Published online 26 October 2016 *Nucleic Acids Research, 2017, Vol. 45, Database issue* ***D145–D150***
doi: 10.1093/nar/gkw1007

TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions

Sebastian Schmeier[1], Tanvir Alam[2], Magbubah Essack[2] and Vladimir B. Bajic[2,*]

[1]Massey University Auckland, Institute of Natural and Mathematical Sciences, Auckland, New Zealand and [2]King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Thuwal, Kingdom of Saudi Arabia

Received August 29, 2016; Revised September 29, 2016; Editorial Decision October 15, 2016; Accepted October 17, 2016

ABSTRACT

Transcription factors (TFs) play a pivotal role in transcriptional regulation, making them crucial for cell survival and important biological functions. For the regulation of transcription, interactions of different regulatory proteins known as transcription co-factors (TcoFs) and TFs are essential in forming necessary protein complexes. Although TcoFs themselves do not bind DNA directly, their influence on transcriptional regulation and initiation, although indirect, has been shown to be significant, with the functionality of TFs strongly influenced by the presence of TcoFs. In the TcoF-DB v2 database, we collect information on TcoFs. In this article, we describe updates and improvements implemented in TcoF-DB v2. TcoF-DB v2 provides several new features that enables exploration of the roles of TcoFs. The content of the database has significantly expanded, and is enriched with information from Gene Ontology, biological pathways, diseases and molecular signatures. TcoF-DB v2 now includes many more TFs; has substantially increased the number of human TcoFs to 958, and now includes information on mouse (418 new TcoFs). TcoF-DB v2 enables the exploration of information on TcoFs and allows investigations into their influence on transcriptional regulation in humans and mice. TcoF-DB v2 can be accessed at http://tcofdb.org/.

INTRODUCTION

Transcriptional regulation in eukaryotes is a complex process. It involves many proteins that assemble at the site of transcription initiation (1,2) either binding DNA directly, e.g. transcription factors (TFs) (3–5), or interacting with TFs, e.g. transcription co-factors (TcoFs) (6–9). These proteins work in unison to facilitate the recruitment of RNA Polymerase II to the site of transcription initiation, which substantially increases the complexity of the regulatory process (10). TcoFs have many functions, such as signal transduction, modulation of TF–DNA binding, and chromatin modification (11,12). However, so far research on the interaction of TFs with other proteins in relation to transcription regulation has not been that intensive.

Protein–protein interactions (PPI) play a crucial role in transcriptional regulation and can take several forms, e.g. the modification of one protein by another protein, the formation of a protein complex by two or more proteins, etc. PPIs of our interest include the physical interaction between two TFs (10,13), the physical interaction between a TF and a TcoF, as well as potentially, the physical interaction between a TF and other nuclear proteins that are not known to function as TFs or TcoFs. A web-based tool to recognise the type of TF binding partner has also been developed (14).

Several resources that collect information on mammalian TFs exist (e.g. JASPAR (15), TRANSFAC (16), HOCOMOCO (17), TFCat (18), DBD (19), TFCONES (20), TFcheckpoint (21), TFdb (22), COMPEL (23), TRANSCompel (16), etc.). While the influence of TFs on many biological processes (24–27) and diseases (26,28–30) is well established and described, research into the influence of TcoFs has not received the same level of attention. However, the recent release of the FARNA database (http://cbrc.kaust.edu.sa/farna) utilised TcoFs (31) to infer functions of over 10 000 non-coding transcripts in humans, showing the importance of TcoFs for the function of their target transcripts. To facilitate investigations into the influence of TcoFs on the regulation of transcriptional initiation, we first developed in 2010 the TcoF-DB database that comprises human TFs and 529 TcoFs that interact with them (31). In 2014, 415 human TcoFs as well as TcoFs from other species were published in AnimalTFDB 2.0 database (32). However, we still do not know the complete list of TFs and TcoFs in mammals. Thus, we developed an updated version of TcoF-DB v2, which increases the number

*To whom correspondence should be addressed. Tel: +966 12 808 2386; Fax: +966 2 802 0127; Email: vladimir.bajic@kaust.edu.sa

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from http://nar.oxfordjournals.org/ by guest on January 5, 2017

of TFs and TcoFs substantially. Since the last TcoF-DB release, several important changes have been made to the resources on which TcoF-DB depends upon, e.g. Gene Ontology (GO) updated its annotations and structure (33), and PPI resources grew substantially. These changes made an update of TcoF-DB necessary. In this article, we describe new features and modifications we have made in the TcoF-DB v2 resource, and the updated version of TcoF-DB.

