TcoF-DB: dragon database for human transcription co-factors and transcription factor interacting proteins

Ulf Schaefer, Sebastian Schmeier and Vladimir B. Bajic*

Computational Bioscience Research Center (CBRC), 4700 King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

ABSTRACT

The initiation and regulation of transcription in eu-karyotes is complex and involves a large number of transcription factors (TFs), which are known to bind to the regulatory regions of eukaryotic DNA. Apart from TF–DNA binding, protein–protein interaction involving TFs is an essential component of the machinery facilitating transcriptional regulation. Proteins that interact with TFs in the context of transcription regulation but do not bind to the DNA themselves, we consider transcription co-factors (TcoFs). The influence of TcoFs 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. While the role of TFs and their interaction with regulatory DNA regions has been well-studied, the association between TFs and TcoFs has so far been given less attention. Here, we present a resource that is comprised of a collection of human TFs and the TcoFs with which they interact. Other proteins that have a proven interaction with a TF, but are not considered TcoFs are also included. Our database contains 157 high-confidence TcoFs and additionally 379 hypothetical TcoFs. These have been identified and classified according to the type of available evidence for their involvement in transcriptional regulation and their presence in the cell nucleus. We have divided TcoFs into four groups, one of which contains high-confidence TcoFs and three others contain TcoFs which are hypothetical to different extents. We have developed the Dragon Database for Human Transcription Co-Factors and Transcription Factor Interacting Proteins (TcoF-DB). A web-based interface for this resource can be freely accessed at http://cbrc.kaust.edu.sa/tcof/ and http://apps.sanbi.ac.za/tcof/.

INTRODUCTION

It is known that initiation and regulation of transcription in eukaryotic cells is a complex process. It is widely accepted that it constitutes a vastly complex process involving a large number of proteins interacting with the regulatory regions of DNA in various ways (1–5). Additionally, the process involves numerous proteins that do not themselves interact with the DNA molecule, but instead interact with DNA binding proteins, forming complexes that could themselves interact with DNA (6,7). This phenomenon significantly increases the complexity of the process of transcriptional regulation and initiation. Proteins that bind DNA and in this way affect the regulation of transcription are in this context referred to as transcription factors (TFs) (8). At the same time proteins that are not DNA binding in the context of transcriptional regulation, but are nevertheless involved in this process through the interaction with TFs are referred to as transcription co-factors (TcoFs) (9,10). The functions that TcoFs are known to have are multiple and include, among others, signal transmission, modulation of TF–DNA binding and chromatin modification (11,12).

Apart from the interaction of TFs with the regulatory regions of the DNA molecule, the interactions between proteins (protein–protein interaction, PPI) are of importance for understanding the mechanisms of transcriptional regulation. PPIs include various forms of relationships between two proteins, e.g. the modification of one protein by another protein, the transport of one protein from one cellular location to another by another protein or the formation of a protein complex by two or more proteins, etc. PPIs of our interest that are important for the processes of transcriptional regulation include the physical interaction between two TFs, the physical

interaction between a TF and a TcoF as well as potentially the physical interaction between a TF and another protein that cannot be regarded as a TcoF.

The specific influence of TFs on transcription is widely accepted to be an important component in gene expression (13). It has been recognized that TFs can have an influence on disease development (14) and they have been previously identified as potential targets in cancer therapy (15–18) or cancer diagnostic biomarkers (19). The functionality of TFs has also been shown to be important for the development of treatments for a number of other human diseases (20–23). However, research on the interaction of TFs with other proteins in relation to transcription regulation has so far not been that intensive. It has previously been shown that PPI can be linked to organism complexity (24). It has also been shown that the same TF can have opposite functions depending on its interactions with TcoFs (25). We therefore believe that in order to get a more complete insight into the complexity of transcriptional regulatory processes and more accurately describe the role of TFs and TcoFs in relation to human diseases it is imperative to investigate TFs in conjunction with their interacting proteins.

With this in mind, we developed a public resource: Dragon Database for Human Transcription Co-Factors and Transcription Factor Interacting Proteins (TcoF-DB). With this database, we wish to provide a comprehensive starting point for investigations into transcriptional regulation that include PPIs between TFs, TFs and TcoFs, as well as between TFs and other proteins. While there are numerous resources available that deal with either the complete human proteome (26–29), human PPI (30–33) or mammalian TFs (34–36), we are not aware of a resource that unifies the data necessary to directly investigate human TFs, human TcoFs and other human proteins that interact with TFs. TcoF-DB attempts to reduce this gap.

