ORIGINAL ARTICLE

Eur. J. Anat. 19 (2): 131-137 (2015)

Computer tools and collaborative translational research in the life sciences: the further advance of genomics and proteomics

Carlos González[1], Pablo J. Álvarez[2,4], Fernando Rodríguez-Serrano F.[2,4], Celia Vélez[2,4], A.R. Rama[3,4], Amelia E. Aránega[3,4] and José C. Prados[2,4]

[1]Franchise at Novartis Oncology Com, [2]Dept. of Anatomy, University of Granada, Spain, [3]Dept. of Health Science, University of Jaén, Spain and [4]Institute of Biopathology and Regenerative Medicine (IBIMER), University of Granada, Spain

SUMMARY

Currently, biomedical research is mainly focused on overcoming the major challenges faced by society, including the development of new therapeutic strategies against highly prevalent diseases. Over the past 20 years, considerable advances in this field have been achieved through an interdisciplinary and collaborative approach, enhanced by the development of computer science and its applications in genomics and proteomics. This study centers on platforms for the data management of research assets with high specialization in genomics and proteomics, analyzing the role of web-based databases in the progress made in these areas and evaluating their impact on global scientific production. The web platforms analyzed have proven to be an important resource for stimulating the integration of research data through information exchange. Specialized web search sites facilitate the obtaining of data in these specific areas, creating a trend in current biomedical research. The importance of these platforms is revealed by their impact on scientific production, with some being referenced in more than 100,000 articles and patents. A wider extension of the use of these tools can be expected within the modern society of information.

Corresponding author: Pablo Alvarez. Institute of Biopathology and Regenerative Medicine (IBIMER), University of Granada, 18071 Granada, Spain. Phone: +34 608376697; Fax: +34 958 246296. E-mail: aranega@ugr.es

Key words: Databases – Genomics – Ontology – Proteomics

INTRODUCTION

Current biomedical research is largely focused on addressing the major challenges faced by society, including the development of new therapeutic strategies against highly prevalent diseases. Considerable advances in this field over the past 20 years can be attributed to an interdisciplinary and collaborative approach and to the development of information technology and specific disciplines such as genomics and proteomics, which have had a significant effect on high-impact areas in biomedicine. The abundant data derived from research in genomics and proteomics have allowed valid and useful conclusions to be drawn. However, this has only been possible by applying the computer tools developed over the past few years for the adequate management of these data (Webster et al., 2011). These developments have strengthened strategies for improving communication and access to data from research groups involved in common interdisciplinary projects, as well as for promoting diffusion of their scientific achievements (Anderson et al., 2012). The scope of research projects is also defined by their social usefulness, and the management of scientific accomplishments should also aim to achieve their recognition by industry and the business world in

Submitted: 12 November, 2014. *Accepted:* 15 December, 2014.

general.

The biomedical scientific community has been able to develop highly innovative therapies through an interdisciplinary approach, with an increasing involvement of basic and translational research in the evaluation of this knowledge (Beskow and Smolek, 2009). There is a need to support the management of this type of specialization of knowledge at these new frontiers of science (Collins, 2011), and web technologies and metadata can play a key role in organizing the knowledge generated by modern collaborative research. Enormous amounts of information are generated in the Life Sciences, including data from genotyping projects (e.g., the Human Genome Project) and on genetic variants and disease predisposition, diagnosis, and the response to treatments. The development of innovative computer tools to extract and organize the data has facilitated the management of these data, endowing metadata with useful meaning and generating new knowledge (Szalay and Gray, 2006; Antezana et al., 2009; Rebholz-Schuhmann and Nenadic, 2010). They have proven especially useful in such strategic areas as gene sequencing and expression (Baux et al., 2009).

The integration of multidisciplinary results is necessary (Antezana et al., 2009) but challenging, because the different types of information obtained in a biological system are computed through different databases. This is a prolix and complex task, and research is required on the optimal methods for managing and interrelating the data (Miyoshi et al., 2013). In genomics and proteomics, data derive from multiple fields, including gene expression, polymorphisms, metabolic maps, and genome sequencing, including the human genome.

