Published online 11 November 2014 *Nucleic Acids Research, 2015, Vol. 43, Database issue* ***D963–D967***
doi: 10.1093/nar/gku1139

PubAngioGen: a database and knowledge for angiogenesis and related diseases

Peng Li[1,†], Yongrui Liu[1,†], Huan Wang[1,†], Yuan He[1,†], Xue Wang[1], Yundong He[1], Fang Lv[1], Huaqing Chen[1], Xiufeng Pang[1], Mingyao Liu[1,2], Tieliu Shi[1,*] and Zhengfang Yi[1,*]

[1]The center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, 500 Dongchuan Road, Shanghai 200241, China and [2]Center for Cancer and Stem Cell Biology, Institute of Biosciences and Technology, Texas A&M University Health Science Center, Houston, TX 77030, USA

Received August 15, 2014; Revised October 18, 2014; Accepted October 27, 2014

ABSTRACT

Angiogenesis is the process of generating new blood vessels based on existing ones, which is involved in many diseases including cancers, cardiovascular diseases and diabetes mellitus. Recently, great efforts have been made to explore the mechanisms of angiogenesis in various diseases and many angiogenic factors have been discovered as therapeutic targets in anti- or pro-angiogenic drug development. However, the resulted information is sparsely distributed and no systematical summarization has been made. In order to integrate these related results and facilitate the researches for the community, we conducted manual text-mining from published literature and built a database named as PubAngioGen (http://www.megabionet.org/aspd/). Our online application displays a comprehensive network for exploring the connection between angiogenesis and diseases at multilevels including protein–protein interaction, drug-target, disease-gene and signaling pathways among various cells and animal models recorded through text-mining. To enlarge the scope of the PubAngioGen application, our database also links to other common resources including STRING, DrugBank and OMIM databases, which will facilitate understanding the underlying molecular mechanisms of angiogenesis and drug development in clinical therapy.

INTRODUCTION

Angiogenesis, the process of generating new blood vessels based on pre-existing ones (1), is a complex multistep dynamic process that comprises matured vessel changes, pericyte detachment, extracellular matrix (ECM) degradation and remodeling, proliferation, migration and assembly of endothelial cells (ECs) into tubule structures (2). The recruitment and differentiation of circulating endothelial progenitor cells (EPCs) also contributes to the vascularization process (3,4).

Angiogenesis itself is controlled by a fine balance between pro- and anti-angiogenic factors (5) involved in many signals such as vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), platelet-derived growth factor (PDGF) and endothelial nitric oxide synthase (eNOS) pathway (6). If the balance is tipped, it leads to some pathological situations, either in deficit conditions (e.g. cardiovascular diseases, diabetic tissue ischemia) or in excess ones (e.g. cancer growth and metastases, atherosclerotic plaque development and diabetic retinopathy). Therefore, angiogenesis plays a key role in main diseases, including cancers (7,8), cardiovascular diseases (9) and complications of diabetes mellitus (10).

The angiogenic switch is a critical progression point in a range of etiology and pathogenesis; therefore, it was assumed that angiogenesis inhibition or promotion of regulating pathway and special target might play a fundamental role in related diseases therapy. Since the first work of Folkman that highlighted the role of angiogenesis as a crucial determinant for tumor development and progression (11), the idea of anti-angiogenesis as a therapeutic strategy has been accounted for several decades. In fact, several anti-angiogenic therapies have recently been approved by Food and Drug Administration (FDA) for cancers, including the humanized antibody bevacizumab (Avastin), which targets VEGF-A, the tyrosine kinase inhibitor sorafenib (Nexavar), which targets Raf and VEGF and PDGF receptors, and the tyrosine kinase inhibitor sunitinib (Sutent), which targets VEGF and PDGF receptors (12,13). In addition, promoting blood vessel growth has been recognized

*To whom correspondence should be addressed. Tel: +86 21 5434 5016; Fax: +86 21 5434 4922; Email: zfyi@bio.ecnu.edu.cn
Correspondence may also be addressed to Tieliu Shi. Tel: +86 21 5434 5020; Fax: +86 21 5434 4922; Email: tieliushi01@gmail.com
†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

