 DOI 10.1002/pmic.200800002

The Gene Ontology – Providing a Functional Role in Proteomic Studies

Emily C. Dimmer[1], Rachael P. Huntley[1], Daniel G. Barrell[1], David Binns[1], Sorin Draghici[2], Evelyn B. Camon[1], Mike Hubank[3], Philippa J. Talmud[4], Rolf Apweiler[1], Ruth C. Lovering[4]

Until recently, biologists have concentrated on studying specific pathways or individual molecules as an approach to unravelling the intricate details of cellular events. However recent advances in high-throughput proteomic methodologies have made it possible to profile the global compositions of entire tissues, organelles or interactomes at specific time points or under particular developmental or disease states. While such methods provide researchers with a greater understanding of large-scale, complex biological changes, it has necessitated an increasing reliance on standardised annotation to link results to known biological activities. The Gene Ontology (GO) Consortium provides standardised functional annotations, which support high-throughput analyses and systems biology by enabling genes or proteins to be classified and grouped according to their function, involvement in a particular process or subcellular location, providing researchers with an indication of underlying mechanisms behind a certain phenotype.

In this article we provide an overview of the Gene Ontology, its uses and the software available for the analysis of proteomic data. Many of the examples presented in this paper relate to cardiovascular studies, the focus of a new GO annotation initiative.

Abbreviations:
GO, Gene Ontology
GOC, Gene Ontology Consortium
GOA, Gene Ontology Annotation database

[1] European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
[2] Wayne State University, Department of Computer Science, Detroit, MI, USA
[3] Molecular Haematology and Cancer Biology Unit, Institute of Child Health, London, UK
[4] Department of Medicine, University College London, Rayne Institute, London, UK

Correspondence:
Dr. Emily Dimmer
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD, UK
Tel.: +44-1223-494-654
Fax: +44-1223-494-468
edimmer@ebi.ac.uk

Introduction to the Gene Ontology

The Gene Ontology Consortium (GOC) is composed of both model organism groups and multi-species databases. The GOCs main objective is to provide three structured vocabularies (or ontologies) of terms to describe the *molecular functions* that gene products can normally carry out (*e.g.* 'lipid phosphatase activity'), the *biological processes* that gene products are involved in (*e.g.* 'cardiac muscle growth') and their subcellular locations (*cellular components*), where gene products occur (*e.g.* 'nuclear matrix'). Since March 2007 almost

25,000 unique GO terms have been created, these provide users with both a broad and detailed set of descriptors for many normal cellular activities or locations. The Gene Ontology (GO) also provides information on the association between genes and their products to GO terms. This is a major annotation resource that is much used to provide functional insights into the results of microarray and proteomics experiments. GO annotations are now cross-referenced by all major biological databases (UniProtKB, Ensembl, EntrezGene (No. 1–3 of Table 1)) and integrated into many third-party analytical tools [1, 2]. In order to be as descriptive and biologically accurate as possible, the creation and development of the GO is tightly managed, with information added and developed by a dedicated team of editors in response to requests from different research and database communities [1, 3].

The GO is structured as three directed acyclic graphs, where each term can have multiple relationships to parent (broader) terms (Figure 1). The relationships between GO terms provide an information-rich structure that can be manipulated (by expanding or contracting nodes of the ontology) by users, facilitating either a broad overview of a set of functional attributes, or a detailed view of specific processes in the ontology. Consequently, investigators are able to find functional insights for large sets of genes or proteins.

GO Annotations

GO annotations are associations between GO terms and gene or protein identifiers. Depending on the amount of functional data available, gene/protein identifiers can be annotated to multiple GO terms from each of the three gene ontologies (Figure 2). Annotations are produced by curation groups applying either manual or computational techniques. Each technique has its own advantages and disadvantages, but both require skilled biologists and software engineers to ensure the creation of conservative, high-quality annotations [4].

The 'manual' annotation of genes using GO terms involves highly-trained curators evaluating published data in the scientific literature and associating the most specific GO terms possible to gene/protein records. Manual annotations provide comprehensive, accurate and information-rich summaries of the functional knowledge for genes/proteins. Manual annotation is however, expensive, and annotation targets must be prioritised to ensure that maximum benefit can be provided to the GO user community [5].