The update of transcription factors

We updated our lists of TFs that form the foundation of TcoF-DB v2. In TcoF-DB v2, we opted for a gene-centric approach as opposed to a protein-centric one, to better support genomics and transcriptomics studies, which in general are gene-centric. We extracted human genes that encode TFs (we denote them as TF genes) from a recent large-scale study of promoter elements (4). After a hand-curation step to remove obsolete entries, the numbers of human TF genes in TcoF-DB v2 increased to a total of 1 758. In addition, we now expanded the database by including 1 507 mouse TF genes (4).

The update of transcription co-factors

We extracted genes and proteins that interact with TFs from three databases: BioGRID v3.4.135 (34), IntAct (accessed April 2016) (35) and Reactome (accessed April 2016) (36). BioGRID stores interactions in terms of gene-to-gene identifiers, although the actual interaction is between proteins the genes encode for. To conform to our gene-centric approach, and to be consistent, we converted the protein identifiers in the PPIs in IntAct and Reactome to gene identifiers. We excluded the formerly used database MINT (37), since from 2013 its data has been integrated into IntAct. All data from the above three databases are experimentally supported. It was of particular importance that this information is presented in the PSI-MI format [molecular interaction standard of the Proteomics Standards Initiative (38)] in order to allow us to focus on PPIs of certain types. We only consider PPIs that represent physical interactions between two proteins [the PSI–MI interaction type 0914:'association' and its descendent terms: 0915:'physical association', 0407:'direct interaction', 0195:'covalent binding', 0408:'disulfide bond', and 0556:'transglutamination reaction'], and where both interacting partners are from the same species.

In this manner, we were able to extract 37 250 (4 428) unique interactions between two human (mouse) genes, where at least one of the participants is a TF gene from our list and the interaction type is one of the above-mentioned physical interactions.

Initially, we considered all genes (denoted as non-TF genes as they have not been part of our TF gene list) that were identified as having a physical interaction with known TF genes. We found that 7 437 distinct human non-TF genes were interacting with 1 397 human TF genes. However, for the remainder of the 361 TF genes we did not find any interaction with non-TF genes in the interaction sources mentioned above. Similarly, we found that 2 062 distinct mouse genes were interacting with 603 mouse TF genes, leaving 904 mouse TF genes without interactions to a non-TF genes.

Out of the 7 437 (2 062) human (mouse) TF interacting genes, we have now identified those that can be considered to encode TcoFs. Similar to our former methodology (31), four conditions had to be met by a gene so that we consider it a TcoF (see Table 1). However, note that the GO in recent years has updated its annotations, thus, there was a need to adjust the methodology in this regard and it is reflected by new GO-terms within condition four (33).

As before, we follow the classical definition of TcoFs from (9) and classify all TcoFs into four groups (31), based on the type of evidence that is present regarding fulfilment of conditions c and d in Table 1 (conditions a and b from Table 1 are compulsory for all TcoF groups):

- High-confidence TcoFs: All TcoFs that have experimental evidence for both, involvement in transcription regulation and for occurrence in the cell nucleus.
- Hypothetical TcoFs (Class 1): All TcoFs that have experimental evidence for involvement in transcription regulation, but only non-experimental evidence (e.g. 'Inferred from Electronic Annotation' or 'Author statement') for occurrence in the cell nucleus.
- Hypothetical TcoFs (Class 2): All TcoFs that have experimental evidence for occurrence in the cell nucleus, but only non-experimental evidence for involvement in transcription regulation.
- Hypothetical TcoFs (Class 3): All TcoFs that have only non-experimental evidence for both, involvement in transcription regulation and for occurrence in the cell nucleus.