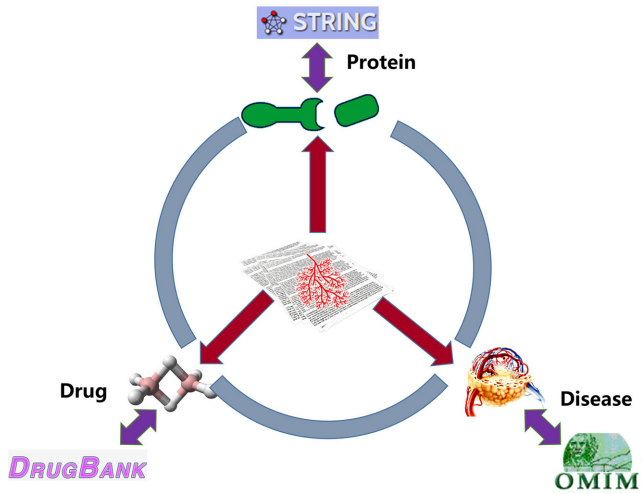


Figure 1. The data sources and relation models.

as a potential therapeutic approach for the treatment of ischemic diseases. To date, therapeutic benefits have also been achieved with anti-angiogenic therapy in the treatment of life-threatening tumors.

Besides drugs such as small molecule inhibitors, researchers have also focused on gene therapy using pro-angiogenic factors and/or cell-based therapy using several types of cells, including bone marrow cells and EPCs, to achieve therapeutic angiogenesis. Therefore, the future of research and drug development involved in anti- or pro- angiogenesis depends on the comprehensive understanding of the mechanisms and relationships between objectives (e.g. compounds or proteins), cells, diseases as well as signaling pathways and targets. Fully integration of those published results will surely facilitate the researches of angiogenesis and diseases.

To this end, for the first time, we developed PubAngioGen, a knowledge-based database, to record those results related to angiogenesis. During data collection, we manually extracted those published relevant results at multilevels including molecular, cellular, animal model and disease. We then comprehensively summarized the relationships between angiogenesis and major diseases, which help users to obtain the latest information of angiogenesis involved in main diseases and signaling pathways. In addition, we also collected those potential drug targets and related compounds, which have been experimentally proven to be effective on angiogenesis. Moreover, those candidates of angiogenic factors that are under clinical trials and the therapeutic targets for anti-angiogenic or pro-angiogenic drug development also have been collected (Table 1).

MATERIALS AND METHODS

PubAngioGen provides information on three different areas related to angiogenesis, proteins, drugs, diseases and the links between them. The data integrated from related resources and manual text-mining from published articles, as well as the relation models were depicted in Figure 1.

The core components were collected mainly through text-mining from published articles and protein–protein interaction was extended based on STRING database (14). Information for drugs and diseases was extended based on DrugBank (15,16) and Online Mendelian Inheritance in Man (OMIM) (17,18), respectively.

The main goal of our system is to build the connections between angiogenesis and diseases through disease, proteins as well as related drugs. Users can query our system easily according to their interests to obtain the information about angiogenesis, diseases, targets, drugs and the mechanisms of drug interactions.

First of all, through manual text-mining, we collected results published in the past decades related to angiogenesis at multilevels including molecular, cellular, animal model and disease. We extracted the information of signaling pathways involved in angiogenesis and summarized the relationships between angiogenesis and major diseases. The information on drugs that FDA approved to treat angiogenesis-related diseases or candidates that are in clinical trials was also extracted. All the information recorded was curated by the experts from National Consortium of GPCR of China and could be traced back to the original publications. Then, we used the well-defined data sources to extend the interaction between proteins or proteins and compounds.

DATABASE ACCESS AND NETWORK DISPLAY

Database query

Users can identify proteins, drugs or diseases when they would like to know their roles in angiogenesis at PubAngioGen through the quick search form (Figure 2A) or the query page (Figure 2B). PubAngioGen allows several types of query keywords consistent with other databases, including UniProtKB/Swiss-Prot ID (19), Entrez Gene name (20,21), Refseq provisional ID (NCBI) (22,23), symbols, disease name or drug name.