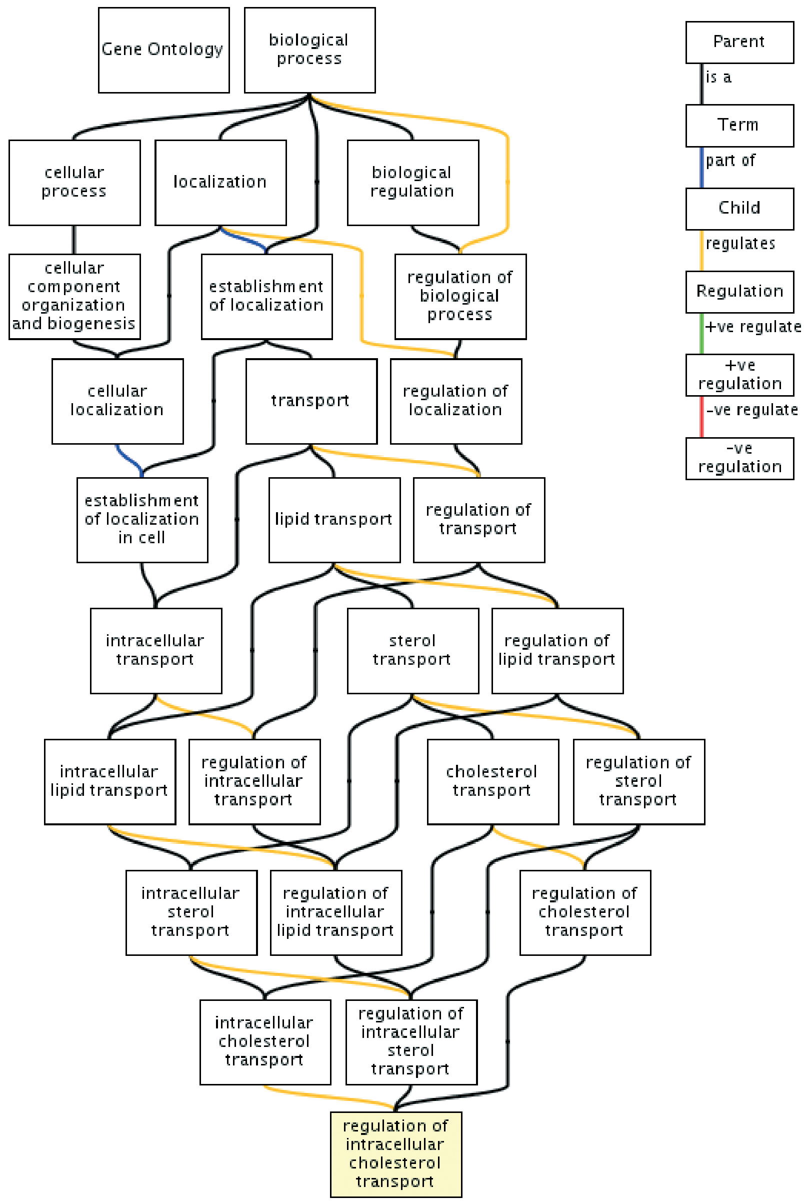
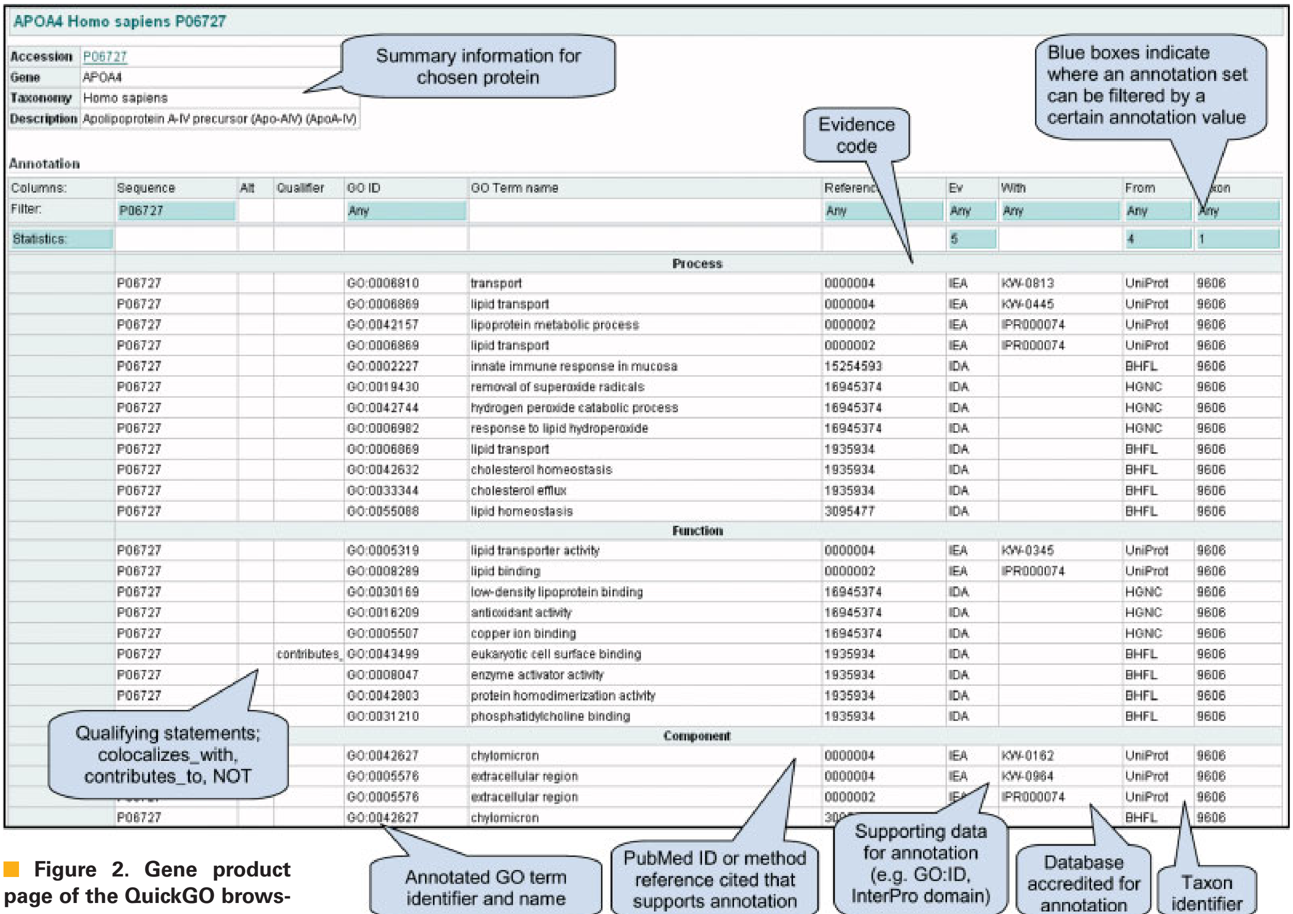


Figure 1. Section of the Gene Ontology for the Biological Process GO term 'regulation of intracellular cholesterol transport' (GO:0032383), showing ancestor terms and the different interconnecting relationships.

As the numbers of genes being identified increases, annotation requirements are also growing, and manual methods alone are unable to provide sufficient coverage. The large-scale assignment of GO terms to proteins using computational methods is a fast and efficient way of associating relatively high-level GO terms to a large number of sequences, and, with conservative usage, these methods can produce reliable, although often less detailed annotations [4, 6, 7].