Major advances in specialized software have facilitated efficient data management (Meslin and Cho, 2010), and novel user interfaces and interactive systems have allowed the convergence and management of large sets of bigdata (Collins, 2011). However, difficulties remain when integrating the heterogeneous data generated in research across different disciplines, and there is a need to overcome these problems. The past two decades have seen the development of invaluable and increasingly versatile biocomputing tools and techniques for integrating heterogeneous data and standardizing terminology (Splendiani et al., 2011). There has been also an exponential growth in instrumental methodologies to facilitate and promote collaborative and interdisciplinary research.

In this study, we investigate data management platforms related to highly-specialized research in genomics and proteomics, exploring the role of web-based databases in advances in these fields and their impact on scientific production.

ONTOLOGY OF DATA

Ontology is widely used in semantic computing and reasoning, and has become a key computing discipline in Life Sciences (Antezana et al., 2009). It makes a major contribution to effective metadata processing and analysis, improving the ability of researchers to make use of the increasing amount of heterogeneous knowledge available (Li et al., 2014). Over the past few years, ontology development has become a priority in Artificial Intelligence laboratories, and has become widespread on the Internet, facilitating information exchange among software researchers or agents. The precision and reutilization of all types of knowledge is improved in the context of the semantic web (Miyoshi et al., 2013), which is in productive use in multiple fields, including software engineering, web engineering, information systems, and applied computing in different areas (medicine, education, etc).

Thus, achievements in genomics and proteomics are recorded in accordance with the detailed ontological protocol established by each database (Brinkman et al., 2010). Each new entry contains a brief but precise description of its characteristic features and its relationship with other resources. Hence, ontologies constitute taxonomic indexes for organizing and accessing content (Goble and Stevens, 2008). Metadata allow the structuring of content, while ontologies offer the semantics for its management, providing a controlled vocabulary common to researchers needing to share information in a given domain (Stein, 2008). Ontology can give meaning to different types of data offered in varied formats in a given area (Benites et al., 2014).

The aim is to turn information into knowledge by its transformation into formalized structures that reference data under a normalized common scheme in a given domain of knowledge. Metadata not only specify the data scheme that appears in each case, but can also include information on how to establish axioms that can be applied in the different domains in which the stored knowledge is managed.

Hence, abundant and precise data can be yielded by searches of ontology-based metadata, which automatically offer information related to the requirements of the user though utilization of standardized web annotation schemes.

CONTRIBUTION TO THE DEVELOPMENT OF GENOMICS AND PROTEOMICS

The above strategies have mainly been used to optimize Internet/Intranet searches for information and to verify and test the validity of data, and the development of web resources has focused on the construction of databases. However, bioscientific research is increasingly carried out in projects that involve networks of different researchers who need

to communicate and share their preliminary results and the *know-how* acquired. The metadata generated in many studies must be adequately managed to draw useful conclusions and avoid the loss of essential information (Roncaglia et al., 2013). Furthermore, they represent a clear potential indicator for the detection of errors and optimization of efforts. These resources can also enhance the societal contribution of scientific achievements generated by researchers with common specific interests (Walk et al., 2008). As an illustrative example, analysis of the human genome led to numerous projects on the relationship of genetic variants with disease predisposition and the response to treatment, among others. These required the development of highly complex computing tools to manage the massive amounts of data from different databases (Bettembourg et al., 2014). These advances have improved the diagnosis and therapy of numerous complex diseases (Meslin and Cho, 2010), and research in human genetic diseases cannot be imagined without biocomputing.

DATABASES

There are two main types of database in genomics and proteomics: primary databases, which include information related to DNA and protein sequences, expression profiles, and 3D structure acquired; and secondary databases, which contain the results obtained from analysis of the primary databases, including the taxonomy of gene and protein families, active motifs or domains, mutations, polymorphisms, and the involvement of all these in disease development (Panet et al., 2011). The information in these databases is generally obtained from scientific journals. Various web-based bibliographic databases allow access to the articles published in these journals, notably (in descending order of access frequency) Pubmed (NCBI), Sciencedirect (Elsevier), Scopus (Elsevier), and Web of Science (Thomson Reuters). These databases facilitate advanced analyses of scientific production based on various classification criteria (Álvarez et al., 2014).