This classification is done based on the type of evidence that is attached to the GO-terms of conditions c and d in Table 1. We distinguish between experimental evidence (GO codes: EXP, IDA, IMP, IGI, IEP and IPI) and non-experimental evidence (all other codes; for more detail on the codes see http://geneontology.org/page/guide-go-evidence-codes). Thus, the predicted TcoFs have a varying degree of reliability, ranging from high-confidence TcoFs to hypothetical TcoFs of classes 1–3, with class 3 being the least confident class of TcoFs. Tables 2 and 3 show the distribution among the four groups for human and mouse TcoFs, as well as the evidence types necessary for a gene to belong to one of the groups. Only TcoFs that have experimental evidence cited for at least one GO annotation relevant to transcriptional regulation, in addition to experimental evidence for the occurrence in the cell nucleus, are considered TcoFs with high confidence. In the TcoF-DB v2 the number of high confidence human TcoFs more than doubled from 155 in TcoF-DB to currently 389 TcoFs. The total number of human TcoFs (high confidence and hypothetical) increased substantially to now 958 TcoFs. Mouse TcoFs were not part of TcoF-DB, but are included in TcoF-DB v2, which now allows for easier exploration of the influence of TcoFs on transcriptional regulation in mouse.

Out of 1 758 human (1 507 mouse) TFs in our database, 1 133 human (432 mouse) TFs have at least one interaction with a TcoF. Supplementary Table S2 shows the top 5 human and mouse TFs in terms of number of interactions to TcoFs. The human TF with most interactions to TcoFs is tumor protein p53 (TP53), which interacts with 208 TcoFs, followed by estrogen receptor 1 (ESR1) and

Downloaded from http://nar.oxfordjournals.org/ by guest on January 5, 2017

Table 1. A gene is only considered a TcoF if it satisfies these four conditions

No.	Condition
a	The gene is not characterized as a TF gene.
b	The gene is shown to bind to a known TF. This binding is supported by experiment and is referenced in scientific literature.
c	The gene is annotated in Gene Ontology with GO:0005634 ('nucleus').
d	The gene is annotated in Gene Ontology with GO:0000988 (molecular function: 'transcription factor activity, protein binding') or one of its descendent terms, or with GO:1903506 (biological process: 'regulation of nucleic acid-templated transcription') or one of its descendent terms. In addition, the gene is not annotated with the Gene Ontology term GO:0003677 ('DNA binding').

Table 2. Distribution of human TcoFs among groups

TcoF groups		Evidence for involvement in transcriptional regulation	
		Experimental	Non-experimental
Evidence for location in nucleus	Experimental	389 (40.6%) High confidence	280 (29.2%) Hypothetical (Class II)
	Non-experimental	83 (8.7%) Hypothetical (Class I)	206 (21.5%) Hypothetical (Class III)

Table 3. Distribution of mouse TcoFs among groups

TcoF groups		Evidence for involvement in transcriptional regulation	
		Experimental	Non-experimental
Evidence for location in nucleus	Experimental	136 (32.5%) High confidence	93 (22.3%) Hypothetical (Class II)
	Non-experimental	43 (10.3%) Hypothetical (Class I)	146 (34.9%) Hypothetical (Class III)

jun proto-oncogene (JUN) which are interacting with 171 TcoFs, each. The mouse TF with the highest number of interactions is high mobility group AT-hook 2 (Hmga2) with 44 interactions to TcoFs, followed by forkhead box P3 (Foxp3) with 43 interactions to TcoFs. On average, based on the data in TcoF-DB v2, a human TF is interacting with ∼5.9 TcoFs, whereas a mouse TF is on average interacting with ∼0.9 TcoFs. Because ∼36% of human (∼71% mouse) TFs do not have any interaction to TcoFs, the median number of interactions of human TFs to TcoFs is only 1 (and for mouse the median is 0 as more than half of the TFs do not have any interaction to TcoFs). The differences in human and mouse might reflect differences in available annotations, specifically on interactions. For example, the databases we used to extract experimentally supported interactions to TFs contain ∼6.5 times as many TF interactions for humans as opposed to mouse.