DATA INTEGRATION AND ANALYSIS

We generated a highly accurate set of human TFs. For this, we rely on data that was previously published by Vaquerizas *et al.* (13) and that we regard to be a gold-standard due to the meticulous way in which it was created. We extracted all TFs from that reference. This set constitutes all human proteins that possess a sequence specific DNA binding domain and additionally have passed a step of manual curation during which each protein was examined by a human curator for being a TF with high confidence. In addition to this, we extracted 70 proteins that are included in TRANSFAC 11.4 (34) but were not included in the list published by Vaquerizas *et al.* We manually checked each of these proteins and added 19 of them to our list of human TFs. To complement our set of TFs further, we finally extracted mouse TF genes from TFCAT (37) whose protein products bind DNA in a sequence specific as well as a non-sequence specific manner. For these mouse genes, we identified human ortholog genes using NCBI's Homologene (38) and their gene products, if available. We manually checked each of

these proteins that were not already included in our set and added eight proteins from TFCAT to our TF list. In total, we identified 1365 human TFs in this way. It should be highlighted that each TF in this list has been hand curated at some point during the process.

Four public databases of human PPIs were used to extract PPIs. These were MINT (accessed July 2010) (39), IntAct (accessed July 2010) (31), BioGRID version 3.0.67 (30) and Reactome version 33 (40). These databases were selected because they allow for the download of data in the format suitable for processing by a computer. All data in these four databases is experimentally confirmed. It was of particular importance that this information is presented in the PSI-MI format [molecular interaction standard of the Proteomics Standards Initiative (41)] in order to allow us to focus on PPI of a certain type. We only consider PPIs that represent physical interactions between two proteins. Thus, considered PSI-MI interaction types were one of the following:

- MI:0195 (covalent binding),
- MI:0407 (direct interaction),
- MI:0915 (physical association).

In this manner, we were able to extract 7045 unique interactions between two proteins where at least one of the participants is a TF from our list and the interaction is one of the above mentioned physical interactions.

All proteins that were identified as having a physical interaction with a known TF and are not themselves included in our reference list of TFs are initially considered 'TF interacting proteins'. Evidence for the interaction is extracted from one of the aforementioned databases and added to our database. In total, 2300 distinct human proteins have been found to interact with one of 712 TF. This leaves 653 TFs from our list for which we were not able to find a known interaction with another protein.

Out of these 2300 proteins, we subsequently endeavoured to identify those proteins that can be considered TcoFs. First, we postulated the necessary condition that a TcoF candidate is known to be located in the cell nucleus. For this, we check whether the protein in question is annotated in Gene Ontology (GO) (42) with term GO:0005634 (cellular component: 'nucleus'). For this we use AmiGO version 1.7 (43). We do not consider proteins that do not possess this annotation as TcoFs. Second, we require that a TcoF candidate is known to be involved in transcriptional regulation. For this condition, we check whether the protein in questions is annotated in GO with either GO:0030528 (molecular function: 'transcription regulator activity') or one of its descendent terms, or with GO:0045449 (biological process: 'regulation of transcription') or one of its descendent terms. All proteins that do not have one of these annotations in either the molecular function or the biological process ontology of GO also are not considered TcoFs. In summary, we only consider a protein to be a TcoF if it satisfies all conditions shown in Table 1.

On the other hand, proteins that satisfy both conditions 1 and 2 but do not satisfy both conditions 3 and 4 are

considered 'TF binding proteins', but not TcoF. In total, 529 TcoFs were identified in this way. This leaves 1771 proteins to be characterized as 'TF binding proteins'.

Since this way of characterizing TcoFs is heavily dependent on protein annotations in GO, we subsequently divided the TcoFs into different groups based on the type of evidence that is given for each relevant annotation in GO. For this purpose, we consider two groups of evidence given in GO: experimental evidence consisting of the evidence codes EXP, IDA, IMP, IGI, IEP and IPI and non-experimental evidence consisting of all other evidence codes (see http://www.geneontology.org/GO.evidence.shtml for details on evidence in GO). We then divide the TcoFs into the following four groups, based on the type of evidence that is present for the fulfilment of conditions 3 and 4 from Table 1:

- 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 leads to a distribution of TcoFs among the four groups as can be seen in Table 2.

