

Database tool

T2D@ZJU: a knowledgebase integrating heterogeneous connections associated with type 2 diabetes mellitus

Zhenzhong Yang[†], Jihong Yang[†], Wei Liu, Leihong Wu, Li Xing, Yi Wang, Xiaohui Fan* and Yiyu Cheng*

Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

*Corresponding author: Tel: +86 571 88208596; Fax: +86 571 88208426; Email: fanxh@zju.edu.cn

*Correspondence may also be addressed to Yiyu Cheng. Tel: +86 571 88208596; Fax: +86 571 88208426; Email: chengyy@zju.edu.cn

[†]These authors contributed equally to this work.

Submitted 2 January 2013; Revised 30 May 2013; Accepted 10 June 2013

Citation details: Yang,Z., Yang,J., Liu,W. *et al.* T2D@ZJU: a knowledgebase integrating heterogeneous connections associated with type 2 diabetes mellitus. *Database* (2013) Vol. 2013: article ID bat052; doi:10.1093/database/bat052.

Type 2 diabetes mellitus (T2D), affecting >90% of the diabetic patients, is one of the major threats to human health. A comprehensive understanding of the mechanisms of T2D at molecular level is essential to facilitate the related translational research. Here, we introduce a comprehensive and up-to-date knowledgebase for T2D, i.e. T2D@ZJU. T2D@ZJU contains three levels of heterogeneous connections associated with T2D, which is retrieved from pathway databases, protein–protein interaction databases and literature, respectively. In current release, T2D@ZJU contains 1078 T2D related entities such as proteins, protein complexes, drugs and others together with their corresponding relationships, which include 3069 manually curated connections, 14 893 protein–protein interactions and 26 716 relationships identified by text-mining technology. Moreover, T2D@ZJU provides a user-friendly web interface for users to browse and search data. A Cytoscape Web-based interactive network browser is available to visualize the corresponding network relationships between T2D-related entities. The functionality of T2D@ZJU is shown by means of several case studies.

Database URL: <http://tcm.zju.edu.cn/t2d>

Introduction

Diabetes mellitus is one of the major threats to human health, which is expected to affect 552 million people by 2030 (1). More than 90% of the diabetic patients are affected with type 2 diabetes mellitus (T2D) (2). Although several mechanisms of T2D have been proposed, including metabolic overload, mitochondrial dysfunction, inflammatory mediators, deposition of toxic amyloid fibrils and etc, the pathogenesis of T2D is still under investigation (3, 4). Hence, a comprehensive understanding of the mechanisms of T2D at molecular level is urgently needed.

Nowadays, the existing knowledge of T2D is located in different formats at scattered places (5, 6). The published

literature, as well as the protein–protein interaction (PPI) databases, provides a huge source of knowledge about T2D, whereas several pathway databases offer highly refined information of the molecule regulation processes. To integrate the existing knowledge, there are only a few attempts, such as T2D-Db and T2DGADB, to develop T2D-related databases. T2D-Db manually curated 330 candidate genes from the Pubmed literature and provided their corresponding information (6). T2DGADB collected 701 publications in T2D genetic association studies (7). Therefore, the development of a comprehensive database able to cover current understandings of T2D is in high demand.

Here, we introduce a comprehensive and up-to-date knowledgebase for T2D, i.e. T2D@ZJU. T2D@ZJU contains three levels of data associated with T2D, which is retrieved from pathway databases, PPI databases and literature, respectively. In comparison with the previous T2D-related databases, which focus on the information of the T2D-related genes, T2D@ZJU organizes the integrated information in a manner of network, which could be further expanded as desired with data of three distinguishable levels, in terms of data reliability and information coverage, to meet the needs of network pharmacology studies. To enhance the readability of these data, T2D@ZJU provides an interactive web tool for users to visualize the network relationships between T2D-related entities, which will offer a better understanding of T2D for basic and clinical researchers and hereby facilitate the T2D-related studies of network pharmacology and multi-target therapeutics.

