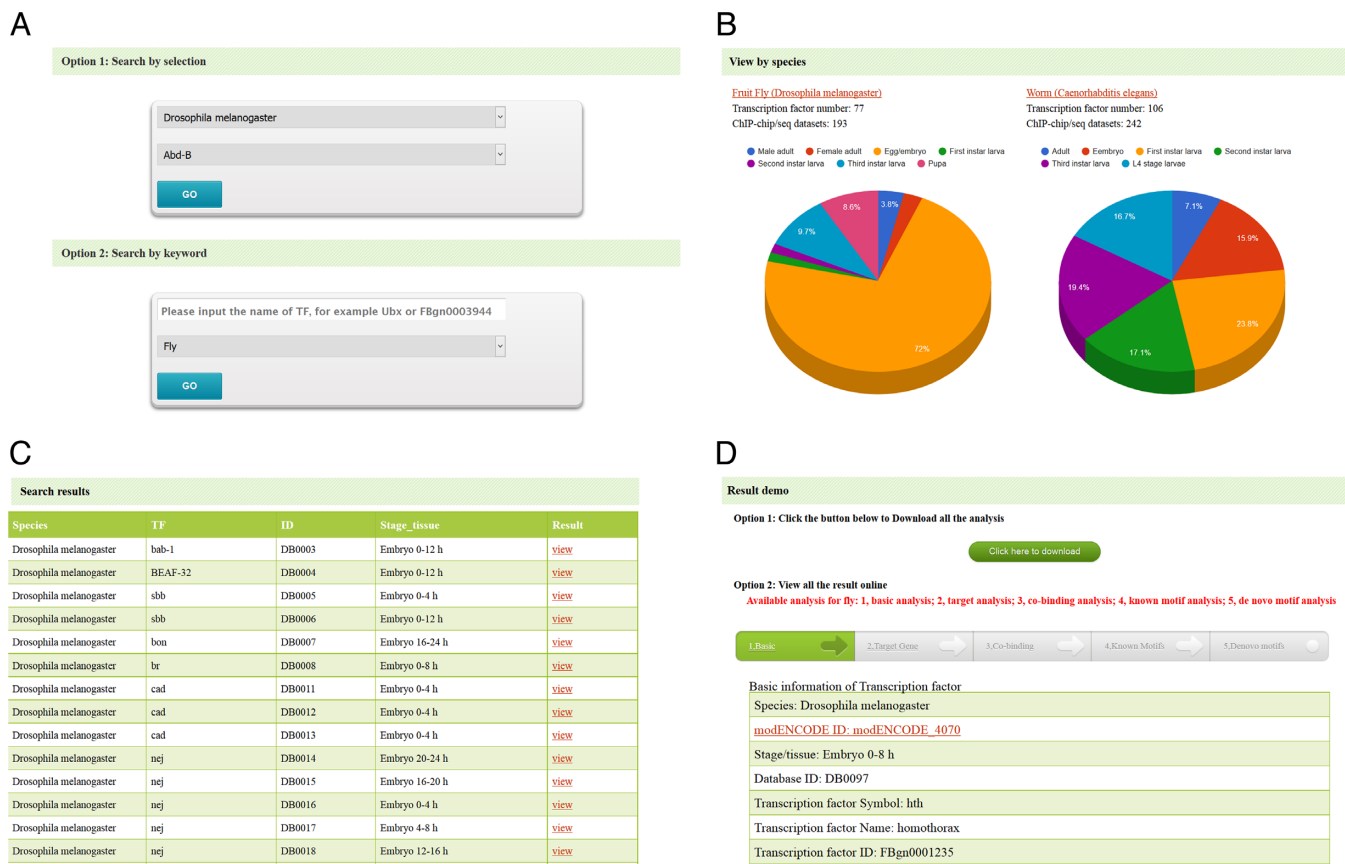


Figure 4. An overview of TFBSbank. (A) Searching page. (B) Browsing page. (C) Listing page. (D) Data viewing page.

less than 0.01, four of which with  $P$ -value  $< 4e-95$ , were Hth motifs (Hth\_Cell\_FBgn0001235, Hth\_SOLEXA\_FBgn0001235, hth\_SOLEXA\_2\_FBgn0001235 and CORE-hth-MA0227.1). Therefore, this result validates these Hth binding sites in *in vivo* (Figure 2E and Supplementary File 7). Additionally, the motif of Exd, a cofactor which has been reported to directly interact with Hth in an essential manner for its nuclear translocation (28), was also found to be enriched in the Hth ChIP peaks.

- v) ‘*De novo* motif analysis’. We also predicted *de novo* motifs in TF ChIP peaks by HOMER (18). In summary, 27 *de novo* motifs, ranked according to their  $P$ -values, (Figure 2F and Supplementary File 8), and the percentage of each motif in Hth peaks versus the rest of the genome were listed in the output webpage. Besides, in order to assist users in dissecting those *de novo* motifs, a link was provided at the end of each row to compare *de novo* motifs with known motifs (Figure 3 and Supplementary File 9).

### Data searching

We supplied both selection-based query and keywords-based query options to facilitate the search (Figure 4A). Users can either build their query by selecting species and the corresponding TF, or by inputting keywords related to TF names and choosing the corresponding species. Af-

ter submitting keywords and species name, a SQL inquiry command would be executed and a list of matched records would be returned in the front webpage in the tabular form (Figure 4C). Query without keywords is possible. In that case, all the records of a given species will be returned.