"The GOCs main objective is to provide three structured vocabularies (or ontologies) of terms to describe the *molecular functions* that gene products can normally carry out, the *biological processes* that gene products are involved in and their subcellular locations (*cellular components*), where gene products occur"



Columns:	Sequence	Alt	Qualifier	GO ID	GO Term name	Referenc	Ev	With	From	xon
Filter:	P06727			Any		Any	Any	Any	Any	Any
Statistics:							5		4	1
	Process									
	P06727			GO:0006810	transport	0000004	IEA	KW-0813	UniProt	9606
	P06727			GO:0006869	lipid transport	0000004	IEA	KW-0445	UniProt	9606
	P06727			GO:0042157	lipoprotein metabolic process	0000002	IEA	IPR000074	UniProt	9606
	P06727			GO:0006869	lipid transport	0000002	IEA	IPR000074	UniProt	9606
	P06727			GO:0002227	innate immune response in mucosa	15254593	IDA		BHFL	9606
	P06727			GO:0019430	removal of superoxide radicals	16945374	IDA		HGNC	9606
	P06727			GO:0042744	hydrogen peroxide catabolic process	16945374	IDA		HGNC	9606
	P06727			GO:0006982	response to lipid hydroperoxide	16945374	IDA		HGNC	9606
	P06727			GO:0006869	lipid transport	1935934	IDA		BHFL	9606
	P06727			GO:0042632	cholesterol homeostasis	1935934	IDA		BHFL	9606
	P06727			GO:0033344	cholesterol efflux	1935934	IDA		BHFL	9606
	P06727			GO:0055088	lipid homeostasis	3095477	IDA		BHFL	9606
	Function									
	P06727			GO:0005319	lipid transporter activity	0000004	IEA	KW-0345	UniProt	9606
	P06727			GO:0008289	lipid binding	0000002	IEA	IPR000074	UniProt	9606
	P06727			GO:0030169	low-density lipoprotein binding	16945374	IDA		HGNC	9606
	P06727			GO:0016209	antioxidant activity	16945374	IDA		HGNC	9606
	P06727			GO:0005507	copper ion binding	16945374	IDA		HGNC	9606
	P06727		contributes_	GO:0043499	eukaryotic cell surface binding	1935934	IDA		BHFL	9606
	P06727			GO:0008047	enzyme activator activity	1935934	IDA		BHFL	9606
	P06727			GO:0042803	protein homodimerization activity	1935934	IDA		BHFL	9606
	[illegible]			GO:0031210	phosphatidylcholine binding	1935934	IDA		BHFL	9606
	Component									
	[illegible]			GO:0042627	chylomicron	0000004	IEA	KW-0162	UniProt	9606
	[illegible]			GO:0005576	extracellular region	0000004	IEA	KW-0984	UniProt	9606
	[illegible]			GO:0005576	extracellular region	0000002	IE[illegible]	IPR000074	UniProt	9606
	P06727			GO:0042627	chylomicron	30[illegible]	[illegible]		BHFL	9606

Figure 2. Gene product page of the QuickGO browser, showing annotations for the human APOA4 protein (P06727). All electronic and manual annotations are displayed. For further information about how some of these annotations are made, see the Annotation Tutorial <http://www.ebi.ac.uk/GOA/annotationexample.html >.

"GO annotations are associations between GO terms and gene or protein identifiers"

One of the mostly widely applied electronic annotation techniques exploits information already added to database entries by other curation groups. This method uses other controlled vocabularies, such as UniProt keyword or Enzyme Commission (E.C.) numbers or InterPro protein domains. When appropriate, terms in these external systems are manually mapped to corresponding terms in GO. Automatically searching a database with such mappings generates a table of GO annotations. Twenty-three different GO mappings are now available from the GO Consortium (No. 4 of Table 1). Different computational mappings provide sets of annotations with different levels of accuracy, coverage and reliability. An evaluation carried out on the annotations that were created using the Swiss-Prot keyword, Enzyme Commission (E.C.) number and InterPro to GO mappings found that while the GO terms predicted were likely to be less specific than those chosen manually, all strategies predicted the corrected GO term 91–100% of the time [4].

As the GOC is a widely geographically distributed, inter-group annotation project, it has been essential to establish well-documented, standardised annotation procedures. While some of the GO annotation conventions are described below, further information is available from the Annotation Guide on the GOC website (No. 5 of Table 1). Whether manually or electronically created, all GO annotations refer to the source of evidence that supports a GO term-protein association (often either a PubMed reference or a reference to a computational method) as well as an 'evidence code', which indicates to the user the category of evidence that was identified in the associated reference.

Evidence Codes

An integral part of a GO annotation is its evidence code. Currently there are 16 evidence codes used to describe the different categories of support that endorse an association between a GO term and a gene/protein identifier (Figure 3b). Broadly speaking, evi-

	Biological Databases	
1	UniProt KnowlegeBase	http://www.uniprot.org/
2	Ensembl	http://www.ensembl.org/
3	Entrez Gene	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene
	GOC and GOC Member Resources	
4	GO Mapping Files	http://www.geneontology.org/GO.indices.shtml
5	GO Annotation Guide	http://www.geneontology.org/GO.annotation.shtml
6	Evidence Code Guide	http://www.geneontology.org/GO.evidence.shtml
7	Qualifier Documentation	http://www.geneontology.org/GO.annotation.conventions.shtml#qual
8	AmiGO browser	http://amigo.geneontology.org/cgi-bin/amigo/go.cgi
9	QuickGO browser	http://www.ebi.ac.uk/quickgo
10	MGI GO browser	http://www.informatics.jax.org/searches/GO_form.shtml
11	GO Annotation downloads	http://www.geneontology.org/GO.current.annotations.shtml
12	GOA Website downloads	http://www.ebi.ac.uk/GOA/downloads.html
13	GO Consortium Tools Listing	http://www.geneontology.org/GO.tools.shtml
	Large-scale Functional Analyses Tools	
14	Onto-Express	http://vortex.cs.wayne.edu/ontoexpress/ (Figure 5)
15	FatiGO	http://babelomics.bioinfo.cipf.es/ (Figure 4)
16	The Ontologizer	http://www.charite.de/ch/medgen/ontologizer/
17	Blast2GO	http://babelomics.bioinfo.cipf.es/
	Cardiovascular Annotation Initiative	
18	Cardiovascular Annotation	http://www.geneontology.org/GO.cardio.shtml
	Initiative home pages	http://www.cardiovasculargeneontology.com
19	Cardiovascular association file	ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/bhf-ucl/gene_association.goa_bhf-ucl.gz
20	Editable wiki pages	http://wiki.geneontology.org/index.php/Cardiovascular
21	Feedback form	http://www.ebi.ac.uk/GOA/contactus.html
22	Cardiovascular mailing list	http://www.geneontology.org/GO.list.cardiovascular.shtml

Table 1. Useful URLs for sources of GO documentation, data and tools.

