Improvements to cardiovascular Gene Ontology

Article *in* Atherosclerosis · December 2008

DOI: 10.1016/j.atherosclerosis.2008.10.014 · Source: PubMed

3 authors, including:

Ruth Caroline Lovering
University College London
92 PUBLICATIONS **4,790** CITATIONS

SEE PROFILE

Philippa J Talmud
University College London
236 PUBLICATIONS **7,125** CITATIONS

SEE PROFILE

Published as: *Atherosclerosis*. 2009 July ; 205(1): 9–14.

Improvements to cardiovascular Gene Ontology

Ruth C. Lovering[a,*], **Emily C. Dimmer**[b], and **Philippa J. Talmud**[a]

[a]Department of Medicine, University College London, Rayne Institute, 5 University Street, London WC1E 6JF, UK.

[b]GOA Project, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

Abstract

Gene Ontology (GO) provides a controlled vocabulary to describe the attributes of genes and gene products in any organism. Although one might initially wonder what relevance a 'controlled vocabulary' might have for cardiovascular science, such a resource is proving highly useful for researchers investigating complex cardiovascular disease phenotypes as well as those interpreting results from high-throughput methodologies. GO enables the current functional knowledge of individual genes to be used to annotate genomic or proteomic datasets. In this way, the GO data provides a very effective way of linking biological knowledge with the analysis of the large datasets of post-genomics research. Consequently, users of high-throughput methodologies such as expression arrays or proteomics will be the main beneficiaries of such annotation sets. However, as GO annotations increase in quality and quantity, groups using small-scale approaches will gradually begin to benefit too. For example, genome wide association scans for coronary heart disease are identifying novel genes, with previously unknown connections to cardiovascular processes, and the comprehensive annotation of these novel genes might provide clues to their cardiovascular link. At least 4000 genes, to date, have been implicated in cardiovascular processes and an initiative is underway to focus on annotating these genes for the benefit of the cardiovascular community. In this article we review the current uses of Gene Ontology annotation to highlight why Gene Ontology should be of interest to all those involved in cardiovascular research.

Keywords

Gene Ontology; Cardiovascular science; High-throughput analysis; Chromosome 9

1 Introduction

Until recently, the study of specific pathways or individual molecules has been the major approach to understanding the intricate molecular and cellular details associated with cardiovascular processes and disease, with thousands of publications each year adding to our accumulated knowledge of these systems. However, genome-sequencing projects have led to the identification of thousands of genes in higher vertebrates, the majority of which are only

*Corresponding author. Tel.: +44 20 7679 6968; fax: +44 20 7679 6212. E-mail: r.lovering@ucl.ac.uk.

characterised by their sequence and genomic location, with their potential involvement in cardiovascular systems awaiting experimental investigation. High-throughput methodologies, such as expression arrays or proteomics are providing substantial information about the properties of these newly identified genes, through the detailed characterisation of the molecular composition of entire tissues, cells or organelles at both specific developmental and specific disease states or through protein binding or cellular location studies. Consequently, such investigations provide researchers with the potential to rapidly increase our understanding of complex interactions and biological functions within the cardiovascular system. However integrating such high-throughput data with the detailed published experimental knowledge about the function of individual genes is an essential step that is necessary to ensure that all experimental approaches make an impact on current research projects. Fortunately, the Gene Ontology Consortium (GOC) has been developing terms to describe the functional attributes of gene products, across all species, in a consistent and computer-friendly manner to enable the integration of all of these data. This system of terms, called Gene Ontology (GO), enables the accumulated knowledge about individual gene products and their functional domains to be included in individual gene records, in biological sequence databases, and within high-throughput analysis software. This information can then be applied by high-throughput analysis software to aid in the interpretation of large datasets. By providing current functional knowledge in a format that can be exploited by high-throughput technologies, the GOC provides a major freely available public annotation resource that can help bridge the gap between data collation and data analysis [1] (www.geneontology.org).

