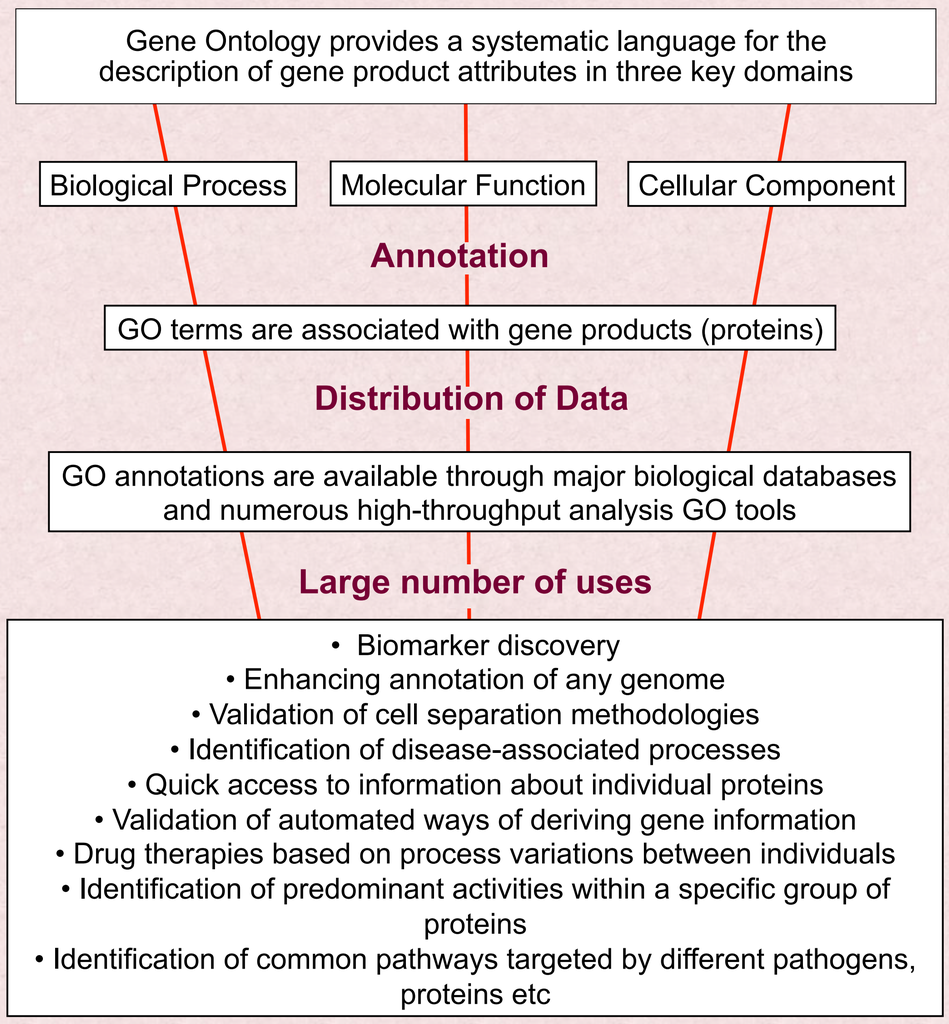
Can HVP GO further? Expanding human Gene Ontology…….

Ruth C Lovering*[1], Varsha K Khodiyar[1], Emily C Dimmer[2], Claire O'Donovan[2], Pete Scambler[3], Mike Hubank[4], Rolf Apweiler[2] and Philippa J Talmud[1]

[1]Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, Rayne Building, 5 University Street, London WC1E 6JF.
[2]UniProt Gene Ontology Annotation Project, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD.
[3]Molecular Medicine Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH.
[4]Molecular Hematology and Cancer Biology Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH.
Corresponding author: r.lovering@ucl.ac.uk

Gene Ontology[1] (GO) provides a controlled vocabulary to describe the attributes of genes and gene products in any organism. This resource is proving highly useful for researchers investigating complex phenotypes such as cardiovascular disease, as well as those interpreting results from high-throughput methodologies. By providing current functional knowledge in a format that can be exploited by high-throughput technologies, the GO Consortium provides a freely available key public annotation resource that can help bridge the gap between data collation and data analysis (www.geneontology.org).



Where can you find GO annotations?

In addition to the GO Consortium browsers QuickGO and AmiGO, which display GO annotation data and ontologies, GO annotations are included in the majority of functional analysis tools and displayed by popular knowledgebases, including NCBI Gene, Ensembl, UniProt, GeneCards and Wikipedia. Here the GO data provides users with a brief summary of the known function of a protein, the processes it is involved in and its location within the cell.