> **"An integral part of a GO annotation is its evidence code. Currently there are 16 evidence codes used to describe the different categories of support that endorse an association between a GO term and a gene/protein identifier."**

dence codes fall into three main groups: annotations based on published experimental data (such as an enzyme assay), non-experimental statements provided by an author or inferred by a curator (for instance inferring a nuclear localisation for an in vitro-characterised transcription factor) and finally evidence from computational predictions (No. 6 of Table 1).

Evidence codes help users to evaluate the sources of data that have been used in an annotation set. For well-annotated genomes, such as yeast, a user may be able to choose to use only manually-created annotations, whereas for other genomes such as pig, bovine or even human, users at present need to use both electronic and manual annotation sets to ensure that their sequences of interest have sufficient annotation data.

Annotation Qualifiers

Manual annotations can also include 'qualifiers', to provide an additional layer of information regarding the relationship between a protein and its associated GO term (No. 7 of Table 1). Three qualifiers are currently available: 'colocalizes_with' (to indicate a transient or peripheral association of the protein with an organelle or complex), 'contributes_to' (where a function of a protein complex is facilitated, but not directly carried out by one of its subunits) and 'NOT' (to indicate conflicting published data, or where in contrast to previous assumptions, a protein is not found to have a particular activity, location or process involvement). It is important to be aware of such qualifiers, for although they are infrequently used, they change the meaning of the associated annotation. Most importantly the 'NOT' qualifier produces the most drastic change in the interpretation of an annotation, and users of large datasets are advised to ensure 'NOT' annotations are appropriately applied by the large-scale functional analyses tool (Table 2).

> **"As the number of proteomic methods has increased, so has the number of ways in which Gene Ontology data has been applied to link from experimental results to current functional knowledge."**

Use of GO in Proteomics Studies

As the number of proteomic methods has increased, so has the number of ways in which Gene Ontology data has been applied to link from experimental results to current functional knowledge. For example, proteomes can simply be analysed with data from GO to provide a broad overview of the predominant activities a specific group of proteins. One such study, by Pasini *et al.*, looked at the locations and functions of proteins extracted from red blood cells, where the analysis with GO data indicated the expected involvement of these proteins in transporter, signal transducer or structural activities, however, it also implicated their involvement in other unexpected transcriptional or translational regulator activities [8]. In contrast, the use of GO data can also contribute towards determining the success of a particular sub-cellular enrichment strategy. For instance, Karsen *et al.* were able to demonstrate that following their membrane enrichment procedure, the number of proteins that were classified by GO as being localised to membranes increased from 23% in whole cell lysates to 49% in the membrane enriched fractions [9].

GO data have proved useful in generating hypotheses for the mechanisms underlying proteome-wide alterations in response to certain diseases or stress states: for instance in cardiac hypertrophy [10] or hypoxia [11]. In such studies, subsets of proteins found to be similarly over- or under-expressed can be clustered to highlight subsets sharing related GO annotation, providing an indication of the underlying cellular mechanisms that may account for an observed phenotype. One study by Pan *et al.* [10] on the adaptations in hyper- or hypocontractile hearts, looked at the composition of microsomal membrane fractions in mutant mouse cardiac tissue and found that over-expressed cardiac proteins were enriched in annotations to GO terms describing fat and carbohydrate metabolism and G-protein-dependent signalling pathways. This use of GO validated the investigators' proteomic methods and the results were consistent with the suggestion that the deregulation of calcium-dependent cardiac contractility resulted in compensatory cellular activities. GO has also helped when investigators need to select a subset of proteins to analyse in greater depth, for instance in investigations aimed at uncovering new sets of biomarkers for a certain disease, whereby enriched GO categories have indicated those processes that are being deregulated by the disease's onset [12, 13].

GO has also been used to provide functional data to protein interactome sets, providing a link between the protein binding network and the activities/locations of the participant proteins, potentially predicting whether a particular interactome is likely to occur *in vivo*. Dyer *et al.* [14] investigated interactions of human proteins with viral pathogens, where a GO analysis indicated that many different pathogens target the same processes in the human cell, such as regulation of apoptosis, even if they interact with different proteins.