Contents and Constructions

As illustrated in Figure 1, in current release, T2D@ZJU contains three levels of data, namely, curated directed connections (CDC) data set, PPI data set and text-mining-based relationships (TMR) data set, which are complementary owing to the different levels of data reliability and information coverage. The web interface of T2D@ZJU has been developed using HTML/CSS, Javascript, Ruby on Rails and Cytoscape. All data in the database have been stored in SQLite tables.

CDC data set

CDC data set was manually integrated from four well-established pathway databases including KEGG (8), Reactome (9), BioCarta (<http://www.biocarta.com>) and PANTHER (10) as well as DrugBank (11). All the human T2D-related pathways in these databases were collected. Individual interactions are the compositions of the biological pathways. The concept and the boundary of pathway are always not

coincident between different pathway databases. The aim of establishing CDC data set was to construct a network derived from the pathway databases. Therefore, individual interactions from different databases were compared with each other and adopted to construct CDC data set. Moreover, novel loci associated with T2D newly identified by genome-wide association studies and other association studies were also curated and integrated to CDC data set. In total, 1078 T2D-related entities (including 885 proteins, 14 protein complexes, 36 drugs and 143 small molecules) and 3069 directed connections are collected and imported to CDC data set.

PPI data set

iRefIndex 9.0 (12), an integrated PPI source containing 14 PPI databases, served as the source of the PPI data set. After downloading iRefIndex 9.0, the human interactions associated with the T2D-related proteins in CDC data set were extracted from it. The classes of the PPI (e.g. experimentally validated, predicted, etc.) along with the specific detection methods, interaction types and PubMed Identifiers (PMIDs) of the origin literature were also labeled. The PPI data set contains 14893 PPIs in current release.

TMR data set

The relationships among the T2D-related proteins were further investigated by a text-mining tool, i.e. SciMiner (13). SciMiner was run on a query of 'type 2 diabetes' and found 53 960 documents as of 31 December 2012. SciMiner uses a dictionary and rule-based approach to recognize biomedical objects (13). The documents with less than two genes or proteins found were removed. The remained documents were used to acquire the relationships. It is considered as a TMR if two proteins were co-occurred in at least one article. The TMR data set so far contains 26 716 relationships associated with the T2D-related proteins in CDC data set. In this work, frequency of the proteins co-occurrence was counted to index the weight of the relationship. The PMIDs of the origin literature were also labeled.

T2D network

T2D@ZJU provides a Cytoscape Web (14) based interactive network browser for users to visualize the network relationships between T2D-related entities. The T2D network constructed on the basis of CDC data set is shown in Figure 2. To improve the readability of the network, entities that play same mechanistic role within the network were aggregated into an entity class, e.g. the entity class 'AKT' includes AKT1, AKT2 and AKT3.

Utilities

T2D@ZJU is freely available to academic users, which provides a user-friendly web interface for users to browse and

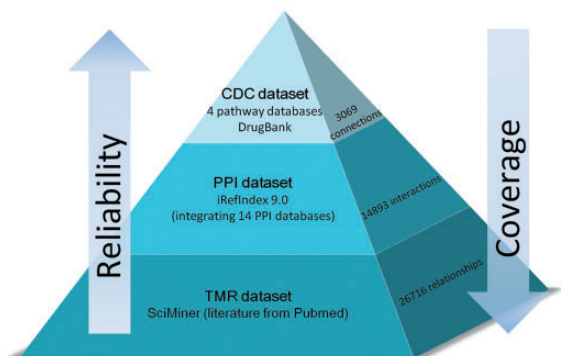


Figure 1. Overview of the T2D@ZJU, which is composed of three data sets, i.e. CDC data set, PPI data set and TMR data set.