Using our method we extracted 958 human (418 mouse) TcoFs (see Tables 2 and 3). Supplementary Table S3 shows the top 5 human and mouse TcoFs in terms of number of interactions to TFs. The human TcoF with the most interactions to TFs is histone deacetylase 1 (HDAC1) with 238 interactions to a TF. For mouse, the TcoF embryonic ectoderm development (Eed) is interacting with 64 TFs. On average a human (mouse) TcoF is interacting with ∼10.8 (∼3.3) TFs. The median number of interactions for human TcoF interactions to TFs is 6, whereas for mouse is 2. In total TcoF-DB v2 collects 10 358 (1 372) interactions between human (mouse) TFs and TcoFs. However, 126 human (69 mouse) TcoFs are responsible for over half the interactions. In Supplementary Figure S4, a clustered association map shows the interactions between human TFs and TcoFs. Interestingly, we can identify clear groups of TFs interacting with groups of TcoFs as is expected, e.g. mediator complex subunit TcoFs or nuclear receptor coactivator TFs (see Supplementary Figure S4). In summary, these statistics highlight the big influence of non-DNA binding TcoFs on transcriptional regulation in general.

Data integration, web interface, new features and utilization of TcoF-DB v2

We extracted information on the association between genes and biological annotation from GO, Reactome (36) and KEGG (39) pathways; cancer and immunological molecular signatures from MSigDB (40); as well as, disease information from DisGeNET (41) and MGI (42). Data for gene to GO annotation was downloaded from NCBI, while associations for genes to other biological annotations was downloaded from the respective sources themselves. We parsed the data to extract all associations of a concept and one of our TFs and TcoFs. This information was integrated into TcoF-DB v2 and is accessible through the web-interface.

The navigation panel on the left-hand site of TcoF-DB v2 allows for browsing particular classes of TcoFs and TFs. Users can browse biological annotations to identify specific TcoFs and TFs associated with them. For each annotation type, we list individual concepts and show the numbers of TFs and TcoFs associated to each. Following the link to one concept, a user can easily identify all TFs and TcoFs that are associated to a particular concept of interest. A global search box is located on the top of the navigation panel. This search box allows a user to search for genes and proteins of interest. All TFs and TcoF are integrated in TcoF-DB v2 in a gene-centric manner. However, from individual TF or TcoF pages, information on all associated proteins from UniProt can be accessed (43), and the search allows for searching genes as well as protein identifiers.

When browsing a particular TcoF class through one of the links in the navigation panel, one can now make use of

Downloaded from http://nar.oxfordjournals.org/ by guest on January 5, 2017

a new feature that lets a user select subsets of genes in the table, and run a gene-set enrichment analysis with the selected set of genes (44) through a button in the upper right corner of the table. Thus, a user interested in a particular set of TcoF and TF genes is able to quickly gain an overview of the biological concepts the set is enriched for.

When selecting a particular gene in one of the TcoF-DB v2 tables for further investigation (through clicking on the arrow to the left or the gene symbol), one is directed to a new gene view. This view consists of two tabs, the 'Information'- and 'Interactions'-tab. On the 'Information'-tab one can quickly gain information on the gene itself. Importantly, we link now to many more relevant external sources. For example, in addition to general gene information from NCBI Gene (https://www.ncbi.nlm.nih.gov/gene/), we also link to pathway information and expression data through the FANTOM5 SSTAR and EBI expression atlas (45,46) and to ZENBU and UCSC genome browsers (47,48), to be able to study the genes genomic context. If the gene is encoding a TF, we link to TF binding site information for the gene in HOCOMOCO and Factorbook.org (49,17). Importantly, in case the gene has been classified as a TcoF, one can gain information on the GO-terms that led to its TcoF classification. Finally, we show all gene-associated proteins from UniProt (43) and link for each protein to relevant external sources, like PDB (50) and PFAM (51).

The second 'Interactions'-tab shows, for the considered gene, tables of interacting TFs and TcoFs (in case the gene under consideration is a TF). For each individual entry in the table of interactions, we show the relevant PSI-MI entries of experiment and interaction types and link them to the original source, e.g. interaction database, and literature that describes the interaction. In this way, one can quickly gain an overview of the interactions of a particular gene to other genes, and use the PSI-MI entries to judge their relevance and use the literature link to get more information on the performed experiment.

TcoF-DB v2 was build using the DJANGO framework (https://www.djangoproject.com/), with a MySQL (https://www.mysql.com/) database in the backend. TcoF-DB v2 can be accessed at http://tcofdb.org and http://www.cbrc.kaust.edu.sa/tcofdb2, while the source-code is freely available at https://gitlab.com/s-schmeier/tcof. All relevant tables in the web-interface can be downloaded in excel-format.