The development of information networks among researchers across different European countries has been a priority EU objective (6th and 7th Framework Programs, and “Horizon 2020”), so that research data can be absorbed, evaluated and utilized by other research groups and by the productive sector. Open access resources are increasingly numerous and influential, and major improvements have been achieved in the design, effectiveness, and usefulness of web-based databases (Herrera-Galeano et al., 2013; Chen et al., 2013), enabling the development of extensive “knowledge networks” to cover the needs of collaboration and information exchange in the scientific community (Lyne et al., 2013).

GenBank (ncbi.nlm.nih.gov/genbank) and EntrezProteins (ncbi.nlm.nih.gov/protein) of the NCB are the leading databases in genomics and proteomics. They offer simple and well- organized access to a large amount of well-referenced information, mainly related to gene/protein sequences, their translation, and their functions. They also offer rapid access to related topics available in other databases, especially other NCBI databases, including: the NCBI taxonomy database (ncbi.nlm.nih.gov /taxonomy), which includes the names and phylogenetic trees of more than 160,000 organisms based on molecular determinants; the above-mentioned Pubmed (ncbi.nlm.nih.gov/pubmed); the MMDB (ncbi.nlm. nih.gov/Structure/MMDB/mmdb. shtml), which provides experimental structural data obtained by crystallography and magnetic resonance imaging (MRI); and the OMIM (ncbi.nlm.nih.gov/omim), which contains valuable information on genetic disorders and other inherited characters, including their clinical manifestations, molecular genetic diagnosis, and prognosis. The HPRD (Human Protein Reference Database; hprd.org) is related to the OMIM and constitutes a global platform for integrating data on domain architecture, post-transational changes, and interaction networks, and on the diseases associated with each protein in the human proteome. The information on protein domains is especially relevant (Goh et al., 2013), presented as 3D units displaying sets of amino acid residues that remain highly preserved throughout the phylogenetic scale and are crucial for the function of the protein.

Other databases include the EMBL of the European Bioinformatics Institute (ebi.ac.uk/ena), DDBJ (DNA Data Bank of Japan; ddbj.nig.ac.jp) of the Japan National Institute of Genetics, and the UniprotKB (uniprot.org/help/uniprotkb). The EMBL offers one of the best repositories in Europe and includes the results of genome sequencing projects and patented applications. The UniprotKB integrates information from three key primary databases (PIR, Swiss-Prot, and TrEMBL). The PIR (Protein Information Resource; pir.georgetown.edu) offers information on protein sequences, while wiss-Prot web expasy org docs swiss- (prot guideline html) contains data on protein function, domain structure, post-translational modifications, and variants. The TrEMBL (Translation of EMBL Nucleotide Sequence Database) is complementary to the Swiss-Prot, providing information on nucleotide sequence translations and allowing large-scale functional characterization. We underline the close collaboration established among the managers of the above databases. Thus, GenBank, DDBJ, and EMBL continuously share their data with the International Nucleotide Sequence Database Collaboration, which was established over 15 years ago.

Other outstanding web consultation databases

include: GeneCards (genecards.org) of the Weizmann Institute of Science, which provides concise information in the fields of genomics, proteomics, transcriptomics, and functional genetics on all known and candidate human genes, using a highly intuitive web interface. When the gene encodes for an enzyme, the information can be completed by using the Enzyme database (enzyme.expasy.org), which offers an in-depth description of the enzymes with an Enzyme Commission (EC) number. This database follows the nomenclature recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB); MotifScan (myhits.isb-sib.ch/cgi-bin/motif_scan) enables sequence analysis and the search for similarities and homologies, and provides links with other databases for additional information, including: Prosite (secondary structure of proteins, domains; prosite.expasy.org), Pfam (relationship among domain configuration, protein family, and function prediction; pfam.sanger.ac.uk); InterPro (functional analysis and classification of proteins; ebi.ac.uk/interpro); and a large group of analysis databases, including Blocks (blocks.fhcrc.org), Prints (bioinf.manchester.ac.uk/dbbrowser/prints/ index.php), the Motif group (motif.stanford.edu/ projects.html), and a database with information related to genes, proteins, and mitochondrial diseases (Mitochondrial Proteome; mitop.de:8080/ mitop2).