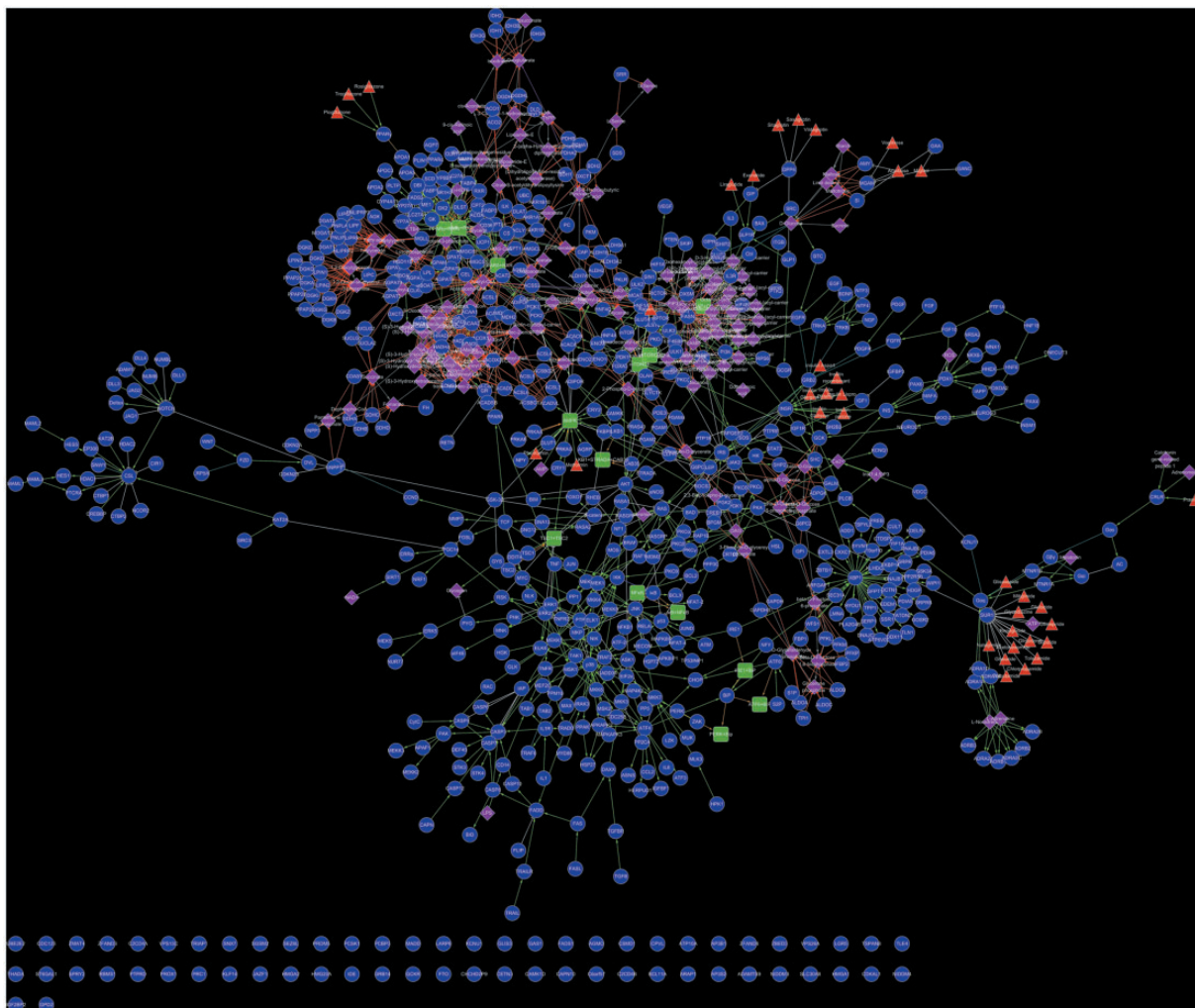


Figure 2. The T2D network constructed on the basis of CDC data set. In the network, blue circle, round rectangle, purple diamond and red triangle represented protein, protein complex, small molecule and drug, respectively.

search data. The main functions of the web interface are illustrated in [Figure 3](#).