Key Questions	Reasons
Does the tool enable the hierarchical structure of GO to be exploited?	GO analysis tools should be designed to improve the identification of functional groups within a dataset by allowing the user to manually consolidate genes associated with highly specific (child) GO terms to those with the higher (parent) GO terms in order to formulate and test specific biological hypotheses.
What is the release date of the data used by the tool?	Each month an average of 240 GO terms is added to the Gene Ontology and 1,500 GO annotations are added to the human GO dataset. Tools that infrequently download GO data will restrict analyses. Remember to include the release date(s) of the GO annotation dataset and ontology file used by the tool in any resulting publications. The tool should provide this information.
Does the tool correctly treat the GO annotations with the qualifier 'NOT' ?	This qualifier reverses the meaning of an annotation (see 'Qualifiers' section), so should be either removed from the analysis or used to calculate the amount of evidence against certain hypotheses involving the terms annotated with NOT.
Does the analysis tool enable concomitant functional profiling for all three GO categories?	In addition, an increasing number of tools also display other annotation data such as TRANSFAC regulatory motifs, BioCarta, KEGG and Reactome pathways
Is the type of identifier used in the assay directly accepted by the tool (*e.g.* probe IDs, RefSeq protein IDs, *etc.*) or will it be necessary to map one identifier type to another?	You may need to convert the identifiers of your gene list into those accepted by the tool. This can be an important source of errors since up to 20% of the identifiers can be routinely lost or incorrectly mapped during identifier translations [22]. Even if the type of identifier is directly accepted, is this the native identifier used in the analysis or is an internal identifier mapping being performed? If internal identifier mappings are being carried out, what are the sources of data and their release dates?
Does the tool test for both enrichment AND depletion of the GO terms?	Some tools only test for over-representation of the differentially expressed genes within the given GO term. However, both significantly enriched as well as depleted GO terms can be biologically meaningful.
Does the tool enable the user to submit their own GO annotation dataset or select specific evidence code-supported annotations for the analysis?	In many cases it will be appropriate for the user to define the background set of proteins used in the analysis. However, there are only a few species where filtering out certain evidence code supported annotations is appropriate (see Evidence Code section)
What is the statistical model used and are there several alternative models that the user may choose from?	One serious and widely neglected problem in GO profiling is that the same data submitted to different tools can provide widely different results for the same GO terms. Having the ability to specify the model allows the user to eliminate one variable to verify their analysis.
What choice of correction factors is available?	To compensate for the propagation of gene associations from each GO term to all their parent GO terms many tools give a choice of correction factors, such as Bonferroni, Holmes, false discovery rate (FDR) and Šidák [20]. Bonferroni or Šidák are suitable when less than 50 unrelated GO categories are involved, Holmes is more appropriate for larger numbers of unrelated GO categories and FDR is a good choice if several GO categories are related, *e.g.* contain several GO terms with a common parent.

Table 2. Considerations for high-throughput data analyses. Key points to consider when using GO based analysis tools.

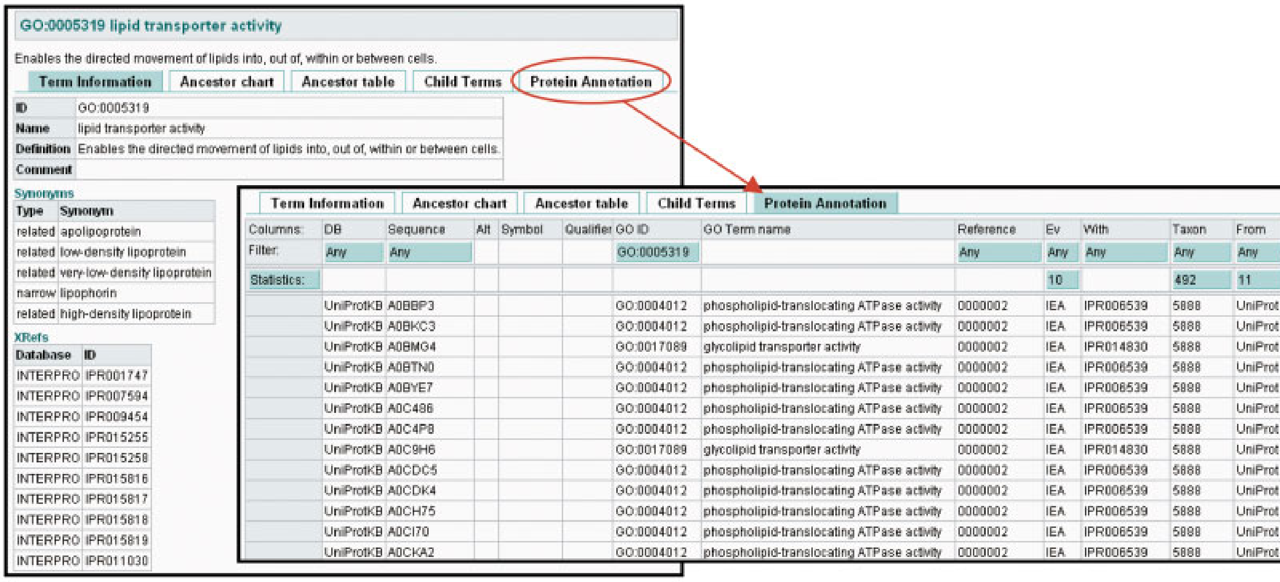


Figure 3. Viewing GO terms in QuickGO. (A) GO term pages of QuickGO, showing information for GO:0005319. The *Term Information* page shows a GO term's identifier, name and definition and any synonyms. The *Protein Annotation* page displays a table of gene products annotated to GO:0005319 and its children. By default all annotations (electronic and manual) are shown. Most of the information in the annotation table is clickable with explanatory text appearing or links available as appropriate. (B) Filtering annotations in QuickGO. Annotations can be filtered using a variety of annotation values, including evidence codes. In this example, the proteins that have been manually annotated with GO:0005319 or its children are displayed simply by clicking on the "Evidence" filter and checking the "Manual All" box. Filtering options are available for all columns which have a blue filter box.