Result page and network display

To virtually display the connection between proteins, drugs or diseases and angiogenesis, we developed network-display tools, which provide more detailed network information. Our network display includes protein–protein, compound–protein interaction and signaling pathways involved in angiogenesis, which help users to check the potential combinational effect. All returned pages inform the user of related protein annotations by texts, graphs and tables (Figure 3). For example, if the user is interested in the epidermal growth factor receptor (EGFR) protein, the 'Related Information For EGFR' shows a summary network of the compounds, peptides, proteins, complexes and diseases nodes with different colors that connected with EGFR (Figure 3A) and users can select a node to see its directed partners (Figure 3B). The first table (Figure 3C) lists the basic information of EGFR and the second table (Figure 3D) presents the detailed related information of EGFR such as interaction type, reaction type and so on. The underlined red parts are hyperlinks that include literature references from PubMed and experimental detection methods extracted from text-mining. Moreover, to explore the distinctive mechanisms of each angiogenesis-related gene, the gene expression levels in different statuses were recorded in our databases based

Table 1. Data statistics for PubAngioGen

Data field	Data source	Amount of data
Compounds/drugs	Text-mining, DrugBank	6693
Gene/protein targets	Text-mining	963
Signal pathways	Text-mining, CST pathways	179
Diseases	Text-mining, OMIM	2364

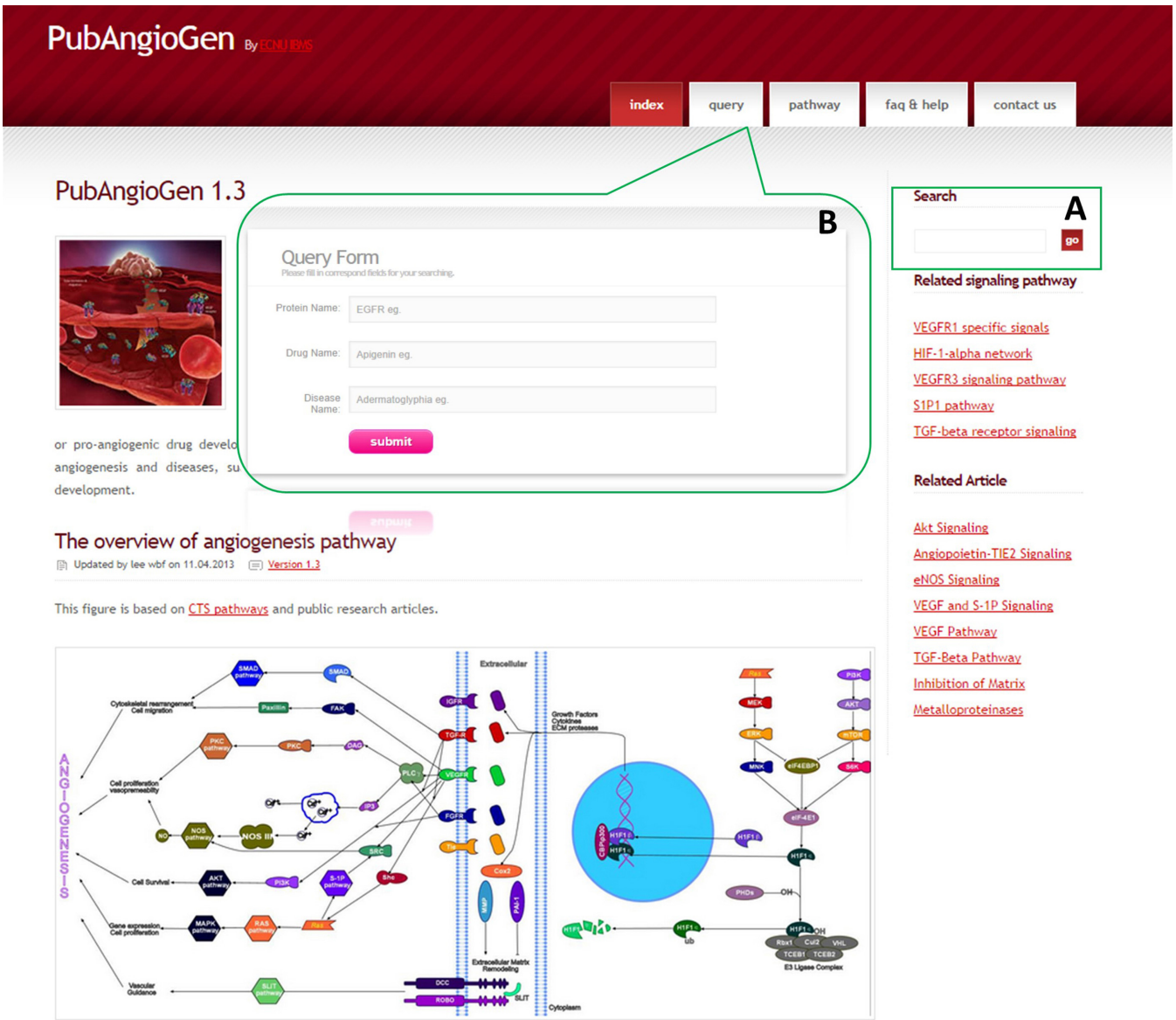
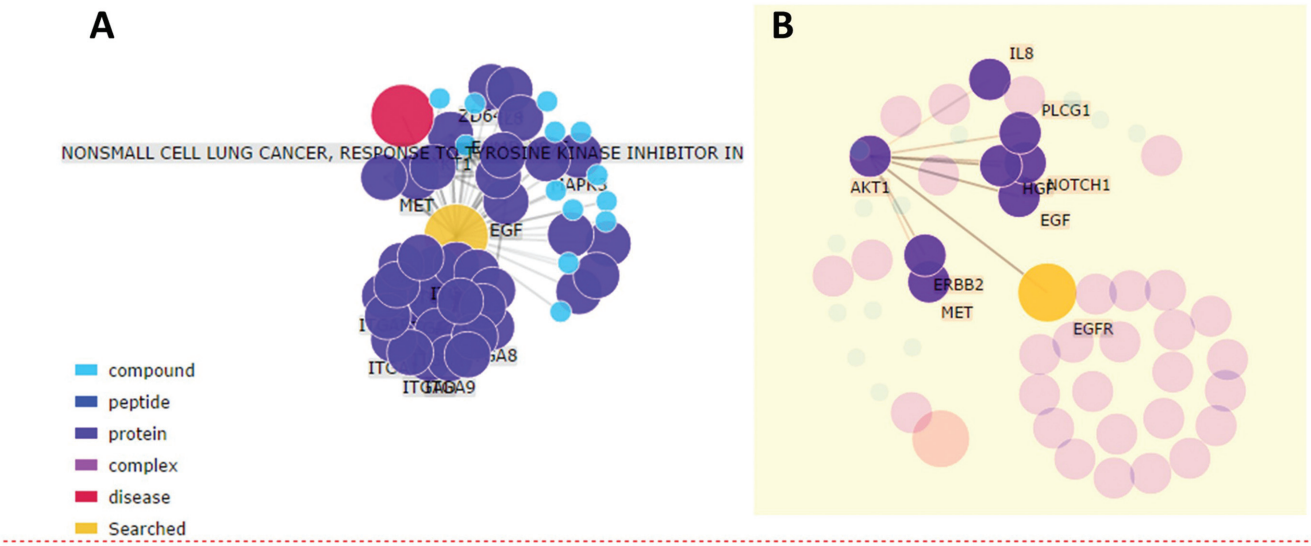