Understandably, one of the most active fields in current biomedical research centers on drug discovery and development (Álvarez et al., 2012), which has been able to gain from major advances in related areas of biomedical sciences, including regenerative and stem cell medicine (Álvarez et al., 2013). It has also taken advantage of novel computing technologies, with web applications (Vihinen, 2014) such as a protein-small molecule *in silico* docking platform (e.g. SwissDock; swissdock.ch/docking). They permit the construction of predictive models for the interaction of proteins and enzymes with candidate molecules that may play an important role in the physiopathology or healing of disease (Krallinger et al., 2012). In this manner, computers become virtual laboratories for rapidly evaluating in probabilistic terms the usefulness of continuing with research on numerous molecules.

The use of a database that provides the true structure of the protein becomes indispensable for work with molecular docking systems, enabling the application of software to determine the possibilities of interaction with the corresponding ligands (Bettembourg et al., 2014). These databases, notably the PDB (Protein Data Bank; wwpdb.org), offer information on the 3D structure of proteins, generally based on X-ray crystallographic studies. This information can be imported into most molec-

Table 1. Impact of genomics and proteomics databases on scientific production. The databases reviewed, number of documents in which they are referenced in Google scholar (including articles and patents), and the main web address of the database. Search dates 5-10 February 2014

	References	**Web Site**
EMBL	415,000	http://www.ebi.ac.uk/
PDB	249,000	http://www.wwpdb.org/
GenBank	243,000	http://www.ncbi.nlm.nih.gov/genbank
NCBI Taxonomy	230,000	http://www.ncbi.nlm.nih.gov/taxonomy
OMIM	162,000	http://www.ncbi.nlm.nih.gov/omim
Swiss-Prot	83,700	http://web.expasy.org/docs/swiss-prot_guideline.html
PIR	82,200	http://pir.georgetown.edu/
DDBJ	53,900	http://www.ddbj.nig.ac.jp/
Hapmap	34,200	http://hapmap.ncbi.nlm.nih.gov/index.html.en
Enzyme	33,600	http://enzyme.expasy.org/
Pfam	33,500	http://pfam.sanger.ac.uk/
EntrezProteins	30,400	http://www.ncbi.nlm.nih.gov/protein
Prosite	22,200	http://prosite.expasy.org/
TrEMBL	18,300	http://www.ebi.ac.uk/
InterPro	14,100	http://www.ebi.ac.uk/interpro/
UniprotKB	12,100	http://www.uniprot.org/help/uniprotkb
Prints	7,720	http://www.bioinf.manchester.ac.uk/dbbrowser/prints/index.php
DGV	6,320	http://projects.tcag.ca/variation/
HGMD	4,950	http://www.hgmd.cf.ac.uk/ac/index.php
HPRD	4,840	http://hprd.org/
Blocks	4,530	http://blocks.fhcrc.org/
GeneCards	4,460	http://genecards.org/
Mitomap	2,910	http://mitomap.org/mitomap
MMDB	2,480	http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdb.shtml
Pupasuite	320	http://pupasuite.bioinfo.cipf.es/
F-SNP	136	http://compbio.cs.queensu.ca/F-SNP/
Mitop	65	http://mitop.de:8080/mitop2/
SYSNPs	20	http://www.sysnps.org/
Total	**1,755,951**	

ular docking platforms merely by specifying the PDB code, greatly simplifying access to the data.