Search

The query interface of T2D@ZJU allows users to perform two types of searches. Users can query using entity/entity class names or external identifiers (e.g. gene symbol, Entrez Gene ID, UniProt ID, etc.) in search box. Once given a query, the background program can find all the directly related entities/entity classes and draw the corresponding networks. Information of all the connections in the network is listed below the network, and the PMIDs of the references are linked to the respective literature sources. Clicking on any node of a given network, the detailed information would be listed with external links. Users also can search the pathway associated with T2D by selecting a pathway name in the dropdown box, and all the

connections related to the corresponding pathway will be returned.

To demonstrate the utility of the search function and the associated knowledge discovery, a use case was applied. A search with 'SOCS3' as the query keyword resulted in five entities (INSR, IRS, LEPR, JAK2 and NF κ B, respectively) related to SOCS3 directly. As displayed in [Figure 4](#), these six entities formed a network through seven interactions. In obesity, SOCS3 is upregulated in the skeletal muscle, liver, adipose tissue and hypothalamus (15–18), which is associated with increases in inflammation. SOCS3 is a major negative regulator of insulin and leptin signaling (15), which is thought to contribute to the pathogenesis of insulin resistance. In adipocytes, SOCS3 deficiency increases insulin-stimulated IRS1 and IRS2 phosphorylation, whereas overexpression of SOCS3 reduces both IRS1 protein levels and the phosphorylation of IRS1 and IRS2 (19).

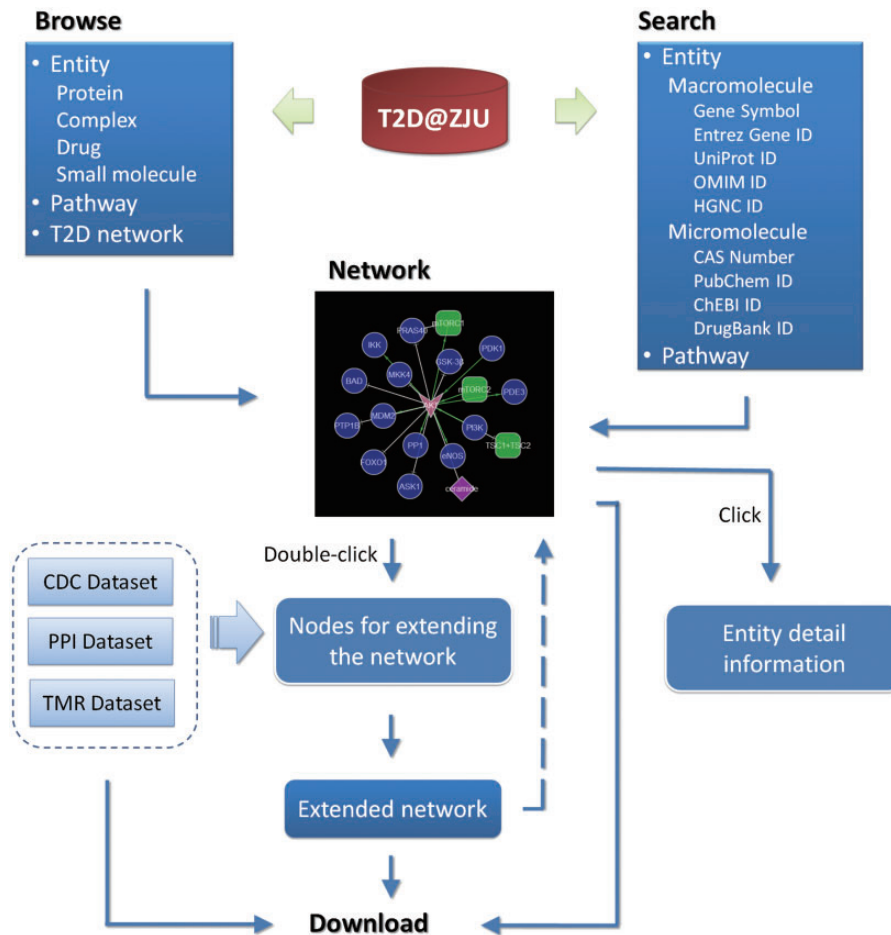


Figure 3. Graphical summary of the main functions of the T2D@ZJU web interface.