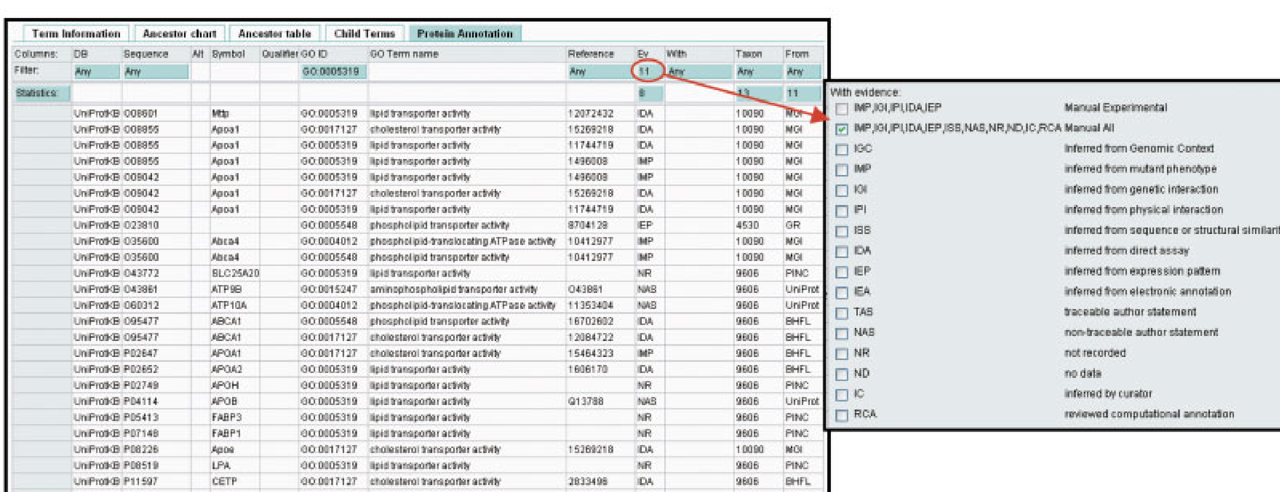
Viewing and Applying Gene Ontology Data

Browsing the GO

To enable the scientific community to effectively use the GO vocabularies and annotations, a number of web-based tools have been developed by members of the GOC and by third parties, to search, browse and view the GO hierarchy and annotations. Eighteen different GO browsers have been developed, and include the official GOC browser AmiGO, as well as the QuickGO and the MGI GO browsers (No.8–10 of Table 1). Each browser has a number of unique features and it is worth trying a couple of different tools initially to compare their functionality. The QuickGO browser, produced by the GOA group at the EBI, enables querying of individual or groups of proteins for associated GO annotations (Figure 2), or individual or multiple GO terms for details of associated proteins (Figure 3a). QuickGO's interface supports a variety of filter options that fully exploit the information contained in GO annotations to enable users to personalise their view of the annotations available (Figure 3b) and their annotation downloads.

Downloading GO Annotations

Large sets of GO annotations can also be downloaded in bulk. GO annotations are supplied by the GO Consortium in simple 15 column, tab-delimited 'gene association files'. Each row of



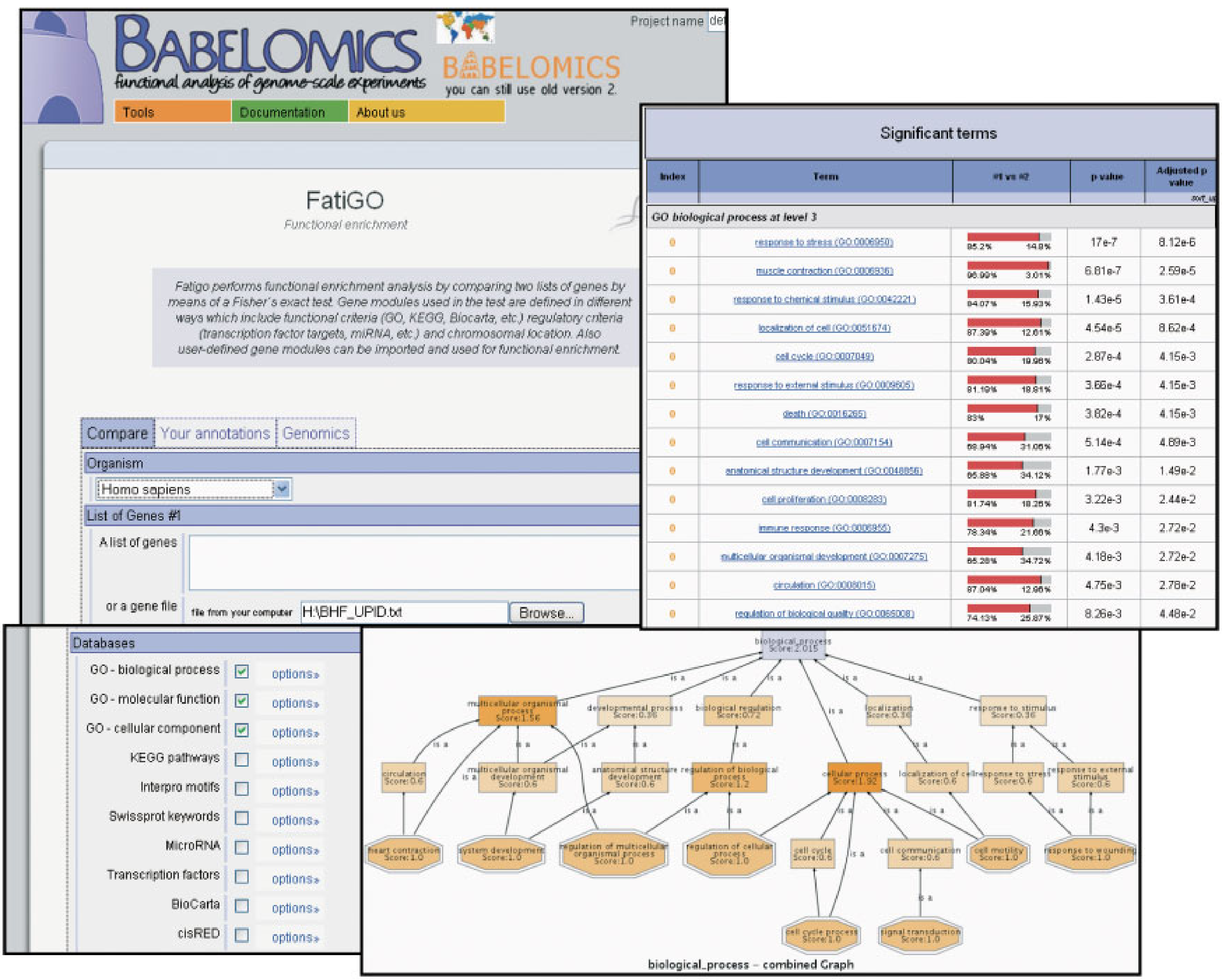
the file contains all the information required for one GO term-gene association including details of the sequence being annotated, the GO term and the reference used.