Most genome variations correspond to alterations in a single nucleotide throughout a sequence (Single Nucleotide Polymorphism). Although changes are small, they are highly relevant and can affect the susceptibility to diseases and the effectiveness of specific treatments, being invaluable for studying the evolution of population groups, among other issues. Various databases gather this type of polymorphism, including pupasuite (pupasuite.bioinfo.cipf.es), F-SNP (compbio.cs.queensu.ca/F-SNP), and SYSNPs (sysnps.org). This information has contributed to the development of other databases for human genetic variants as a function of their geographic distribution (Hapmap, hapmap.org/index.html.en) or their genotype and phenotype (Database of Genomic Variants, DGV, projects.tcag.ca/variation).

The importance of mutations in oncogenesis, especially in proto-oncogenes and tumor suppressor genes, is reflected in the considerable space they occupy in databases devoted to genomics and proteomics, as in the case of the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff (hgmd.cf.ac.uk/ac/index.php) and the Human Mitochondrial Genome Database (Mitomap; mitomap.org), which gathers information on variations and mutations in mitochondrial DNA.

All of these databases are valuable e-science tools that enable advanced searches for specific information on the subject of analysis and facilitate management of these data (Arighi, 2011). They provide valuable and specific sets of biological data and protocols from well-equipped and -supported state-of-the-art laboratories and highly qualified research teams (Smith et al., 2013). These new portals enrich not only the scientific community but also those investigating computing systems related to the biomedical information ecosystem.

IMPACT OF GENOMIC AND PROTEOMIC WEB-DATABASES

The impact of databases in a given field of science is demonstrated by the frequency of their citation by articles in specialist journals, indicating their visibility and usefulness. Among the various generic databases used to search for scientific articles, including Pubmed, Sciencedirect, or Scopus, we selected Google Scholar. This platform offers major advantages relevant to our objectives. It allows combined searches in articles and patents and, unlike other platforms, it focuses on articles published in the scientific literature. Additionally, searches in the other databases are limited to searches based on titles, abstracts, and key words. An increasing number of studies have described Google Scholar as a search engine that surpasses even *Pubmed* in article classification in terms of their relevance and number of citations and the impact of the journals in which they are published (Nourbakhsh et al., 2014). Table 1 exhibits the results of our searches. The EMBL, PDB, GenBank, NCBI Taxonomy, and OMIM databases are all cited in over 100,000 references, with EMBL being cited more than 400,000 times. Other databases are available that are less cited, attributable to their high specialization. The usefulness of the databases based on generic genomic and proteomic ontologies is reflected by their more than 1,700,000 times that they have been cited.

CONCLUSIONS

We explored the role and importance of search engines based on the data management platforms of high specialization research assets or knowledge networks, centering on genomics and proteomics as high-value areas in biomedical research.

The platforms proved to be an important asset for stimulating the integration of research and collaborative data by information exchange. Highly specialized web search sites facilitate the retrieval and integration of data in specific areas and represent a current trend in life sciences research.

Advances in life science research in general, and in proteomics and genomics in particular, require the development of e-science in the form of specific search engines and websites. The nature of this type of investigation makes collaboration and interdisciplinarity essential; therefore, computing tools that permit working with and recording the large amount of information generated and improving data integration and exchange become indispensable. The knowledge management of 'big data' as such requires the development of computationally processable solutions.

The magnitude and complexity of research in these areas implies collaboration with multidisciplinary teams *via* extensive scientific networks, and this trend will continue to grow. Research groups show an increasing capacity for collaboration and interaction with others, linked to working practices geared towards the Internet. This change in conceptual framework has been responsible for introducing changes in the shaping of the competences of present-day scientists. The aim of an effective, efficient and collaborative scientific community is achievable based on the experience accumulated in this field on the utilization of novel tools and other innovations.

ACKNOWLEDGEMENTS

This article is related to the PhD thesis of C. González. This study was supported by the University of Granada and the Regional Government of Andalusia (Project P11-CTS-7651 and assis-

tance to the CTS-107 group).