The success of GO rests on the philosophy behind it; GO was designed by biologists to improve data integration and consequently enables genes to be classified and grouped together according to their functional properties [2–4]. At times the English language can be rather vague, with the majority of words having a variety of subtly different meanings. Similarly, scientific terms or phrases can have dual meanings. Consequently, one of the primary aims of GO is to create a single, explicit definition for each biological term so that these terms can be applied and interpreted consistently by all biologists. All such terms are provided as three structured vocabularies of terms (ontologies) that describe the *molecular functions* that gene products normally carry out, the *biological processes* that gene products are involved in and lastly the *subcellular locations* (*cellular components*) where gene products are active. For example, the annotations for cholesteryl ester transfer protein (*CETP*) include the *Molecular Function* term: ‘cholesterol transporter activity’, the *Biological Process* term: ‘reverse cholesterol transport’ and the *Cellular Component* term: ‘high-density lipoprotein particle’; whereas the annotations for troponin C type 1 (*TNNC1*) include the *Molecular Function* term: ‘troponin I binding’, the *Biological Process* term: ‘regulation of muscle contraction’ and the *Cellular Component* term: ‘troponin complex’.

The terms in GO are structured as directed acyclic graphs, where each term can have multiple relationships to broader ‘parent’ and more specific ‘child’ terms (Fig. 1). This hierarchical structure produces a representation of biology that allows a greater amount of flexibility in data analysis than would be afforded by a format based on a simple list of terms. Users can manipulate the structure to see either a broad overview of the general functional attributes presented by a set of data, or focus in on specific sections in the ontology to investigate in greater detail.

The second resource supplied by the GOC are datasets of GO terms associated with the appropriate genes and their products, thus providing a resource of diverse detailed functional annotation for many different species [1] (www.geneontology.org/GO.current.annotations.shtml). These annotations are created by 13 different annotation groups, including Gene Ontology Annotation @ EBI (GOA), FlyBase, and the Mouse Genome Database. Depending on the amount of published data available, gene/

protein identifiers can be annotated with multiple GO terms from any, or all, of the three gene ontologies (Fig. 1). Annotations can be produced either by a curator reading published scientific papers and manually creating each association or by a software engineer applying computational techniques to predict associations [5]. These two broad categories of techniques have their own advantages and disadvantages, but both require skilled biologists and software engineers to ensure that conservative, high-quality annotations are created. The annotation of each gene is therefore a potentially long laborious process, which for a highly studied gene like *TNF* could take several days (Fig. 2) or, for a more recently described gene like *CDKN2B*, may only take a few hours (Fig. 1).

2 Use of GO in high-throughput studies

As the number of high-throughput methodologies has increased, so has the number of ways in which GO annotation data has been exploited to link experimental results to current functional knowledge.

Proteomes and differentially regulated mRNAs can be analysed with GO data to provide an overview of the predominant activities the constituent proteins are involved in or where they are normally located. For example Ashley et al. [6] used GO to compare the genes up-regulated in *de novo* atherosclerosis with those associated with in-stent restenosis. They found a significant proportion of genes up-regulated following *de novo* atherosclerosis were associated with inflammatory processes, whereas a high proportion of in-stent restenosis up-regulated genes had GO terms indicating an involvement with cell growth and association with the extracellular matrix [6].

Often the generation of hypotheses to explain proteome-wide alterations in response to certain diseases, such as cardiac hypertrophy [7], or stress states, such as hypoxia [8], rely on the use of GO annotation data. In such studies an indication of underlying cellular mechanisms that may account for an observed phenotype can be obtained using GO to cluster subsets of proteins that share related GO annotation, and found to be similarly over- or under-expressed in the disease or stress state. For example, Pan et al. [7] found over-expressed cardiac microsomal membrane proteins in mouse hyper- or hypocontractile hearts were enriched with GO terms describing fat and carbohydrate metabolism and G-protein-dependent signalling pathways [7]. Enrichment of these GO terms validated the investigators proteomic method and was consistent with the suggestion that the deregulation of calcium-dependent cardiac contractility resulted in compensatory growth activities.