Annotation of the human proteome at UCL

To date, the Cardiovascular GO Annotation Initiative, funded by the British Heart Foundation (BHF), has associated 17,200 GO terms with 2000 human proteins, providing 10% of the manual human GO data. In the process the initiative has instigated the development of 1500 new GO terms.

Community annotation appeal

Despite the cardiovascular and other, focused efforts to capture the molecular function, biological processes and cellular locations of the human proteome using GO terms, the annotation of many proteins still does not reflect our understanding of these proteins.

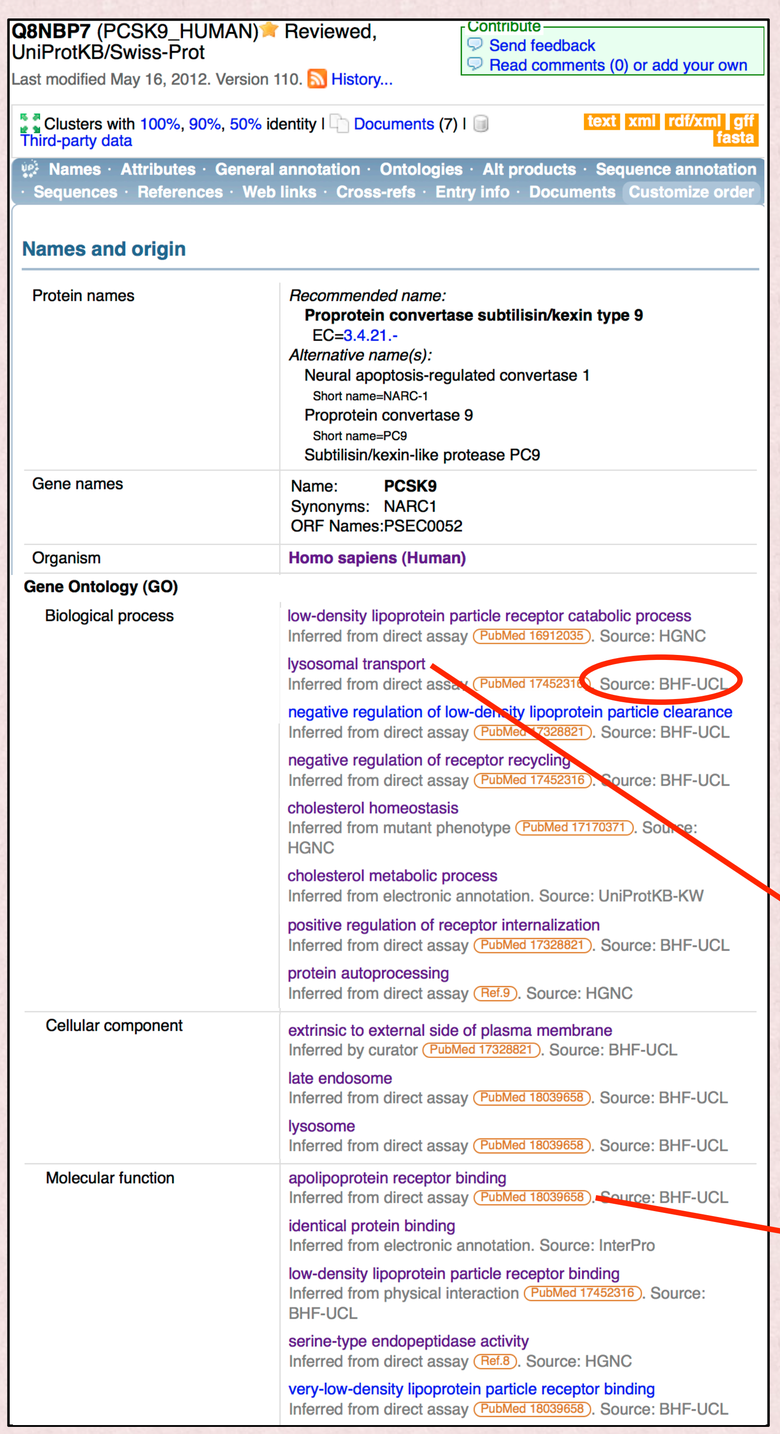
Curators are faced with the challenge of identifying which publications provide experimental evidence to support annotations, amongst the hundreds or thousands, which have accumulated for each protein. Consequently, although the basic function of a protein, as well as the cellular locations, may be described by GO, the biological processes a protein is associated with may not fully described. In addition, very little of the knowledge gleaned from human variation data is currently being captured by GO.