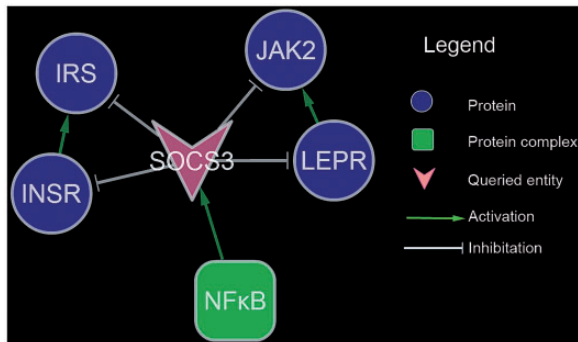


Figure 4. The outcome of the search with 'SOCS3' as the query keyword.

Genetic deletion of SOCS3 in mouse liver results in increased insulin sensitivity owing to increased IRS1 phosphorylation (20). SOCS3 has also been reported to co-immunoprecipitate with both INSR and IRS1 in skeletal muscle (21). SOCS3 inhibits leptin signaling by binding

to phosphorylated tyrosine residue of the leptin receptor (22). SOCS3 is a major regulator of JAK signaling, which binds and inhibits the catalytic domains of JAK2 via an evolutionarily conserved motif unique to JAKs (23).

Therapies aimed at inhibiting SOCS3 in skeletal muscle may be a favorable strategy to reverse glucose intolerance and insulin resistance (24). NFκB plays an important role in regulating PIM2 induced SOCS3 expression (25). Therefore, inhibiting NFκB might be an alternative way to reduce SOCS3 expression. Moreover, Salicylates have been reported to inhibiting NFκB by inhibiting IκB kinase-β (26–28). Reduce of SOCS3 expression might be involved in the mechanism of the anti-inflammatory drug through which it works.

Browse and network extension

Users are allowed to browse the T2D networks. Moving the mouse over any node of the network, all the directly related interactions will be highlighted, whereas clicking on a node, details about the node will be listed below.

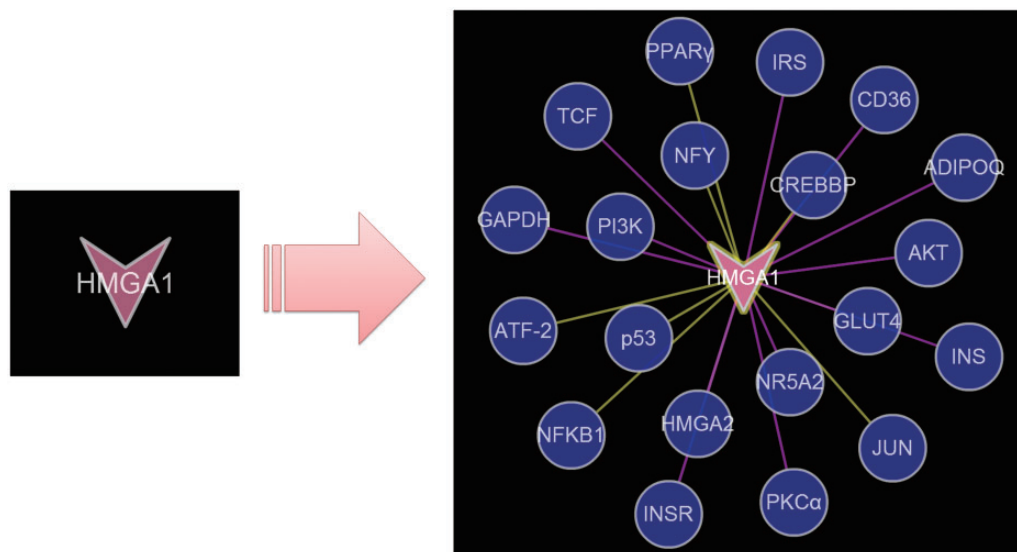


Figure 5. The extending of HMGA1 with PPI and TMR data sets. The yellow edges represented relationships derived from PPI data set, whereas the pink represented relationships from TMR data set.