The ability to review experimental results with respect to known functional information has also proved useful when investigators need to select a subset of proteins to analyse in greater depth in order to identify new sets of biomarkers for a certain disease. This approach has enabled investigators of buccal carcinoma [9], Parkinson disease [10] and chronic kidney disease [11] to identify new biomarkers for these diseases. Furthermore, in all of these reports the enriched GO categories indicated disease-associated deregulated processes.

GO can also be used to provide a link between the protein binding network and the activities/locations of the participant proteins. Use of cellular component GO annotations can aid data visualisation or confirm whether a particular set of interactions is likely to occur *in vivo*. Dyer et al. [12] used GO data to investigate interactions of human proteins with viral pathogens and found that many different pathogens target the same processes in the human cell, such as regulation of apoptosis, even though they may interact with different proteins. Similarly, many studies have focused on a ‘guilt-by-association’ hypothesis, where the involvement of proteins in a particular pathway can be hypothesised in relation to the processes their interacting proteins

carry out. To this end GO annotations are integrated in the GEOMI [13] and Cytoscape [14] network visualization tools.

A number of proteomic investigations have found that GO data provides an indispensable resource to indicate the success of a particular subcellular enrichment strategy or large scale confocal microscopy analyses [15–18]. Kislinger et al. [15] used GO data to verify that their subcellular fractionation protocol efficiently isolated subcellular compartments. For example, of the nearly 600 proteins detected exclusively in the nuclear fractions, nearly half were either annotated solely to the nucleus or had a function known to be localized within the nucleus (e.g. transcription factors [15]). Barbe et al. [18] applied GO annotation to validate the protein subcellular locations identified using protein-specific antibodies and large scale confocal microscopy analysis, and in this case 80% of the subcellular locations identified in human cell lines were supported by existing GO annotation data [18].

Despite the wide variety of applications that GO is used for, there are many aspects to biological processes that are not addressed by this database. In particular, GO only describes the normal, physiological function of a protein, rather than the pathological function of a protein in a diseased situation. Furthermore, the dynamic relationships between protein function, its cellular location (including its intracellular location, cell specificity and developmental specific expression) and how this relationship influences the biological processes a protein is involved in are not currently represented by GO. Protein interaction databases such as BioGrid [19], Biomolecular Interaction Network Database [20], Human Protein Reference Database [21] and IntAct [22] enable complex protein interaction networks to be investigated. However, at present there is no single database that enables complex biological relationships to be investigated. The content of GOC database is influenced by the curation groups who are submitting data, therefore some groups of organisms, such as viruses, are not well represented by this database. Details about some of the other available ontologies, such as cell type or human disease ontologies, are available at the Open Biomedical Ontologies web site (www.obofoundry.org).

3 Where to find Gene Ontology and GO annotations

A large range of applications have been developed specifically for the visualization of GO and its associated annotation data, and the computational and statistical analysis of large datasets with GO [4,23]. Currently there are 48 tools for gene expression and microarray analysis and 20 GO browsers, all of which are listed on the GOC tools web page (www.geneontology.org/GO.tools.shtml). In addition to their inclusion in dedicated GO browsers, GO annotations are also imported into many of the top biological databases, including UniProtKB, Ensembl, EntrezGene and GeneCards.

To enable the scientific community to effectively use the GO vocabularies and annotations, a number of web-based tools have been developed by both members of the GOC and third parties to search, browse and view the GO hierarchy and annotations (Table 1). Such browsers include AmiGO (amigo.geneontology.org/cgi-bin/amigo/go.cgi), QuickGO (www.ebi.ac.uk/ego) and the MGI GO browser (www.informatics.jax.org/searches/GO_form.shtml). Each of these browsers has its own unique features and varies slightly in the functionality available to users. All three of these browsers have a variety of hyperlinks from the protein records to the GO term records and to the appropriate publication references, and AmiGO and QuickGO both provide a variety of filtering options to modify the search query, enabling users to tailor their download set. In AmiGO, the GOC browser, the GO term records clearly show the number of proteins annotated with that term by all GOC databases, however the protein records only include manual annotation data. The QuickGO browser, produced by the GOA group at the European Bioinformatics Institute (EBI), provides a simple view of the GO annotation data available, including data derived from both manual and electronic pipelines. Users can also