Figure 2. PubAngioGen homepage and query interface. (**A**) Quick search form on homepage. (**B**) Detail search form on query page.

on the experiments from Gene Expression Omnibus (GEO) (24); more experiments are under evaluation and the results will be processed and stored in PubAngioGen soon.

Signaling pathway

During the process of angiogenesis-related diseases, the signaling pathway stimulation or repression acts in the different ways. To incorporate those diversities of signaling pathways, our database displays the latest singling pathways involved in angiogenesis from our collected data and other relevant databases (e.g. Cell Signaling Technology Pathway Database, http://www.cellsignal.com/). Users can submit those interested proteins, diseases or other components to view their roles in the process of diseases of angiogenesis. For example, when users enter the 'pathway' in home page of our website and click 'EGFR-dependent Endothelin signaling events', the returned page displays the network and components involved in this pathway.



Unit Name	Unit Type	Chromosome	Full Name C
EGFR	protein	7p12	epidermal growth factor receptor
Experission: [FPKM_SRP008496] NON tumor 5.15, Tumor Esophageal Squamous Cell Carcinoma 63.1 [GSE4290]:NON tumor 7.98, Tumor: astrocytomas 10.53, glioblastomas 11.09, oligodendrogliomas 10.98			

Interaction type	Reaction type	Related unit	Unit type	Related DataSource D
protein-protein	downstream component Protein interaction	ITGAD	protein	TEXT MINING:21765015
protein-protein	Same Level Collaboration Activation	AKT1	protein	TEXT MINING:19061355
protein-protein	downstream component Protein interaction	ITGB6	protein	TEXT MINING:21765015
protein-protein	downstream component Protein interaction	ITGA3	protein	TEXT MINING:21765015
compound-protein	downstream component Inhibition	ZD6474	compound	TEXT MINING:17889445

Figure 3. Query result for EGFR. (**A**) Network viewer to show the proteins, drugs and diseases related to EGFR and the interactions among all of them. (**B**) When the node in the network is selected, all the directed partners are highlighted to facilitate users to investigate the function of specific node; double click will return to the original network. (**C**) The annotation table for queried entry. (**D**) The interaction table for all related information.

DISCUSSION

In order to better elucidate the pathological angiogenesis in these diseases, such as cancers, cardiovascular diseases and diabetes mellitus, we extracted abundant angiogenesis-related results published in past decades and comprehensively summarized their relationships at multilevels including molecular, cellular, animal model and disease in our database. On one hand, we provided the latest information of angiogenesis involved in main diseases and signaling pathways like HIF-1 (symbol: SETD2) network, VEGFR1 (symbol: FLT1), VEGFR2 (symbol: KDR), VEGFR3 (symbol: FLT4) specific signals, transforming growth factor beta (TGFβ) (symbol: TGFB1) receptor signaling pathway and S1P1 (symbol: S1PR1) pathway (25–27). On the other hand, we provided the first integrated resources for exploring the potential drug targets, clinical compounds that are effective on angiogenesis as well as related research progresses for drug development. Therefore, PubAngioGen database would be a promising angiogenic online searcher to some extent.

Meanwhile, the PubAngioGen is supported by National Consortium of GPCR of China; many experts in the field are involved in curating angiogenesis-related articles. While the amount of data entries is increasing daily, many exploring experiments are performed in our collaborators' labs including high-throughput experiments like RNA-Seq (28) and LC MS/MS (29) to validate or extend our findings in PubAngioGen. Along with the accumulation of data, in the future, we will further extend and redefine angiogenesis-

related pathways to help other researchers to better understand the mechanism of it.

In conclusion, PubAngioGen database systemically proposes the relationships between angiogenesis and diseases. Our data retrieve and display system makes researchers access the latest information efficiently. All of these will promote PubAngioGen to be referred by modern medicine researchers for novel discovery in clinical medicine and research work in the near future.

ACKNOWLEDGEMENTS

We are also grateful for the permission from 'Cell Signaling Technology' for the related pathway information and the help from MegaSci Inc. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

FUNDING

Funding for open access charge: Major State Basic Research Development Program of China [2015CB910400]; National Natural Science Foundation of China [81272463, 81472788, 81330049, 31401133]; Science and Technology Commission of Shanghai Municipality [14YF1404400].
Conflict of interest statement. None declared.