Example of application

In this section, we illustrate the reliability of the data housed in TcoF-DB v2. For this example, we will look for TcoFs implicated in rheumatoid arthritis. We expand the 'TcoFs/TFs by annotation' on the left sidebar that allows search by 'Gene Ontology', 'Pathways', 'MSigDB' or 'Diseases', select 'Diseases' and search for 'rheumatoid arthritis'. The search retrieves 32 TFs and ten TcoFs associated with the disease. Selecting 'Rheumatoid Arthritis (human)' gives the user an overview of the associated TFs and TcoFs. Note that the categorization of TcoFs into different classes for this case based on GO evidences is provided in Supplementary Table S1. They include: dihydrofolate reductase (DHFR), hematopoietic cell-specific Lyn substrate 1 (HCLS1), interleukin 1 receptor associated kinase 1 (IRAK1), protein kinase C theta (PRKCQ), transducin like enhancer of split 3 (TLE3), TNF receptor associated factor 6 (TRAF6), stromal antigen 1 (STAG1), thioredoxin interacting protein (TXNIP), peptidylarginine deiminase type IV (PADI4), and PHD finger protein 19 (PHF19). Note that PADI4 and DHFR are not yet classified as high-confidence TcoFs. However, we do not consider them false-positives as the existing evidence points towards a possibility for their involvement in transcriptional regulation and activity in the nucleus. Both are classified as hypothetical TcoFs (DHFR of class 3; PADI4 of class 2).

Suzuki *et al.* demonstrated the expression of PADI4 in hematological and rheumatoid arthritis synovial tissues and that the PADI4 haplotype associated with susceptibility to rheumatoid arthritis increases production of autoantigens, citrullinated peptides; thus autoimmunity to PADI4 precedes clinical onset of rheumatoid arthritis (52). A 70kb region that includes a portion of PHF19, all of TRAF1, and the majority of the TRAF1-C5 intergenic region, has been identified as a rheumatoid arthritis-susceptibility site (53). Rheumatoid arthritis risk loci have been identified in PADI4, IRAK1, TLE3, PRKCQ and TRAF6 (54). TXNIP is a known key regulatory molecule of cartilage destruction in rheumatoid arthritis while HCLS1 was shown to be involved in a rheumatoid arthritis-specific mechanism (55,56). In addition, Methotrexate (MTX) is used for the treatment of rheumatoid arthritis and polymorphisms, within genes (that include DHFR) in the purine biosynthesis pathway, were identified as potential biomarkers in predicting treatment effectiveness of MTX (57). Finally, STAG1, PADI4 and PHF19 have been linked to rheumatoid arthritis phenotypes in a GWAS datasets from the GWAS Catalog SNP-Phenotype Associations dataset (http://amp.pharm.mssm.edu/Harmonizome/gene_set/Rheumatoid+arthritis/GWAS+Catalog+SNP-Phenotype+Associations). Thus, all ten TcoFs that have been suggested by TcoF-DB v2 have reported implications in 'rheumatoid arthritis' in the scientific literature. Thus, we believe that insights gained using TcoF-DB v2 may help in many cases of studies related to gene regulation (references and evidence for the involvement of these genes/proteins in rheumatoid arthritis are given in Supplementary Table S1 in Supplementary Material).

CONCLUSIONS

TcoF-DB v2 includes several key improvements that make the resource far more usable. We expanded substantially the number of human TcoFs, mostly due to new confirmed interactions. We also enhanced the database by adding mouse TFs and TcoFs to enable transcriptional regulatory research for rodents.

We integrated additional biological annotation data in the TcoF-DB v2 for TFs and TcoFs, allowing the user to identify TF and TcoFs that may be important for a particular biological concept (e.g. list all TF or TcoFs that are part of a particular pathway or associated to a particular disease). We also added the possibility to run customized gene-set enrichment analyses of a user-defined set of TFs and TcoFs, which allows the user to investigate the gene set

Downloaded from http://nar.oxfordjournals.org/ by guest on January 5, 2017

influence on particular biological concepts like pathways, diseases, etc. Although this resource is focused on human and mouse organisms, in future updates we intend to apply the similar strategy to other eukaryotes, such as other vertebrates, insects, plants and algae and include the relevant information in TcoF-DB. The information that would be included will depend on the organism. For example, for algae one may be more interested in metabolic pathways that algae can produce and the transcriptional control of genes encoding the necessary enzymes.