tailor the annotation set displayed by specifying lists of sequence identifiers or GO terms, and use QuickGO to map between sequence identifier types (www.ebi.ac.uk/GOA/annotationexample.html). The MGI database is the primary source of mouse GO annotations, and its 'Gene Ontology Classifications' records provide three different views of the associated terms, a text summary, a tabulated list and a graphical view. Alternatively, for high-throughput analysis tools entire sets of annotations can be downloaded from the GOC (www.geneontology.org/GO.current.annotations.shtml) or EBI (ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/) and following the launch of the Cardiovascular GO Annotation Initiative (www.cardiovasculargeneontology.com) [24] funded by the British Heart Foundation, a cardiovascular-specific dataset is now also available as a single download (ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/bhf-ucl/).

Many tools have been developed to allow users to perform a bulk query of GO using a list of gene, protein or probe identifiers, that have been identified from proteomics or other high-throughput experiments [4,23]. A full list of these tools is available at the GOC web site (www.geneontology.org/GO.tools.shtml#micro).

4 Manual annotation—the way forwards?

In theory, text-mining software can be used to extract Gene Ontology terms from the scientific literature and associate these to individual gene records [25,26]. However, the written style of scientific papers represents a substantial obstacle that needs to be overcome for such 'automated' methods to be successful. Papers are written as interesting text rather than as a list of results and conclusions. Consequently, a single function for one gene product may be discussed using a series of similar yet non-overlapping descriptions. The result is that although text mining is proving to be a useful curation aid for locating papers, it is unable to provide correct, detailed descriptions of the functions and processes with which a gene product is involved [27].

Manual gene annotation is an expensive and time-consuming alternative, but does ensure high-quality annotation is achieved. The most accurate, detailed source of functional information for annotation would ideally be provided by scientists working in relevant fields, rather than by a curator who may be unfamiliar with the specific details of a set of genes. However, as annotation guidelines tend to be highly detailed, and annotation a highly time-consuming activity for many bench scientists, bringing the research community closer to the annotation process is a difficult yet valuable goal for curation projects [28,29].

The Cardiovascular GO Annotation Initiative is prioritising the manual annotation of human genes implicated with cardiovascular processes. In order to do this a list of over 4000 cardiovascular-associated genes has been compiled, drawing on the expertise of several advisors for the project [24] (www.cardiovasculargeneontology.com). A concentrated effort to improve the information content provided in the manual GO annotation of genes involved in the cardiovascular system is now underway, this will ensure a comprehensive, up-to-date summary of the current literature for each gene is available [24].

4.1 Request for community participation

For the Cardiovascular GO Annotation Initiative to have a substantial impact in this area of biology, it is important that experts from the cardiovascular community are consulted to ensure that the current accumulated knowledge has been comprehensively reviewed and correctly summarised by the dedicated curation team. Consequently a variety of online facilities have been put in place to encourage scientists to contribute to this initiative [24] (Table 1). Scientists interested in contributing can either simply supply the curators with details of key experimental publications for curation or can provide more detailed information or commentary, such as

reviewing particular annotation sets and pointing out any experimental data that might be missing, wrong or controversial. This information can be sent to GO curators either by direct email to the GO Annotation teams based at UCL, GOannotation@ucl.ac.ukail, or EBI, goa@ebi.ac.uk, using the online user feedback form www.ebi.ac.uk/GOA/contactus.html, or by editing wiki pages set up for this purpose wiki.geneontology.org/index.php/Cardiovascular. Submitted data is regularly reviewed and scientists are notified when their contributions have been incorporated.

5 Conclusion

The Cardiovascular GO Annotation Initiative is currently the only GO annotation project focused on a specific field of human biology. The analysis of the large datasets post-genomics research offers the potential to rapidly advance our understanding of both the basic and clinical aspects of atherosclerosis. Hopefully the cardiovascular community will recognise the potential of this resource, actively contribute to it and reap the rewards in the near future!