### Data browsing

To enhance user experience, we supplied an interface containing the list of five species, which were hyperlinked to a species-specific data browsing page. In fly and worm, a picture of the life cycle of the animal was provided, from which the users can specify certain development stages by clicking corresponding icon (i.e. embryos, larva, pupa and adults). Alternatively, results from all developmental stages will be displayed.

### Data demonstration

A ‘result.php’ file was generated at the end of the analysis, presenting results retrieved from all five functional modules (Figure 4D). Those results were organized into flowcharts on the top of the page. The content will be displayed by clicking the name of the analysis in the flowchart.

### DISCUSSION

TFBSbank is a comprehensive and dedicated platform for the functional annotation of TF–DNA interactions. Cur-

rently, it includes 1870 data sets from five species. Apart from basic annotation of ChIP peaks (GO term/pathway enrichment, genomic features), TFBSbank provides the analysis of co-binding, motif enrichment (known motifs and *de novo* motifs) and cofactors/collaborators prediction, which are crucial to TF binding and activity regulation. As the genome-wide binding data of TFs increases rapidly, we will continue maintaining and updating our database, to provide researchers with the latest annotations. Furthermore, more functional modules will be integrated into TFBSbank to enhance its functionality and usefulness. Users are encouraged to submit their own ChIP-chip/Seq data to fang\_li@tongji.edu.cn, which would be integrated to our database after revision and analysis. We hope that TFBSbank would benefit the research community by providing more insights into the functions of TFs binding.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the ENCODE, modENCODE and UCSC for sharing their data. We would like to thank Bettina Fisher for her expertise in programming and Min Li for helping us collecting the data. Thanks are also due to Jiao Sima who provided instructive feedbacks on our project.

## FUNDING

Science and Technology Committee of Shanghai [124119a3801, 124119a3800 to F.L.]; China Scholarship Council (to D.C., S.J., X.M.). Funding for open access charge: Science and Technology Committee of Shanghai [124119a3801, 124119a3800 to F.L.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Jørgensen, H.F., Azuara, V., Amoils, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B.S., Ramsahoye, B., Merckenschlager, M. and Fisher, A.G. (2007) The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. *Genome Biol.*, **8**, R169.
- Ray, L.B. (2009) Chromatin modifier modulates gene expression. *Sci. Signal.*, **2**, ec175.
- Levine, M. and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature*, **424**, 147–151.
- Bartel, D. (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Castel, S.E. and Martienssen, R.A. (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.*, **14**, 100–112.
- Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.*, **10**, 94–108.
- Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F. *et al.* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*, **330**, 1787–1797.
- The ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T. *et al.* (2015) ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.*, **43**, D1113–D1116.
- Leinonen, R., Sugawara, H. and Shumway, M. (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
- Chen, L., Wu, G. and Ji, H. (2011) hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data. *Bioinformatics*, **27**, 1447–1448.
- Yang, J.-H., Li, J.-H., Jiang, S., Zhou, H. and Qu, L.-H. (2013) ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res.*, **41**, D177–D187.
- Falcon, S. and Gentleman, R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics*, **23**, 257–258.
- Gallo, S.M., Gerrard, D.T., Miner, D., Simich, M., Des Soye, B., Bergman, C.M. and Halfon, M.S. (2011) REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. *Nucleic Acids Res.*, **39**, D118–D123.
- Yu, J., Pacifico, S., Liu, G. and Finley, R.L. (2008) DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics*, **9**, 461.
- Stojnic, R. and Diez, D. (2015) PWMEnrich: PWM enrichment analysis. R package version 4.10.0. <https://bioconductor.org/packages/release/bioc/html/PWMEnrich.html>.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
- Contrino, S., Smith, R.N., Butano, D., Carr, A., Hu, F., Lyne, R., Rutherford, K., Kalderimis, A., Sullivan, J., Carbon, S. *et al.* (2012) modMine: flexible access to modENCODE data. *Nucleic Acids Res.*, **40**, D1082–D1088.
- Torres, M., Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G. and Martínez-A, C. (1999) Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature*, **402**, 425–429.
- Casares, F. and Mann, R.S. (1998) Control of antennal versus leg development in Drosophila. *Nature*, **392**, 723–726.
- Inbal, A., Halachmi, N., Dibner, C., Frank, D. and Salzberg, A. (2001) Genetic evidence for the transcriptional-activating function of Homothorax during adult fly development. *Development*, **128**, 3405–3413.
- Artavanis-Tsakonas, S. (1999) Notch signaling: cell fate control and signal integration in development. *Science (80-)*, **284**, 770–776.
- Aaronson, D.S. and Horvath, C.M. (2002) A road map for those who don't know JAK-STAT. *Science*, **296**, 1653–1655.
- Nusse, R. (2005) Wnt signaling in disease and in development. *Cell Res.*, **15**, 28–32.
- Nusse, R. and Varmus, H.E. (1992) Wnt genes. *Cell*, **69**, 1073–1087.
- Ingham, P.W., Nakano, Y. and Seger, C. (2011) Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.*, **12**, 393–406.
- Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M. and Mann, R.S. (1997) Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell*, **91**, 171–83.
- Carlson, M. and Maintainer, B.P. (2015) TxDb.Dmelanogaster.UCSC.dm3.ensGene: Annotation package for TxDb object(s). R package version 3.2.2. <http://bioconductor.org/packages/release/data/annotation/html/TxDb.Dmelanogaster.UCSC.dm3.ensGene.html>.
- Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A. and Loraine, A.E. (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, **25**, 2730–2731.