Taken together, TcoF-DB v2 represents a valuable and easy to navigate resource for studying the effects of TcoFs on transcriptional regulation via interactions with TFs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

King Abdullah University of Science and Technology (KAUST) base research fund (to T.A. and V.B.B.); KAUST Office of Sponsored Research (OSR) [URF/1/1976-02 to M.E.]. Funding for open access charge: KAUST.
Conflict of interest statement. None declared.

REFERENCES

1. Lenhard,B., Sandelin,A. and Carninci,P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.*, **13**, 233–245.
2. Sandelin,A., Carninci,P., Lenhard,B., Ponjavic,J., Hayashizaki,Y. and Hume,D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev.*, **8**, 424–436.
3. Latchman,D.S. (1997) Transcription factors: an overview. *Int. J. Biochem. Cell Biol.*, **29**, 1305–1312.
4. Forrest,A., Kawaji,H., Rehli,M., Baillie,K., de Hoon,M., Haberle,V., Lassmann,T., Kulakovskiy,I., Lizio,M., Itoh,M. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462–470.
5. Vaquerizas,J.M., Kummerfeld,S.K., Teichmann,S. a. and Luscombe,N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
6. Roeder,R.G. (1998) Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 201–218.
7. Gamble,M.J. and Freedman,L.P. (2002) A coactivator code for transcription. *Trends Biochem Sci*, **27**, 165–167.
8. Xu,L., Glass,C.K. and Rosenfeld,M.G. (1999) Coactivator and corepressor complexes in nuclear receptor function. *Curr Opin Genet Dev*, **9**, 140–147.
9. Naar,A.M., Lemon,B.D. and Tjian,R. (2001) Transcriptional Coactivator Complexes. *Annu. Rev. Biochem.*, **70**, 475–501.
10. Ravasi,T., Suzuki,H., Cannistraci,C.V., Katayama,S., Bajic,V.B., Tan,K., Akalin,A., Schmeier,S., Kanamori-Katayama,M., Bertin,N. *et al.* (2010) An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, **140**, 744–752.
11. Martinez,E. (2002) Multi-protein complexes in eukaryotic gene transcription. *Plant Mol. Biol.*, **50**, 925–947.
12. Thomas,M.C. and Chiang,C.-M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.*, **41**, 105–178.
13. Schmeier,S., Jankovic,B. and Bajic,V.B. (2011) Simplified method to predict mutual interactions of human transcription factors based on their primary structure. *PLoS One*, **6**, e21887.
14. Piatek,M.J., Schramm,M.C., Burra,D.D., Binshbreen,A., Jankovic,B.R., Chowdhary,R., Archer,J.A.C. and Bajic,V.B. (2013) Simplified method for predicting a functional class of proteins in transcription factor complexes. *PLoS One*, **8**, e68857.
15. Mathelier,A., Fornes,O., Arenillas,D.J., Chen,C.-Y., Denay,G., Lee,J., Shi,W., Shyr,C., Tan,G., Worsley-Hunt,R. *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **44**, D110–D115.
16. Matys,V., Kel-Margoulis,O. V, Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
17. Kulakovskiy,I. V, Vorontsov,I.E., Yevshin,I.S., Soboleva,A. V, Kasianov,A.S., Ashoor,H., Ba-Alawi,W., Bajic,V.B., Medvedeva,Y.A., Kolpakov,F.A. *et al.* (2016) HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.*, **44**, D116–D125.