Acknowledgements

Thanks to Varsha Khodiyar for critical reading of this manuscript. The Cardiovascular GO Annotation Initiative, funded by the British Heart Foundation (SP/07/007/23671) and the GOA Project, funded by a P41 grant from the National Human Genome Research Institute (NHGRI, grant HG002273) and the British Heart Foundation (SP/07/007/23671), are both GO Consortium members. GO annotations made by the authors are included in the GOC database.

References

[1]. Consortium T.G.O. Creating the gene ontology resource: design and implementation. Genome Res 2001;11:1425–1433. [PubMed: 11483584]

[2]. Ashburner M. Ball C.A. Blake J.A. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–29. [PubMed: 10802651]

[3]. Lomax J. Get ready to GO! A biologist's guide to the Gene Ontology. Brief Bioinform 2005;6:298–304. [PubMed: 16212777]

[4]. Dimmer EC, Huntley RP, Barrell DG, Binns D, Draghici S, Camon EB, Hubank M, Talmud PJ, Apweiler R, Lovering RC. The Gene Ontology; providing a functional role in proteomic studies. Proteomics 2008; July [Epub ahead of print].

[5]. Camon E. Magrane M. Barrell D. The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. Genome Res 2003;13:662–672. [PubMed: 12654719]

[6]. Ashley E.A. Ferrara R. King J.Y. Network analysis of human in-stent restenosis. Circulation 2006;114:2644–2654. [PubMed: 17145989]

[7]. Pan Y. Kislinger T. Gramolini A.O. 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. [PubMed: 14982994]

[8]. Boraldi F. Annovi G. Carraro F. Hypoxia influences the cellular cross-talk of human dermal fibroblasts. A proteomic approach. Biochim Biophys Acta 2007;1774:1402–1413. [PubMed: 17904921]

[9]. Staab C.A. Ceder R. Jagerbrink T. Bioinformatics processing of protein and transcript profiles of normal and transformed cell lines indicates functional impairment of transcriptional regulators in buccal carcinoma. J Proteome Res 2007;6:3705–3717. [PubMed: 17696463]

[10]. Shi M. Jin J. Wang Y. Mortalin: a protein associated with progression of Parkinson disease? J Neuropathol Exp Neurol 2008;67:117–124. [PubMed: 18219256]

[11]. Perco P. Wilflingseder J. Bernthaler A. Biomarker candidates for cardiovascular disease and bone metabolism disorders in chronic kidney disease: a systems biology perspective. J Cell Mol Med 2008;12:1177–1187. [PubMed: 18266955]

[12]. Dyer M.D. Murali T.M. Sobral B.W. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog 2008;4:e32. [PubMed: 18282095]

[13]. Ho E. Webber R. Wilkins M.R. Interactive three-dimensional visualization and contextual analysis of protein interaction networks. J Proteome Res 2008;7:104–112. [PubMed: 18020406]

[14]. Cline M.S. Smoot M. Cerami E. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2007;2:2366–2382. [PubMed: 17947979]

[15]. Kislinger T. Rahman K. Radulovic D. Cox B. Rossant J. Emili A. PRISM, a generic large scale proteomic investigation strategy for mammals. Mol Cell Proteomics 2003;2:96–106. [PubMed: 12644571]

[16]. Cao R. He Q. Zhou J. High-throughput analysis of rat liver plasma membrane proteome by a nonelectrophoretic in-gel tryptic digestion coupled with mass spectrometry identification. J Proteome Res 2008;7:535–545. [PubMed: 18166008]

[17]. Stevens S.M. Duncan R.S. Koulen P. Prokai L. Proteomic analysis of mouse brain microsomes: identification and bioinformatic characterization of endoplasmic reticulum proteins in the mammalian central nervous system. J Proteome Res 2008;7:1046–1054. [PubMed: 18271522]

[18]. Barbe L. Lundberg E. Oksvold P. Toward a confocal subcellular atlas of the human proteome. Mol Cell Proteomics 2008;7:499–508. [PubMed: 18029348]