Improve the annotation of your favourite gene

One of the most time consuming aspects of GO annotation is finding the right paper to use for annotation. This time could be reduced if members of the Human Variome Project were willing to spend a small amount of time sending details of key publications to GO curators. Additionally, GO annotations would be improved if expert scientists reviewed the GO annotations available for their favourite gene and sent comments about missing or wrong annotations to the GO curators. In this way we could ensure comprehensive GO annotation of HVP genes. For example, the number of experimentally supported annotations associated with PCSK9 doubled from 9 to 20 following the submission of just 3 papers (Figure 1) to the BHF supported GO curators.

How to contribute to GO

Annotation or publication suggestions can be sent through our **feedback form,** www.ucl.ac.uk/cardiovasculargeneontology/feedback, or by **email** to goannotation@ucl.ac.uk



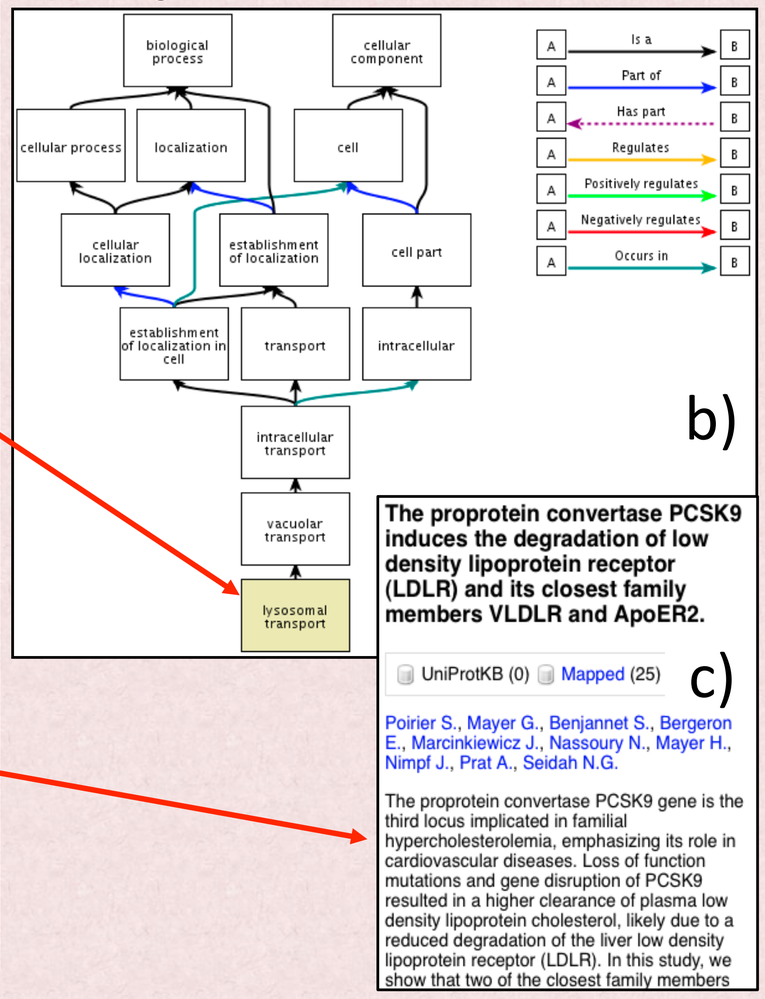


Figure 1. GO annotation of PCSK9. a) screen shot of part of the UniProtKB PCSK9 protein record, www.uniprot.org, illustrating some of the GO terms associated with this protein. The annotations created by the BHF-funded curators are indicated by 'source: BHF-UCL', circled in red. Webpage hyperlinks are indicated with red arrows to: **b)** the QuickGO, www.ebi.ac.uk/QuickGO, view of the ancestor chart for the GO term lysosomal transport; **c)** the literature citation supporting the annotation.

References

1. The Gene Ontology Consortium (2012) The Gene Ontology: enhancements for 2011. Nucleic acids research 40: D559-564.

Grant: SP/07/007/23671

BEATING HEART DISEASE TOGETHER

www.ucl.ac.uk/cardiovasculargeneontology
www.ucl.ac.uk/cardiovasculargeneontology/feedback