18. Fulton,D.L., Sundararajan,S., Badis,G., Hughes,T.R., Wasserman,W.W., Roach,J.C. and Sladek,R. (2009) TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol.*, **10**, R29.
19. Wilson,D., Charoensawan,V., Kummerfeld,S.K. and Teichmann,S.A. (2008) DBD–taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.*, **36**, D88–D92.
20. Lee,A.P., Yang,Y., Brenner,S. and Venkatesh,B. (2007) TFCONES: a database of vertebrate transcription factor-encoding genes and their associated conserved noncoding elements. *BMC Genomics*, **8**, 441.
21. Chawla,K., Tripathi,S., Thommesen,L., Lægreid,A. and Kuiper,M. (2013) TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics*, **29**, 2519–2520.
22. Kanamori,M., Konno,H., Osato,N., Kawai,J., Hayashizaki,Y. and Suzuki,H. (2004) A genome-wide and nonredundant mouse transcription factor database. *Biochem. Biophys. Res. Commun.*, **322**, 787–793.
23. Kel-Margoulis,O. V, Romashchenko,A.G., Kolchanov,N.A., Wingender,E. and Kel,A.E. (2000) COMPEL: a database on composite regulatory elements providing combinatorial transcriptional regulation. *Nucleic Acids Res.*, **28**, 311–315.
24. Roy,S., Schmeier,S., Arner,E., Alam,T., Parihar,S.P., Ozturk,M., Tamgue,O., Kawaji,H., de Hoon,M.J.L., Itoh,M. *et al.* (2015) Redefining the transcriptional regulatory dynamics of classically and alternatively activated macrophages by deepCAGE transcriptomics. *Nucleic Acids Res.*, **43**, 6969–6982.
25. Roy,S., Guler,R., Parihar,S.P., Schmeier,S., Kaczkowski,B., Nishimura,H., Shin,J.W., Negishi,Y., Ozturk,M., Hurdayal,R. *et al.* (2015) Batf2/Irf1 Induces Inflammatory Responses in Classically Activated Macrophages, Lipopolysaccharides, and Mycobacterial Infection. *J. Immunol.*, **194**, 6035–6044.
26. Suzuki,H., Forrest,A.R.R., van Nimwegen,E., Daub,C.O., Balwierz,P.J., Irvine,K.M., Lassmann,T., Ravasi,T., Hasegawa,Y., de Hoon,M.J.L. *et al.* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.*, **41**, 553–562.
27. Schmeier,S., MacPherson,C.R., Essack,M., Kaur,M., Schaefer,U., Suzuki,H., Hayashizaki,Y. and Bajic,V.B. (2009) Deciphering the transcriptional circuitry of microRNA genes expressed during human monocytic differentiation. *BMC Genomics*, **10**, 595.
28. Lee,T.I., Young,R.A., Abramson,J., Giraud,M., Benoist,C., Mathis,D., Adelman,K., Lis,J.T., Aguilo,F., Zhou,M.M. *et al.* (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
29. Engelkamp,D. and van Heyningen,V. (1996) Transcription factors in disease. *Curr. Opin. Genet. Dev.*, **6**, 334–342.
30. Kaur,M., MacPherson,C.R., Schmeier,S., Narasimhan,K., Choolani,M. and Bajic,V.B. (2011) In Silico discovery of transcription factors as potential diagnostic biomarkers of ovarian cancer. *BMC Syst. Biol.*, **5**, 144.
31. Schaefer,U., Schmeier,S. and Bajic,V. (2011) TcoF-DB: dragon database for human transcription co-factors and transcription factor interacting proteins. *Nucleic Acids Res.*, **39**, D106–D110.
32. 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.