The GOC maintains a central repository where GO annotations that have been contributed by member databases are stored (No. 11 of Table 1). It is important to recognise that GOC member databases annotate to a number of different gene or protein identifiers. For instance the GOA group annotates to UniProtKB accessions (*e.g.* P43291) whereas MGI annotate to MGI identifiers (*e.g.* MGI:1913363). Therefore, depending on what the annotations will be used for, a mapping between identifiers may need to be carried out.

Gene association files can also be downloaded from member database sites, which sometimes contain different groupings of annotations. For example, users can download the GOA-UniProt gene association file from the GOA website, (or subsets of annotations *via* the QuickGO browser), where manual and electronic GOA annotations have been supplemented with manual annotations from twelve other model organism databases

"To enable the scientific community to effectively use the GO vocabularies and annotations, a number of web-based tools have been developed by members of the GOC and by third parties, to search, browse and view the GO hierarchy and annotations."

Figure 4. The FatiGO Functional Enrichment tool. FatiGO performs functional enrichment analysis by comparing two lists of genes by means of a Fisher's exact test. Queries can be performed over the three GO categories simultaneously. The results table shows significant GO terms in the list of interest together with the percentage of genes from each list annotated to the term and their *p*-values. The DAG viewer tool allows visualisation of the significant GO terms as a GO graph.



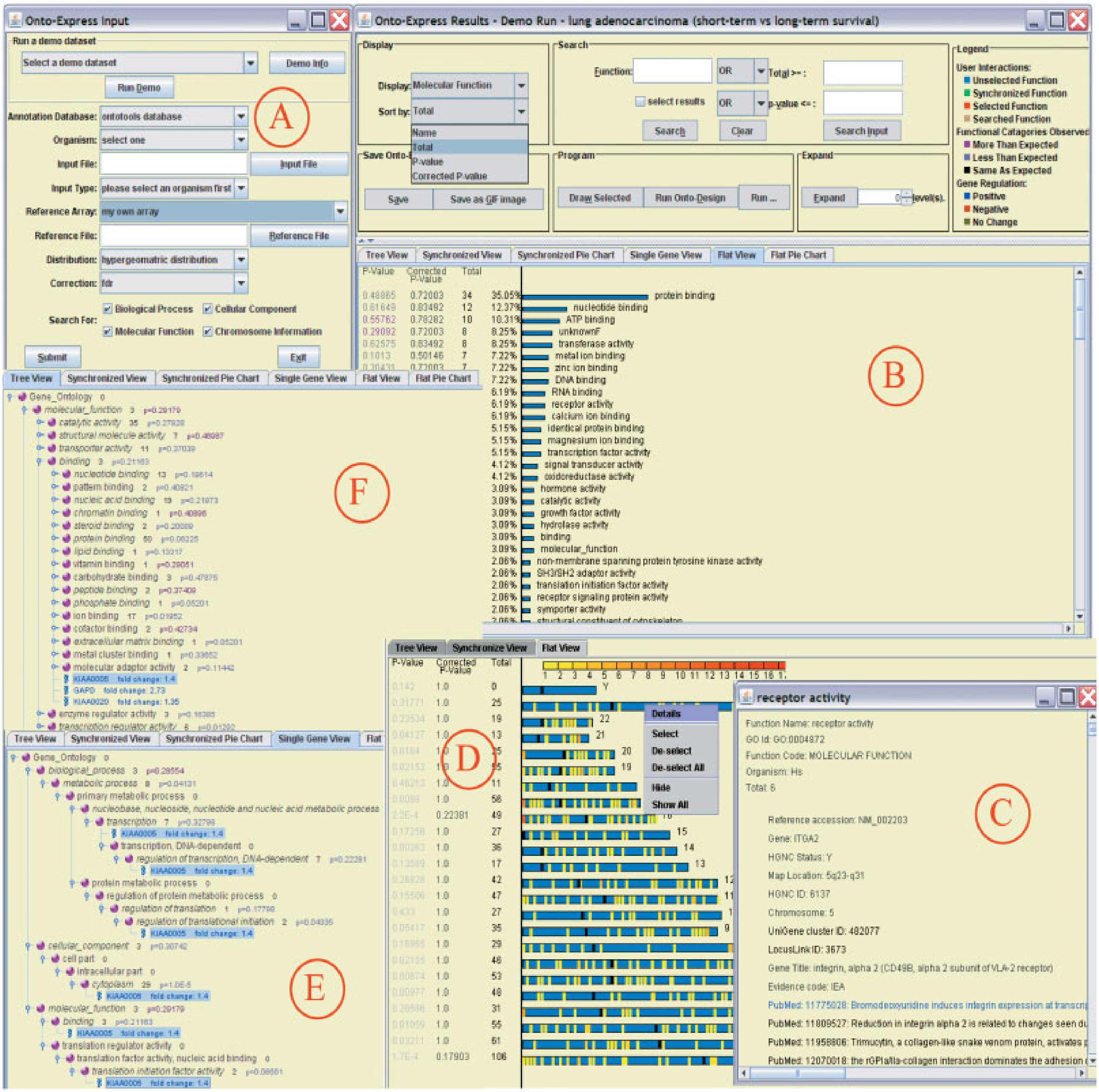


Figure 5. Onto-Express. Onto-Express (OE) generates functional profiles for a list of differentially expressed genes. (A) Users can choose from more than 300 microarrays from 10 manufacturers or submit their own array as a reference. The analysis produces both a 'Flat View' (in which the analysis is done with the terms with which the gene are directly annotated (shown in B), as well as a ,Tree view' showing the results in the context of the GO hierarchy and allowing the users to perform a custom cut if needed by expanding or collapsing nodes (shown in F). (C) Details of genes annotated to each term; (D) A chromosomal location analysis; (E) View of GO annotations for a gene under the 'Single Gene View' tab.

(including MGI, RGD and AgBase), creating a valuable multi-species GO annotation resource [15] (No. 12 of Table 1).