[19]. Stark C. Breitkreutz B.J. Reguly T. Boucher L. Breitkreutz A. Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res 2006;34:D535–539. [PubMed: 16381927]

[20]. Alfarano C. Andrade C.E. Anthony K. The Biomolecular Interaction Network Database and related tools 2005 update. Nucleic Acids Res 2005;33:D418–424. [PubMed: 15608229]

[21]. Mishra G.R. Suresh M. Kumaran K. Human protein reference database—2006 update. Nucleic Acids Res 2006;34:D411–414. [PubMed: 16381900]

[22]. Kerrien S. Alam-Faruque Y. Aranda B. IntAct—open source resource for molecular interaction data. Nucleic Acids Res 2007;35:D561–565. [PubMed: 17145710]

[23]. Khatri P. Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics 2005;21:3587–3595. [PubMed: 15994189]

[24]. Lovering R.C. Dimmer E. Khodiyar V.K. Cardiovascular GO annotation initiative year 1 report: why cardiovascular GO? Proteomics 2008;8:1950–1953. [PubMed: 18491309]

[25]. Aerts S. Haeussler M. van Vooren S. Text-mining assisted regulatory annotation. Genome Biol 2008;9:R31. [PubMed: 18271954]

[26]. Cohen K.B. Hunter L. Getting started in text mining. PLoS Comput Biol 2008;4:e20. [PubMed: 18225946]

[27]. Camon E.B. Barrell D.G. Dimmer E.C. An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. BMC Bioinform 2005;6(Suppl 1):S17.

[28]. Ashburner M. Bergman C.M. Drosophila melanogaster: a case study of a model genomic sequence and its consequences. Genome Res 2005;15:1661–1667. [PubMed: 16339363]

[29]. Bieri T. Blasiar D. Ozersky P. WormBase: new content and better access. Nucleic Acids Res 2007;35:D506–510. [PubMed: 17099234]