REFERENCES

ÁLVAREZ P, BOULAIZ H, VÉLEZ C, RODRÍGUEZ-SERRANO F, ORTIZ R, MELGUIZO C, CARRILLO E, MARTÍNEZ-AMAT A, PRADOS J (2014) Qualitative and quantitative analyses of anatomists' research: evaluation of multidisciplinarity and trends in scientific production. *Scientometrics*, 98: 447-456.

ÁLVAREZ P, CARRILLO E, VÉLEZ C, HITA-CONTRERAS F, MARTÍNEZ-AMAT A, RODRÍGUEZ-SERRANO F, BOULAIZ H, ORTIZ R, MELGUIZO C, PRADOS J, ARÁNEGA A (2013) Regulatory systems in bone marrow for hematopoietic stem/progenitor cells mobilization and homing. *Bio Med Res Int*, 2013: 312656.

ÁLVAREZ P, MARCHAL J, BOULAIZ H, CARRILLO E, VÉLEZ C, RODRÍGUEZ-SERRANO F, MELGUIZO C, PRADOS J, MADEDDU R, ARANEGA A (2012) 5-Fluorouracil derivatives: a patent review. *Expert Opin Ther Pat*, 22: 107-123.

ANDERSON N, BRAGG C, HARTZLER A AND EDWARDS K (2012) Participant-centric initiatives: Tools to facilitate engagement in research. *ApplTransl Genomics*, 1: 25-29.

ANTEZANA E, KUIPERAND M, MIRONOV V (2009) Biological knowledge management: The emerging role of the Semantic Web TECHNOLOGIES. *Brief Bioinform*, 10: 392-407.

ARIGHI CN (2011) A tutorial on protein ontology resources for proteomic studies. *Methods Mol Biol*, 694: 77-90.

BAUX D, FAUGÈRE V, LARRIEU L, LE GUÉDARD-MÉREUZE S, HAMROUN D, BÉROUD C, MALCOLM S, CLAUSTRES M, ROUX AF (2008) UMD-USHbases: a comprehensive set of databases to record and analyse pathogenic mutations and unclassified variants in seven Usher syndrome causing genes. *Hum Mutat*, 29: E76-87.

BENITES F, SIMON S, SAPOZHNIKOVA E (2014) Mining rare associations between biological ontologies. *PLoS OnE;* 9: e84475.

BESKOW LM, SMOLEK SJ (2009) Prospective biorepository participants' perspectives on access to research results. *J Empir Res Hum Res Ethics*, 4: 99-111.

BETTEMBOURG C, DIOT C, DAMERON O (2014) Semantic particularity measure for functional characterization of gene sets using gene ontology. *PLoS One*, 9: e86525.

BRINKMAN RR, COURTOT M, DEROM D, FOSTEL JM, HE Y, LORD P, MALONE J, PARKINSON H, PETERS B, ROCCA-SERRA PH, RUTTENBERG A, SANSONE S, SOLTADOVA L, STOECKERT CH, TURNER JA, ZHENG J (2010) Modeling biomedical experimental processes with OBI. *J Biomed Semant*, 1: S7.

CHEN G, WANG C, SHI L, QU X, CHEN J, YANG J, SHI C, CHEN L, ZHOU P, NING B, TONG W, SHI T (2013) Incorporating the human gene annotations in different databases significantly improved transcriptomic and genetic analyses. *RNA*, 19: 479-489.

COLLINS, FS (2011) Reengineering translational Science: The time is right. *Sci Transl Med*, 3: 90, article 90cm17.

GOBLE C, STEVENS R (2008) State of the nation in data integration for bioinformatics. *J Biomed Inform*, 41: 687-693.

GOH WWB, FAN M, LOW HS, SERGOT M, WONG L (2013) Enhancing the utility of Proteomics Signature Profiling (PSP) with Pathway Derived Subnets (PDSs), performance analysis and specialized ontologies. *BMC Genomics*, 14: 1, 35.

HERRERA-GALEANO JE, HIRSCHBERG DL, MOKASHI V, SOLKA J (2013) OGA: An ontological tool of human phenotypes with genetic associations. *BMC Res Notes*, 6: 1, 511.