Double-clicking on any node of a given network, all the directly related entities from different data sets (CDC, PPI or TMR data set) will be listed, and users are able to select any of listed entities to extend the network.

Nodes for which knowledge is still limited have fewer connections with other entities in the T2D network. However, some of them are proposed to be associated with enhanced risk of T2D, such as HMGA1 (29,30). Connections from PPI and TMR data sets provide valuable clues, as shown in Figure 5, allowing researchers to propose hypotheses that can be tested by further experiments.

Download

All three data sets are freely available at the download page. We also provide an illustration and examples to guide the users through the different functions of the web interface at the help page.

Other application

T2D@ZJU could also offer a view of T2D for visualization of 'Omics' data, such as transcriptomics data and proteomics data. For instance, users could download the T2D network and map the 'Omics' data into it. To represent this functionality, we used transcriptomics data from van Tienen *et al.* (31) comparing skeletal muscle from longstanding T2D patients against that from controls matched by age and body mass index (Gene Expression Omnibus accession number GSE19420). The fold changes in the gene expression of the T2D group relative to the controls were projected onto the T2D network. The color of the node

coded the difference of the gene expression. In detail, red and green represented up- and downregulation, respectively, whereas gray meant the data were lacking. The status of the gene expression was intuitively shown in Figure 6A. In general, fold changes were small. The expression of genes involved in citrate cycle was displayed in detail in Figure 6B. In the T2D group, distinctly, almost all the expression of genes involved in citrate cycle was decreased compared with controls. It demonstrated that the ability of using glucose as energy source was lower in T2D patients, which was consistent with the previous understanding (32).

Conclusions

In summary, T2D@ZJU provides a comprehensive knowledgebase about T2D, including a three levels layout on the associated connections. It builds a basis for the further study of T2D in system biology, which helps to better understand of the complex disease, as well as give clues facilitating researchers to propose testable hypotheses. Furthermore, the T2D-related studies of network pharmacology and multi-target therapeutics would also be accelerated.

The T2D associated information is expected to grow with the continuing effort of the researchers. Therefore, T2D@ZJU will be updated periodically. Specific network analysis tools would be developed for the further research. Semantic data mining is intended to be used to extract the Pubmed literature related to T2D to improve the confidence levels of the TMR data set.