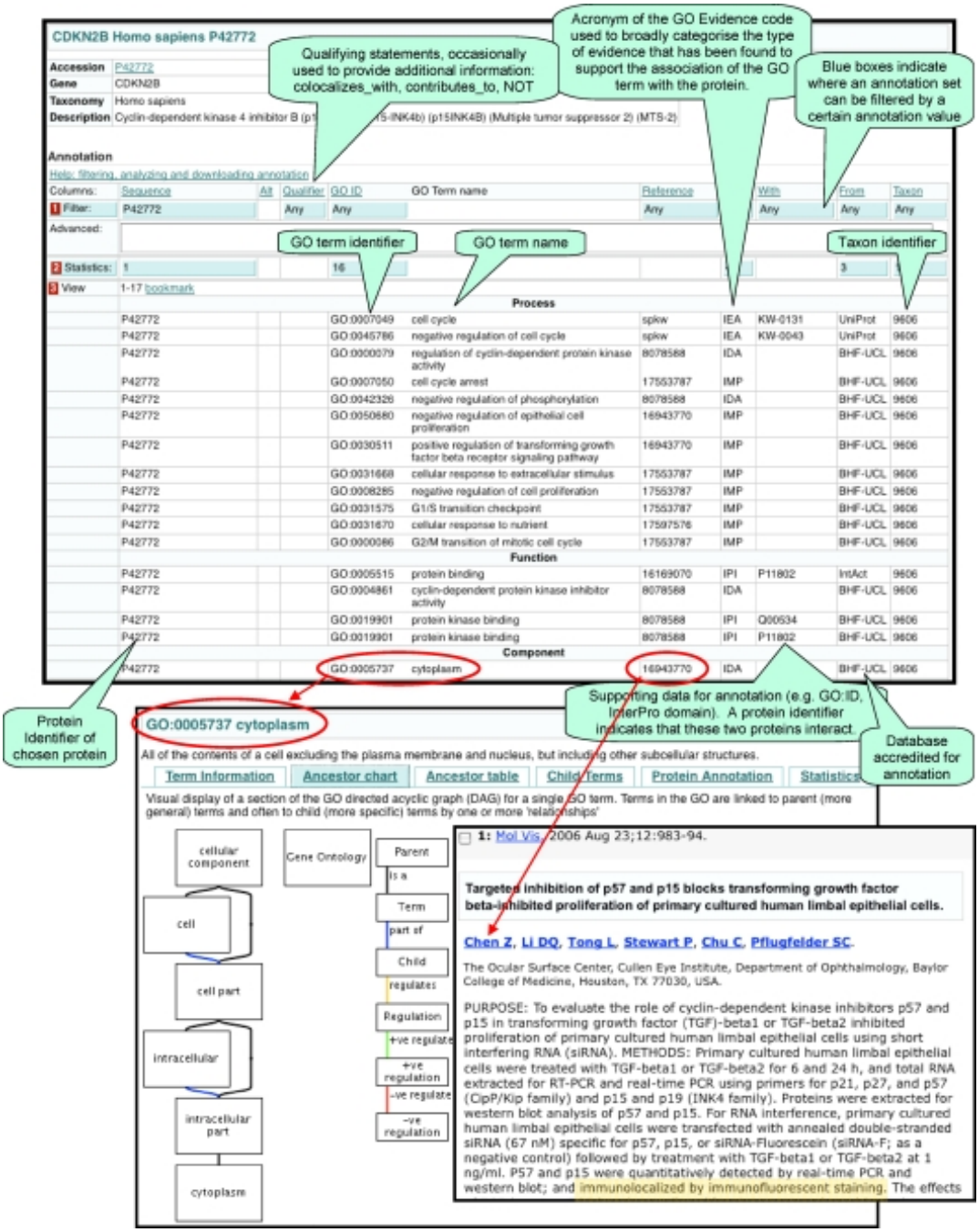

Fig. 1.

Protein record page of the QuickGO browser, showing all annotations for the human CDKN2B protein (www.ebi.ac.uk/ego/GProtein?ac=P42772). The QuickGO protein record page displays a short summary for the chosen UniProtKB protein accession number, together with associated GO annotations. Descriptions about the information in each column are provided in green call-outs, red arrows and circles show some of the additional information hyperlinked to this protein record. In the QuickGO GO term ancestor chart the term 'cytoplasm' is shown to be a child of the broader parent term intracellular part. The GO term 'cytoplasm' was associated with the CDKN2B record based on the direct assay 'immunofluorescent staining' and given the evidence code 'IDA', an acronym for 'inferred from direct assay' (text in the

abstract highlighted yellow). 17 different evidence codes can be used to categorise the type of evidence available to support the annotation, one of which (IEA) is use to indicate an 'electronic' source for the annotation, the remaining 16 are all used for manually curated annotations. For further information about the use of evidence codes see www.geneontology.org/GO.evidence.shtml and for an explanation about how some of these annotations are made, see the Annotation Tutorial at EBI (www.ebi.ac.uk/GOA/annotationexample.html). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

TNF Homo sapiens P01375

Accession	P01375
Gene	TNF
Taxonomy	Homo sapiens
Description	Tumor necrosis factor precursor (TNF-alpha) (Tumor necrosis factor ligand superfamily member 2) (TNF-a) (Cachectin) [Contains: Tumor necrosis factor, membrane form; Tumor necrosis factor, soluble form]