Functional Analysis of Large Datasets

Many tools have been developed to allow users to query GO annotation data with lists of gene, protein or probe identifiers identified from proteomic or other high-throughput experiments (Figures 4 and 5). A list of tools for high-throughput analysis is available from the GOC tools page (No. 13 of Table 1) and, as with the GO browsers, it is worth looking at several analysis tools to find one that suits your needs. Some examples of analysis tools popularly used by proteomics groups include: Onto-Express [16], FatiGO [17], the Ontologizer [18] and Blast2GO [19] (No. 14–17 of Table 1).

Several publications are available that provide useful reviews about the GO annotation tools that are available and how to use GO annotations [2, 20]. However, key points to consider when using GO based analysis tools have been outlined in Table 2.

Acknowledgments

The Cardiovascular GO Annotation Initiative is funded by the British Heart Foundation (SP/07/007/23671). The GOA Project is funded by a P41 grant from the National Human Genome Research Institute (HG002273), and the British Heart Foundation (SP/07/007/23671).

The authors have declared no conflict of interest.

Supporting Information

Microsoft PowerPoint presentation with animations and more detail see online version

Call for Community Participation

In order for the GOC to be able to provide the best possible set of annotations, it is important that scientific experts are able to offer feedback on GO annotation and ontology development activities, to ensure that current knowledge has been comprehensively reviewed and correctly summarised. Consequently a variety of systems have been put in place to encourage scientists to contribute to the GOC data. Groups wishing to contribute experimentally derived GO annotations, can either contact the GOC directly via the mailing list (gohelp@geneontology.org) or alternatively contact a member database that most closely represents the experimental species (for instance scientists interested in human GO annotations should contact the GOA group (goa@ebi.ac.uk) or the Cardiovascular Annotation Initiative (goannotation@ucl.ac.uk)). Individual scientists interested in contributing GO annotation information can do so via editable wiki pages; a feedback form or directly email a GOC member group [21] (No. 18–22 of Table 1).

References

[1] Bada, M., Stevens, R., Goble, C., Gil, Y. *et al.*, A short study on the success of the Gene Ontology. *J. Web Semantics* 2004, *1*, 235–240.

[2] Khatri, P. and Drăghici, S., Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* 2005, *21*, 3587–3595.

[3] Diehl, A.D., Lee, J.A., Scheuermann, R.H., Blake, J.A., Ontology development for biological systems: Immunology. *Bioinformatics* 2007, *23*, 913–915.

[4] Camon, E.B., Barrell, D.G., Dimmer, E.C., Lee V. *et al.*, An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. *BMC Bioinformatics*, 2005, *6*, S17.

[5] Dimmer, E., Berardini, T.Z., Barrell, D., Camon, E., Methods for gene ontology annotation. *Methods Mol. Biol.* 2007, *406*, 495–520.

[6] Biswas, M., O'Rourke, J.F., Camon, E., Fraser, G. *et al.*, Applications of InterPro in protein annotation and genome analysis. *Brief Bioinform.* 2002, *3*, 285–295.

[7] Done, B., Khatri, P., Done, A., Drăghici, S., Predicting novel Human Gene Ontology Annotations using semantic analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2008, DOI:10.1109/TCBB.2008.29.

[8] Pasini, E.M., Kirkegaard, M., Mortensen, P., Lutz, H.U. *et al.*, In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood*, 2006, *108*, 791–801.

[9] Karsan, A., Blonder, J., Law, J., Yaquian, E. *et al.*, Proteomic analysis of lipid microdomains from lipopolysaccharide-activated human endothelial cells. *J. Proteome Res.* 2005, *4*, 349–357.

[10] Pan, Y., Kislinger, T., Gramolini, A.O., Zvaritch, E. *et al.*, Identification of biochemical adaptations in hyper- or hypocontractile hearts from phospholamban mutant mice by expression proteomics. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 2241–2246.

[11] Boraldi, F., Annovi, G., Carraro, F., Naldini, A. *et al.*, Hypoxia influences the cellular cross-talk of human dermal fibroblasts. A proteomic approach. *Biochim. Biophys. Acta* 2007, *1774*, 1402–1413.

[12] Shi, M., Jin, J., Wang, Y., Beyer, R.P. *et al.*, Mortalin: A protein associated with progression of Parkinson disease? *J. Neuropathol. Exp. Neurol.* 2008, *67*, 117–124.

[13] Fliser, D., Novak, J., Thongboonkerd, V., Argilés, A. *et al.*, Advances in urinary proteome analysis and biomarker discovery. *J. Am. Soc. Nephrol.* 2007, *18*, 1057–1071.

[14] Dyer, M.D., Murali, T.M., Sobral, B.W., The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathogens*, 2008, *4*, e.32.

[15] Camon, E., Magrane, M., Barrell, D., Lee V. *et al.*, The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* 2004, *32*, D262–266.

[16] Khatri, P., Bhavsar, P., Bawa, G., Draghici, S., Onto-Tools: An ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. *Nucleic Acids Res.* 2004; *32*, W449–456.

[17] Al-Shahrour, F., Díaz-Uriarte, R., Dopazo, J., FatiGO: A web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, 2004, *20*, 578–580.

[18] Robinson, P.N., Wollstein, A., Böhme, U., Beattie, B., Ontologizing gene-expression microarray data: Characterizing clusters with Gene Ontology. *Bioinformatics*, 2004, *20*, 979–981.

[19] Conesa, A., Götz, S., García-Gómez, J.M., Terol, J. *et al.*, Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 2005, *21*, 3674–3676.

[20] Rhee, S.Y., Wood, V., Dolinski, K., Draghici, S., Use and misuse of the Gene Ontology (GO) Annotations. *Nature Review Genetics*, 2008 (in press).

[21] Lovering, R., Dimmer, E., Khodiyar, V.K., Barrell, D.G. *et al.*, Cardiovascular GO Annotation Initiative Year 1 Report: Why Cardiovascular GO? *Proteomics*, 2008 (in press).

[22] Draghici, S., Sellamuthu, S. and Khatri, P., Babel's tower revisited: A universal resource for cross-referencing across annotation databases. *Bioinformatics*, 2006, *22*, 2934–2939.