Downloaded from http://nar.oxfordjournals.org/ by guest on January 5, 2017

33. Consortium, Gene Ontology (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
34. Chatr-Aryamontri,A., Breitkreutz,B.-J., Oughtred,R., Boucher,L., Heinicke,S., Chen,D., Stark,C., Breitkreutz,A., Kolas,N., O'Donnell,L. *et al.* (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*, **43**, D470–D478.
35. Licata,L. and Orchard,S. (2016) The MIntAct Project and Molecular Interaction Databases. *Methods Mol. Biol.*, **1415**, 55–69.
36. Fabregat,A., Sidiropoulos,K., Garapati,P., Gillespie,M., Hausmann,K., Haw,R., Jassal,B., Jupe,S., Korninger,F., McKay,S. *et al.* (2016) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **44**, D481–D487.
37. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
38. Orchard,S. and Kerrien,S. (2010) Molecular interactions and data standardisation. *Methods Mol. Biol.*, **604**, 309–318.
39. 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.
40. Liberzon,A., Birger,C., Thorvaldsdóttir,H., Ghandi,M., Mesirov,J.P. and Tamayo,P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
41. Piñero,J., Queralt-Rosinach,N., Bravo,À., Deu-Pons,J., Bauer-Mehren,A., Baron,M., Sanz,F. and Furlong,L.I. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford).*, **2015**, bav028.
42. Bult,C.J., Eppig,J.T., Blake,J.A., Kadin,J.A., Richardson,J.E. and Mouse Genome Database Group (2016) Mouse genome database 2016. *Nucleic Acids Res.*, **44**, D840–D847.
43. UniProt Consortium, T.U. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
44. Kuleshov,M. V, Jones,M.R., Rouillard,A.D., Fernandez,N.F., Duan,Q., Wang,Z., Koplev,S., Jenkins,S.L., Jagodnik,K.M., Lachmann,A. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
45. Petryszak,R., Keays,M., Tang,Y.A., Fonseca,N.A., Barrera,E., Burdett,T., Füllgrabe,A., Fuentes,A.M.-P., Jupp,S., Koskinen,S. *et al.* (2016) Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.*, **44**, D746–D752.
46. Abugessaisa,I., Shimoji,H., Sahin,S., Kondo,A., Harshbarger,J., Lizio,M., Hayashizaki,Y., Carninci,P., Forrest,A., FANTOM consortium *et al.* (2016) FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki. *Database (Oxford)*, **2016**, doi:10.1093/database/baw10.
47. Speir,M.L., Zweig,A.S., Rosenbloom,K.R., Raney,B.J., Paten,B., Nejad,P., Lee,B.T., Learned,K., Karolchik,D., Hinrichs,A.S. *et al.* (2016) The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.*, **44**, D717–D725.
48. Severin,J., Lizio,M., Harshbarger,J., Kawaji,H., Daub,C.O., Hayashizaki,Y., Bertin,N., Forrest,A.R.R. and FANTOM Consortium (2014) Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nat. Biotechnol.*, **32**, 217–219.
49. Wang,J., Zhuang,J., Iyer,S., Lin,X.-Y., Greven,M.C., Kim,B.-H., Moore,J., Pierce,B.G., Dong,X., Virgil,D. *et al.* (2013) Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res.*, **41**, D171–D176.
50. Zardecki,C., Dutta,S., Goodsell,D.S., Voigt,M. and Burley,S.K. (2016) RCSB Protein Data Bank: a resource for chemical, biochemical, and structural explorations of large and small biomolecules.
51. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
52. Suzuki,A., Yamada,R., Chang,X., Tokuhiro,S., Sawada,T., Suzuki,M., Nagasaki,M., Nakayama-Hamada,M., Kawaida,R., Ono,M. *et al.* (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.*, **34**, 395–402.
53. Chang,M., Rowland,C.M., Garcia,V.E., Schrodi,S.J., Catanese,J.J., van der Helm-van Mil,A.H., Ardlie,K.G., Amos,C.I., Criswell,L.A., Kastner,D.L. *et al.* (2008) A large-scale rheumatoid arthritis genetic study identifies association at chromosome 9q33.2. *PLoS Genet.*, **4**, e1000107.
54. Eyre,S., Bowes,J., Diogo,D., Lee,A., Barton,A., Martin,P., Zhernakova,A., Stahl,E., Viatte,S., McAllister,K. *et al.* (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.*, **44**, 1336–1340.
55. Andreas,K., Lübke,C., Häupl,T., Dehne,T., Morawietz,L., Ringe,J., Kaps,C. and Sittinger,M. (2008) Key regulatory molecules of cartilage destruction in rheumatoid arthritis: an in vitro study. *Arthritis Res. Ther.*, **10**, R9.
56. Lequerré,T., Bansard,C., Vittecoq,O., Derambure,C., Hiron,M., Daveau,M., Tron,F., Ayral,X., Biga,N., Auquit-Auckbur,I. *et al.* (2009) Early and long-standing rheumatoid arthritis: distinct molecular signatures identified by gene-expression profiling in synovia. *Arthritis Res. Ther.*, **11**, R99.
57. Lee,Y.C., Cui,J., Costenbader,K.H., Shadick,N.A., Weinblatt,M.E. and Karlson,E.W. (2009) Investigation of candidate polymorphisms and disease activity in rheumatoid arthritis patients on methotrexate. *Rheumatology (Oxford).*, **48**, 613–617.

Downloaded from http://nar.oxfordjournals.org/ by guest on January 5, 2017