GO ID	GO Term name	Reference	Ev	With
	Process			
GO:0050901	leukocyte tethering or rolling	10820279	IDA	
GO:0033209	tumor necrosis factor-mediated signaling pathway	10748004	IMP	
GO:0032800	receptor biosynthetic process	10443688	IDA	
GO:0043193	positive regulation of gene-specific transcription	14512626	IDA	
GO:0006919	caspase activation	14512626	IDA	
GO:0002439	chronic inflammatory response to antigenic stimulus	14512626	IMP	
GO:0051384	response to glucocorticoid stimulus	10443688	IDA	
GO:0050715	positive regulation of cytokine secretion	10443688	IDA	
GO:0032715	negative regulation of interleukin-6 production	10443688	IDA	
GO:0002740	negative regulation of cytokine secretion during immune response	10443688	IDA	
GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	17922812	IDA	
GO:0042346	positive regulation of NF-kappaB import into nucleus	17922812	IDA	
GO:0001819	positive regulation of cytokine production	17922812	IDA	
GO:0009615	response to virus	10490959	IDA	
GO:0045080	positive regulation of chemokine biosynthetic process	10490959	IDA	
GO:0051091	positive regulation of transcription factor activity	10748004	IDA	
GO:0045941	positive regulation of transcription	10748004	IDA	
GO:0043122	regulation of I-kappaB kinase/NF-kappaB cascade	10748004	IDA	
GO:0006954	inflammatory response	10748004	IDA	
GO:0000187	activation of MAPK activity	10748004	IDA	
GO:0001934	positive regulation of protein amino acid phosphorylation	10748004	IDA	
GO:0006916	anti-apoptosis	10748004	IDA	
GO:0045071	negative regulation of viral genome replication	10490959	IDA	
GO:0032722	positive regulation of chemokine production	10490959	IDA	
GO:0048661	positive regulation of smooth muscle cell proliferation	16518841	IDA	
GO:0006927	transformed cell apoptosis	3883195	IDA	
GO:0016481	negative regulation of transcription	16895791	IDA	
GO:0000060	protein import into nucleus, translocation	16280327	IDA	
GO:0051092	positive regulation of NF-kappaB transcription factor activity	16280327	IDA	
GO:0051092	positive regulation of NF-kappaB transcription factor activity	15790681	IDA	
	Function			
GO:0042802	identical protein binding	14512626	IDA	
GO:0005164	tumor necrosis factor receptor binding	14512626	IDA	
GO:0005125	cytokine activity	10748004	IDA	
GO:0005515	protein binding	2848815	IPI	P19438
	Component			
GO:0005615	extracellular space	10443688	IDA	
GO:0005615	extracellular space	3932069	IDA	
GO:0055037	recycling endosome		ISS	P06804
GO:0001891	phagocytic cup		ISS	P06804
GO:0009897	external side of plasma membrane		ISS	P06804

Fig. 2.

Protein record page of the QuickGO browser, showing the manual annotations for the human TNF protein (www.ebi.ac.uk/ego/GProtein?ac=P01375). TNF is described in over 55,000 publications, however the annotations currently associated with human TNF have been extracted from only 15 of these publications. A review of these annotations by scientists working on TNF could increase the number and quality of annotations associated with this gene.

Table 1

Useful Gene Ontology links.

Action	URL
Choose the right software tool for your dataset [4,23]	www.geneontology.org/GO.tools.shtml
Download GO annotations	www.geneontology.org/GO.current.annotations.shtml, wiki.geneontology.org/index.php/Defining_CV-associated_gene_list, ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/
Find out about the GO annotation process [3–5]	www.ebi.ac.uk/GOA/annotationexample.html
Browse GO terms and GO hierarchy	amigo.geneontology.org/cgi-bin/amigo/go.cgi, www.ebi.ac.uk/ego
View gene product specific GO annotations	www.ebi.ac.uk/ego/GProtein?ac=P42772, amigo.geneontology.org/cgi-bin/amigo/gp-assoc.cgi?gp=UniProtKB:P01375
Understand GO evidence codes	www.geneontology.org/GO.evidence.shtml
Find out about the Cardiovascular GO Annotation Initiative [24]	www.cardiovasculargeneontology.com
Search the cardiovascular gene list	www.ebi.ac.uk/GOA/CVI
Suggest improvements to specific gene annotations	wiki.geneontology.org/index.php/Cardiovascular, www.ebi.ac.uk/GOA/contactus.html, www.ucl.ac.uk/silva/cardiovasculargeneontology/feedback
Suggest improvements to specific GO terms	www.ebi.ac.uk/GOA/contactus.html, www.geneontology.org/GO.contacts.shtml
Contact GO curators	GOannotation@ucl.ac.ukail, goa@ebi.ac.uk

Published as: *Atherosclerosis*. 2009 July ; 205(1): 9–14.