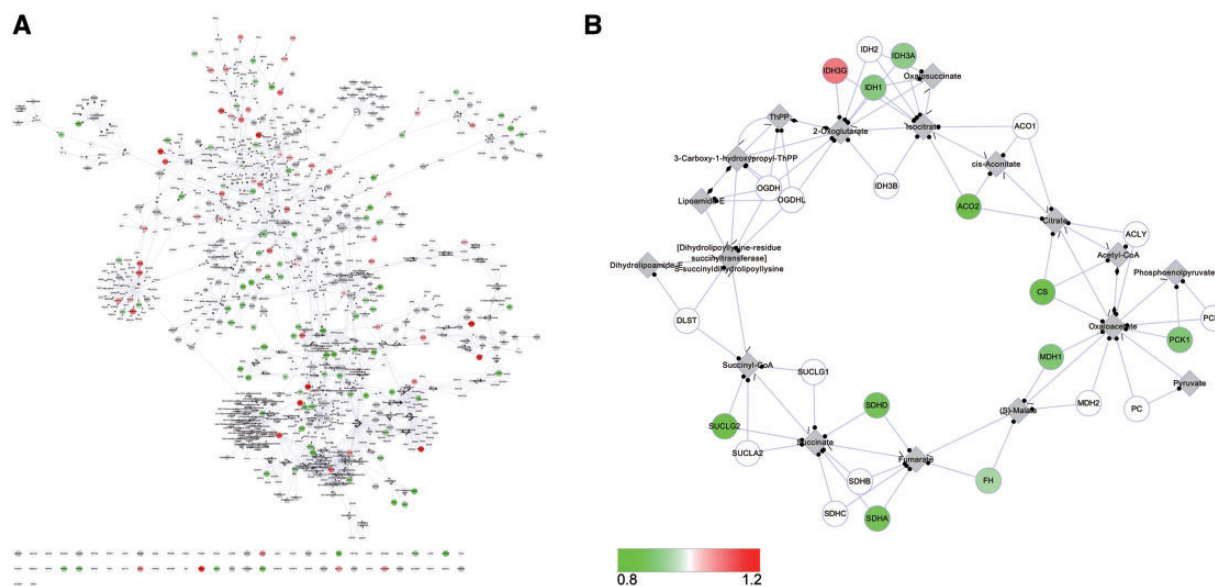


Figure 6. Visualization of transcriptomics data on the T2D network. (A) The gene expression status of the T2D network. (B) The expression of genes involved in citrate cycle. The fold changes of the gene expression were depicted by colors, as indicated by the bar. The networks were constructed in Cytoscape.

Funding

National S&T Major Project (No. 2012ZX09503001), Program for New Century Excellent Talents in University (NCET-12-0488) and Fundamental Research Funds for the Central Universities. Funding for open access charge: National S&T Major Project (No. 2012ZX09503001).

Conflict of interest. None declared.

References

- Whiting,D.R., Guariguata,L., Weil,C. *et al.* (2011) IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res. Clin. Pract.*, **94**, 311–321.
- Chang,K.-H., Chuang,T.-J., Lyu,R.-K. *et al.* (2012) Identification of gene networks and pathways associated with Guillain-Barre syndrome. *PLoS One*, **7**, 80–89.
- Muoio,D.M. and Newgard,C.B. (2008) Molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes. *Nat. Rev. Mol. Cell Biol.*, **9**, 193–205.
- Ashcroft,F.M. and Rorsman,P. (2012) Diabetes mellitus and the beta cell: the Last ten years. *Cell*, **148**, 1160–1171.
- Liechti,R., Csardi,G., Bergmann,S. *et al.* (2010) EuroDia: a beta-cell gene expression resource. *Database*, 2010, baq024.
- Agrawal,S., Dimitrova,N., Nathan,P. *et al.* (2008) T2D-Db: an integrated platform to study the molecular basis of Type 2 diabetes. *BMC Genomics*, **9**, 320.
- Lim,J.E., Hong,K.-W., Jin,H.-S. *et al.* (2010) Type 2 diabetes genetic association database manually curated for the study design and odds ratio. *BMC Med. Inform. Decis. Mak.*, **10**, 76.
- Kanehisa,M., Goto,S., Furumichi,M. *et al.* (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
- Matthews,L., Gopinath,G., Gillespie,M. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.
- Mi,H., Dong,Q., Muruganujan,A. *et al.* (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.*, **38**, D204–D210.
- Knox,C., Law,V., Jewison,T. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
- Razick,S., Magklaras,G. and Donaldson,I. (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics*, **9**, 405.
- Hur,J., Schuyler,A.D., States,D.J. *et al.* (2009) SciMiner: web-based literature mining tool for target identification and functional enrichment analysis. *Bioinformatics*, **25**, 838–840.
- Lopes,C.T., Franz,M., Kazi,F. *et al.* (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
- Yang,Z.G., Hulver,M., McMillan,R.P. *et al.* (2012) Regulation of insulin and leptin signaling by muscle suppressor of cytokine signaling 3 (SOCS3). *PLoS One*, **7**, e47493.
- Gu,H.L., Liu,L., Ma,S. *et al.* (2009) Inhibition of SOCS-3 in adipocytes of rats with diet-induced obesity increases leptin-mediated fatty acid oxidation. *Endocrine*, **36**, 546–554.
- Reed,A.S., Unger,E.K., Olofsson,L.E. *et al.* (2010) Functional role of suppressor of cytokine signaling 3 upregulation in hypothalamic leptin resistance and long-term energy homeostasis. *Diabetes*, **59**, 894–906.
- Emanuelli,B., Peraldi,P., Filloux,C. *et al.* (2001) SOCS-3 inhibits insulin signaling and is up-regulated in response to tumor necrosis factor-alpha in the adipose tissue of obese mice. *J. Biol. Chem.*, **276**, 47944–47949.