REFERENCES

1. Song,Y., Dai,F., Zhai,D., Dong,Y., Zhang,J., Lu,B., Luo,J., Liu,M. and Yi,Z. (2012) Usnic acid inhibits breast tumor angiogenesis and growth by suppressing VEGFR2-mediated AKT and ERK1/2 signaling pathways. *Angiogenesis*, **15**, 421–432.
2. Rundhaug,J.E. (2005) Matrix metalloproteinases and angiogenesis. *J. Cell. Mol. Med.*, **9**, 267–285.
3. Fam,N.P., Verma,S., Kutryk,M. and Stewart,D.J. (2003) Clinician guide to angiogenesis. *Circulation*, **108**, 2613–2618.
4. Hristov,M., Erl,W. and Weber,P.C. (2003) Endothelial progenitor cells: mobilization, differentiation, and homing. *Arterioscler. Thromb. Vasc. Biol.*, **23**, 1185–1189.
5. Anand,S. (2013) A brief primer on microRNAs and their roles in angiogenesis. *Vasc. Cell*, **5**, 2.
6. Dong,Y., Lu,B., Zhang,X., Zhang,J., Lai,L., Li,D., Wu,Y., Song,Y., Luo,J., Pang,X. *et al.* (2010) Cucurbitacin E, a tetracyclic triterpenes compound from Chinese medicine, inhibits tumor angiogenesis through VEGFR2-mediated Jak2-STAT3 signaling pathway. *Carcinogenesis*, **31**, 2097–2104.
7. Zhang,X., Song,Y., Wu,Y., Dong,Y., Lai,L., Zhang,J., Lu,B., Dai,F., He,L., Liu,M. *et al.* (2011) Indirubin inhibits tumor growth by antitumor angiogenesis via blocking VEGFR2-mediated JAK/STAT3 signaling in endothelial cell. *Int. J. Cancer*, **129**, 2502–2511.
8. Li,J., Xu,Y., Jiao,H., Wang,W., Mei,Z. and Chen,G. (2014) Sumoylation of hypoxia inducible factor-1alpha and its significance in cancer. *Sci. China Life Sci.*, **57**, 657–664.
9. Deveza,L., Choi,J. and Yang,F. (2012) Therapeutic angiogenesis for treating cardiovascular diseases. *Theranostics*, **2**, 801–814.
10. Tahergorabi,Z. and Khazaei,M. (2012) Imbalance of angiogenesis in diabetic complications: the mechanisms. *Int. J. Prev. Med.*, **3**, 827–838.
11. Folkman,J. (1971) Tumor angiogenesis: therapeutic implications. *N Engl. J. Med.*, **285**, 1182–1186.
12. Carmeliet,P. and Jain,R.K. (2011) Molecular mechanisms and clinical applications of angiogenesis. *Nature*, **473**, 298–307.
13. Jain,R.K. (2013) Normalizing tumor microenvironment to treat cancer: bench to bedside to biomarkers. *J. Clin. Oncol.*, **31**, 2205–2218.
14. Franceschini,A., Szklarczyk,D., Frankild,S., Kuhn,M., Simonovic,M., Roth,A., Lin,J., Minguez,P., Bork,P., von Mering,C. *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.
15. Wishart,D.S., Knox,C., Guo,A.C., Cheng,D., Shrivastava,S., Tzur,D., Gautam,B. and Hassanali,M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
16. Law,V., Knox,C., Djoumbou,Y., Jewison,T., Guo,A.C., Liu,Y., Maciejewski,A., Arndt,D., Wilson,M., Neveu,V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.
17. Hamosh,A., Scott,A.F., Amberger,J., Bocchini,C., Valle,D. and McKusick,V.A. (2002) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **30**, 52–55.
18. Amberger,J., Bocchini,C. and Hamosh,A. (2011) A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). *Hum. Mutat.*, **32**, 564–567.
19. UniProt,C. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.
20. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
21. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.
22. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
23. Pruitt,K.D., Brown,G.R., Hiatt,S.M., Thibaud-Nissen,F., Astashyn,A., Ermolaeva,O., Farrell,C.M., Hart,J., Landrum,M.J., McGarvey,K.M. *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, D756–D763.
24. 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.
25. Veikkola,T., Karkkainen,M., Claesson-Welsh,L. and Alitalo,K. (2000) Regulation of angiogenesis via vascular endothelial growth factor receptors. *Cancer Res.*, **60**, 203–212.
26. Maceyka,M., Payne,S.G., Milstien,S. and Spiegel,S. (2002) Sphingosine kinase, sphingosine-1-phosphate, and apoptosis. *Biochim. Biophys. Acta*, **1585**, 193–201.
27. Li,J.M., Shen,X., Hu,P.P. and Wang,X.F. (1998) Transforming growth factor beta stimulates the human immunodeficiency virus 1 enhancer and requires NF-kappaB activity. *Mol. Cell. Biol.*, **18**, 110–121.
28. Chen,G. and Shi,T. (2013) Next-generation sequencing technologies for personalized medicine: promising but challenging. *Sci. China Life Sci.*, **56**, 101–103.
29. Liu,Y., Yin,H. and Chen,K. (2013) Platelet proteomics and its advanced application for research of blood stasis syndrome and activated blood circulation herbs of Chinese medicine. *Sci. China Life Sci.*, **56**, 1000–1006.