Only TcoFs that have experimental evidence cited for at least one GO annotation relevant to the regulation of transcription and have experimental evidence cited for occurrence in the cell nucleus are considered TcoFs with high confidence. If experimentally confirmed evidence for one of these is missing the protein is considered to be a hypothetical TcoF. The level of confidence for a hypothetical TcoF decreases the higher the number of the hypothetical TcoF class.

In a final step, all data have been incorporated in a relational database (PostgreSQL version 8.4) and a simple web-interface has been implemented. The database can be accessed through http://cbrc.kaust.edu.sa/tcof/ or http://apps.sanbi.ac.za/tcof/. The interface allows for the search of human proteins by their name or Uniprot identifiers (Uniprot ID or accession number). Alternative names for proteins, as given in the Uniprot Knowledge Base (26), are included in the name search wherever possible. The possibility of viewing predefined sets of proteins like TFs and TcoFs also exists. The viewing of TcoFs can be narrowed by viewing only TcoFs of certain groups. The results page will give a comprehensive overview of all interactions. For each TF it is shown what other TFs it interacts with and what interactions with TcoFs and other interacting proteins are evident. For each TcoF and for each other protein interacting with a TF, a list of TFs that it interacts with is displayed. On each page outlining protein details the evidence for a protein's status is also displayed.

Online help and a user manual are also available through the aforementioned URL. These will aid the user in successfully exploiting this database resource. In addition, we provide the core data of our database for download in a format that can be easily processed by a computer. The database will be updated annually.

DISCUSSION

To the best of our knowledge, this database constitutes the only currently available resource that provides comprehensive information about PPIs of human TFs, human TcoF and other human proteins that interact with human TFs. While there are numerous resources regarding PPI publicly available (44), none of them provides information specifically related to protein involvement in transcriptional regulation, and neither do they provide any mechanism to directly search for such information.

The TcoF-DB database is developed with the understanding that a TcoF is a protein that has a proven binding interaction with a known TF, but does itself not directly bind to the regulatory DNA region. A TcoF is

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

No.	Condition
1	The protein is not characterized as a TF.
2	It is shown to bind to a known transcription factor. This binding was demonstrated by experiment and is referenced in scientific literature.
3	It is annotated in Gene Ontology with GO:0005634 ('nucleus').
4	It is annotated in Gene Ontology with GO:0030528 ('transcription regulator activity') or one of its descendent terms or with GO:0045449 ('regulation of transcription') or one of its descendent terms.

Table 2. Distribution of TcoFs among groups

TcoF groups		Evidence for involvement in transcription regulation	
		Experimental (%)	Non-experimental (%)
Evidence for location in cell nucleus	Experimental	155 (29.3) High confidence	96 (18.1) Hypothetical (Class2)
	Non-experimental	62 (11.7) Hypothetical (Class 1)	216 (40.8) Hypothetical (Class 3)

also required to be reported to be involved in the regulation of transcription. While it is true that the mechanism of transcription regulation involves larger protein complexes, we have concentrated our efforts on interactions between two proteins one of which is a known TF. The method by which we have chosen TFs provides a high accuracy of the TF set. We rely on a previously published set of TFs that was manually curated and we added a small number of manually curated TFs to this set. Thus, all TFs in our set have been hand-curated at some stage during the process of establishing our set of known TF.

Although, we have applied stringent requirements for a protein to be considered a TcoF, we still rely on annotations in GO for determining whether a protein is involved in the regulation of transcription. We are aware that GO is in some cases unreliable due to incomplete or erroneous annotations, but we are not aware of an equally rich data source for functional annotations for human proteins. In order to enable users to judge the reliability of the annotations of a given protein, we have included detailed information about the evidence cited for each annotation used to classify a protein as a TcoF, including GO evidence code, evidence type and a reference pointing to the data source. Thus, it is possible for the user to quickly distinguish between TcoFs whose classification solely depends on non-experimental evidence such as automatic annotations or author statements from those whose annotation has been experimentally confirmed.

It must also be said that the reliability of the identification of TcoFs that we performed is also dependent on the availability of PPI data in one of the resources we utilized. Missing or erroneous data in these resources will inevitably lead to the misclassification of TcoFs in TcoF-DB. However, we benefit from using four resources in parallel in the extraction of PPI data.