19. Shi,H., Cave,B., Inouye,K. et al. (2006) Overexpression of suppressor of cytokine signaling 3 in adipose tissue causes local but not systemic insulin resistance. *Diabetes*, **55**, 699–707.
20. Sachithanandan,N., Fam,B.C., Fynch,S. et al. (2010) Liver-specific suppressor of cytokine signaling-3 deletion in mice enhances hepatic insulin sensitivity and lipogenesis resulting in fatty liver and obesity. *Hepatology*, **52**, 1632–1642.
21. Yaspelkis,B.B., Kvasha,I.A. and Figueroa,T.Y. (2009) High-fat feeding increases insulin receptor and IRS-1 coimmunoprecipitation with SOCS-3, IKK alpha/beta phosphorylation and decreases PI-3 kinase activity in muscle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, **296**, R1709–R1715.
22. Bjorbaek,C., Lavery,H.J., Bates,S.H. et al. (2000) SOCS3 mediates feedback inhibition of the leptin receptor via Tyr(985). *J. Biol. Chem.*, **275**, 40649–40657.
23. Babon,J.J., Kershaw,N.J., Murphy,J.M. et al. (2012) Suppression of cytokine signaling by SOCS3: characterization of the mode of inhibition and the basis of its specificity. *Immunity*, **36**, 239–250.
24. Jorgensen,S.B., O'Neill,H.M., Sylow,L. et al. (2013) Deletion of skeletal muscle SOCS3 prevents insulin resistance in obesity. *Diabetes*, **62**, 56–64.
25. Narayana,Y., Bansal,K., Sinha,A.Y. et al. (2009) SOCS3 expression induced by PIM2 requires PKC and PI3K signaling. *Mol. Immunol.*, **46**, 2947–2954.
26. Fleischman,A., Shoelson,S.E., Bernier,R. et al. (2008) Salsalate improves glycemia and inflammatory parameters in obese young adults. *Diabetes Care*, **31**, 289–294.
27. Hundal,R.S., Petersen,K.F., Mayerson,A.B. et al. (2002) Mechanism by which high-dose aspirin improves glucose metabolism in type 2 diabetes. *J. Clin. Invest.*, **109**, 1321–1326.
28. Yuan,M.S., Konstantopoulos,N., Lee,J.S. et al. (2001) Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of IKK beta. *Science*, **293**, 1673–1677.
29. Garg,A. (2011) HMG1A1, a novel locus for type 2 diabetes mellitus. *JAMA*, **305**, 938–939.
30. Liu,L., Ding,H., Wang,H.R. et al. (2012) Polymorphism of HMG1A1 is associated with increased risk of type 2 diabetes among Chinese individuals. *Diabetologia*, **55**, 1685–1688.
31. van Tienen,F., Praet,S., de Feyter,H. et al. (2012) Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes. *J. Clin. Endocrinol. Metab.*, **97**, 3261–3269.
32. Schrauwen,P. and Hesselink,M.K.C. (2008) Reduced tricarboxylic acid cycle flux in type 2 diabetes mellitus? *Diabetologia*, **51**, 1694–1697.