IHINEN M (2014) Variation Ontology for annotation of variation effects and mechanisms. *Genome Res*, 24: 356-364.

KRALLINGER M, LEITNER F, VAZQUEZ M, SALGADO D, MARCELLE C, TYERS M, VALENCIA A, CHATR-ARYAMONTRI A (2012) How to link ontologies and protein-protein interactions to literature: Text-mining approaches and the BioCreative experience. *Database (Oxford) 2012*, bas017.

LI C-X, WANG R-J, CHEN P, HUANG H, SU Y-R (2014) Interaction relation ontology learning. *J Comput Biol*, 21: 80-88.

LYNE M, SMITH RN, LYNE R, ALEKSIC J, HU F, KALDERIMIS A, STEPAN R, MICKLEM G (2013) Metabolic Mine: An integrated genomics, genetics and proteomics data warehouse for common metabolic disease research. *Database (Oxford) 2013*, bat060 Epub 2013 Aug 9.

MESLIN EM, CHO MK (2010) Research ethics in the era of personalized medicine: Updating science's contract with society. *Public Health Genomics,* 13: 378-384.

MIYOSHI NS, PINHEIRO DG, SILVA JR WA, FELIPE JC (2013) Computational framework to support integration of biomolecular and clinical data within a translational approach. *BMC Bioinformatics*, 14: 180.

NOURBAKHSH E, NUGENT R, WANG H, CEVIK C, NUGENT K (2012) Medical literature searches: A comparison of PubMed and Google Scholar. *Health Inf Libr J*, 29: 214-222.

PAN M, CONG P, WANG Y, LIN C, YUAN Y, DONG J, BANERJEE S, ZHANG T, CHEN Y, ZHANG T, CHEN M, HU P, ZHENG S, ZHANG J, QI M (2011) Novel LOVD databases for hereditary breast cancer and colorectal cancer genes in the Chinese population. *Hum Mutat*, 32: 1335-1340.

REBHOLZ-SCHUHMANN D, NENADIC G (2010) Biomedical Semantics: the Hub for Biomedical Research 2.0. *J Biomed Semant*, 1: 1.

RONCAGLIA P, MARTONE ME, HILL DP,

BERARDINI TZ, FOULGER RE, IMAM FT, DRABKIN H, MUNGALL CH, LOMAX J (2013) The gene ontology (go) cellular component ontology: integration with SAO (Subcellular Anatomy Ontology) and other recent developments. *J Biomed Semant*, 4: 20.

SMITH JR, PARK CA, NIGAM R, LAULEDERKIND SJ, HAYMAN GT, WANG S-J, LOWRY TF, PETRI V, PONS JD, TUTAJ M, LIU W, WORTHEY EA, SHIMOYAMA M, DWINELL MR (2013) The clinical measurement, measurement method and experimental condition ontologies: expansion, improvements and new applications. *J Biomed Semant*, 4: 26.

SPLENDIANI A, GÜNDEL M, AUSTYN JM, CAVALIERI D, SCOGNAMIGLIO C, BRANDIZI M (2011) Knowledge sharing and collaboration in translational research, and the DC-THERA directory. *Brief Bioinform*, 12: 562-575.

STEIN LD (2008) Towards a cyberinfrastructure for the biological sciences: Progress, visions and challenges. *Nat Rev Genet,* 9: 678-688.

SZALAY A, GRAY J (2006) 2020 computing: Science in an exponential world. *Nature*, 440: 413-414.

WALK S, PÖSCHKO J, STROHMAIER M, ANDREWS K, TUDORACHE T, NOY NF, NYULAS C, MUSEN M (2013) An interactive tool for visualizing the creation process behind collaboratively engineered ontologies. *Int J Semantic Web Inf Syst*, 9: 45-78.

WEBSTER YW, DOW ER, KOEHLER J, GUDIVADA RC, PALAKAL MJ (2011) Leveraging health social networking communities in translational research. *J Biomed Inform*, 44: 536-544.