The average numbers of interactions per TF and TcoF show that interactions happen abundantly in transcriptional regulation. The fact that more than half of the known human TFs have at least one proven PPI underlines the importance of PPI and TcoF for the understanding of transcriptional regulation. The arithmetic mean for the number of interactions per TFs is 1.8 for TF–TcoFs interactions, 1.7 for TF–TF interactions and 2.7 for interactions between TFs and other proteins. Because for each of these interaction types more than half our TFs do not have an interaction, the median values for these interactions are all zero. This means that the average TF interacts with 1.7 other TFs, 1.8 TcoFs and 2.7 other proteins. At the same time more than half of the TFs do not have an interaction with another protein of a given type. Considering all protein types (TFs, TcoFs and other proteins) 712 out of 1365 TFs have an interaction. The difference between the arithmetic mean and the median values described above also illustrates the fact that there are a number of TFs which are very promiscuous regarding PPI. The TF 'cellular tumour antigen p53' is one of the best-studied human TFs (13). It is also one of the most promiscuous TFs regarding reported PPI. For this TF, we report 313 unique interacting proteins, 35 other TFs, 99 TcoFs (31 out of the high-confidence group) and 179 other proteins. The human TF 'TATA-box binding protein' is also promiscuous, interacting with 71 TcoFs (17 out of the high-confidence group), 33 other TFs and 26 other proteins. These TFs are two of the most interacting TFs in our database. With 1365 TFs and 529 TcoFs (155 high-confidence TcoFs, 374 hypothetical TcoFs) indentified in this database, we can confirm a previous estimate that ∼10% of the human proteome is directly involved in transcriptional regulation (45). With human TFs having a total of 7045 interactions with other human proteins, the number of TF–protein complexes possibly involved in transcription is significantly larger than the number of TFs.

One has to keep in mind that our classification of TcoFs depends on annotation data which in turn depends on experiments and data submissions. Such data is often inaccurate and incomplete, frequently focusing only on a specific cell location or specific function while alternative cell locations in which the protein expresses and/or alternative protein functions remain elusive. For this reason, proteins that have a proven interaction with a known TF, but are not classified as a TcoF in our database, because they lack the annotation to support such a claim, play an important role. The reason for this is that in some cases they might represent candidates for proteins with roles in transcriptional regulation.

REFERENCES

1. 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.
2. Lemon,B. and Tjian,R. (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.*, **14**, 2551–2569.
3. Lee,T.I. and Young,R.A. (2000) Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.*, **34**, 77–137.
4. Nikolov,D.B. and Burley,S.K. (1997) RNA polymerase II transcription initiation: a structural view. *Proc. Natl Acad. Sci. USA*, **94**, 15–22.
5. Mitchell,P.J. and Tjian,R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371–378.
6. Gamble,M. (2002) A coactivator code for transcription. *Trends Biochem. Sci.*, **27**, 165–167.
7. 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.
8. Latchman,D.S. (1997) Transcription factors: an overview. *Int. J. Biochem. Cell Biol.*, **29**, 1305–1312.
9. Gill,G. (2001) Regulation of the initiation of eukaryotic transcription. *Essays Biochem.*, **37**, 33–43.
10. Xu,L., Glass,C. and Rosenfeld,M. (1999) Coactivator and corepressor complexes in nuclear receptor function. *Curr. Opin. Genet. Dev.*, **9**, 140–147.
11. Thomas,M.C. and Chiang,C. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.*, **41**, 105–178.

12. Martinez,E. (2002) Multi-protein complexes in eukaryotic gene transcription. *Plant Mol. Biol.*, **50**, 925–947.
13. 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.
14. Engelkamp,D. and van Heyningen,V. (1996) Transcription factors in disease. *Curr. Opin. Genet. Dev.*, **6**, 334–342.
15. Mauro,C., Zazzeroni,F., Papa,S., Bubici,C. and Franzoso,G. (2009) The NF-kappaB transcription factor pathway as a therapeutic target in cancer: methods for detection of NF-kappaB activity. *Methods Mol. Biol.*, **512**, 169–207.
16. Morgan,R., Plowright,L., Harrington,K.J., Michael,A. and Pandha,H.S. (2010) Targeting HOX and PBX transcription factors in ovarian cancer. *BMC Cancer*, **10**, 89.
17. Wang,C., Chen,J.J. and Yang,P. (2006) Multifunctional transcription factor YY1: a therapeutic target in human cancer? *Expert Opin. Ther. Targets*, **10**, 253–266.
18. Lu,H. and Huan,C. (2007) Transcription factor NFAT, its role in cancer development, and as a potential target for chemoprevention. *Curr. Cancer Drug Targets*, **7**, 343–353.
19. Kothandaraman,N., Bajic,V.B., Brendan,P.N., Huak,C.Y., Keow,P.B., Razvi,K., Salto-Tellez,M. and Choolani,M. (2010) E2F5 status significantly improves malignancy diagnosis of epithelial ovarian cancer. *BMC Cancer*, **10**, 64.
20. Cuadrado,A., Moreno-Murciano,P. and Pedraza-Chaverri,J. (2009) The transcription factor Nrf2 as a new therapeutic target in Parkinson's disease. *Expert Opin. Ther. Targets*, **13**, 319–329.
21. Qian,L., Mohapatra,B., Akasaka,T., Liu,J., Ocorr,K., Towbin,J.A. and Bodmer,R. (2008) Transcription factor neuromancer/TBX20 is required for cardiac function in Drosophila with implications for human heart disease. *Proc. Natl Acad. Sci. USA*, **105**, 19833–19838.
22. Desplats,P.A., Lambert,J.R. and Thomas,E.A. (2008) Functional roles for the striatal-enriched transcription factor, Bcl11b, in the control of striatal gene expression and transcriptional dysregulation in Huntington's disease. *Neurobiol. Dis.*, **31**, 298–308.
23. Citron,B.A., Dennis,J.S., Zeitlin,R.S. and Echeverria,V. (2008) Transcription factor Sp1 dysregulation in Alzheimer's disease. *J. Neurosci. Res.*, **86**, 2499–24504.
24. Xia,K., Fu,Z., Hou,L. and Han,J.J. (2008) Impacts of protein-protein interaction domains on organism and network complexity. *Genome Res.*, **18**, 1500–1508.
25. Giangrande,P.H., Kimbrel,E.A., Edwards,D.P. and McDonnell,D.P. (2000) The opposing transcriptional activities of the two isoforms of the human progesterone receptor are due to differential cofactor binding. *Mol. Cell. Biol.*, **20**, 3102–3115.
26. The UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
27. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
28. Pontén,F., Jirström,K. and Uhlen,M. (2008) The human protein atlas–a tool for pathology. *J. Pathol.*, **216**, 387–393.
29. Kandasamy,K., Keerthikumar,S., Goel,R., Mathivanan,S., Patankar,N., Shafreen,B., Renuse,S., Pawar,H., Ramachandra,Y.L., Acharya,P.K. *et al.* (2009) Human Proteinpedia: a unified discovery resource for proteomics research. *Nucleic Acids Res.*, **37**, D773–D781.
30. Breitkreutz,B., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bähler,J., Wood,V. *et al.* (2008) The BioGRID interaction database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
31. Aranda,B., Achuthan,P., Alam-Faruque,Y., Armean,I., Bridge,A., Derow,C., Feuermann,M., Ghanbarian,A.T., Kerrien,S., Khadake,J. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.
32. Pagel,P., Kovac,S., Oesterheld,M., Brauner,B., Dunger-Kaltenbach,I., Frishman,G., Montrone,C., Mark,P., Stümpflen,V., Mewes,H. *et al.* (2005) The MIPS mammalian protein-protein interaction database. *Bioinformatics*, **21**, 832–834.
33. Han,K., Park,B., Kim,H., Hong,J. and Park,J. (2004) HPID: the human protein interaction database. *Bioinformatics*, **20**, 2466–2470.
34. 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.
35. 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.
36. 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.
37. 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.
38. Sayers,E.W., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.
39. Ceol,A., Chatr Aryamontri,A., Licata,L., Peluso,D., Briganti,L., Perfetto,L., Castagnoli,L. and Cesareni,G. (2010) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.*, **38**, D532–D539.
40. Matthews,L., Gopinath,G., Gillespie,M., Caudy,M., Croft,D., de Bono,B., Garapati,P., Hemish,J., Hermjakob,H., Jassal,B. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.
41. Orchard,S. and Kerrien,S. (2010) Molecular interactions and data standardisation. *Methods Mol. Biol.*, **604**, 309–318.
42. The Gene Ontology Consortium (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.*, **38**, D331–D335.
43. Carbon,S., Ireland,A., Mungall,C.J., Shu,S., Marshall,B. and Lewis,S. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.
44. Mathivanan,S., Periaswamy,B., Gandhi,T.K., Kandasamy,K., Suresh,S., Mohmood,R., Ramachandra,Y.L. and Pandey,A. (2006) An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatics*, **7(Suppl 5)**, S19.
45. Babu,M.M., Luscombe,N.M., Aravind,L., Gerstein,M. and Teichmann,S.A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.*, **14**, 283–291.