Bringing Gene Ontology to Cardiovascular Research

Ruth Lovering[1], Varsha Khodiyar[1], Tony Sawford[2], Emily Dimmer[2] Pete Scambler[3], Mike Hubank[4], Rolf Apweiler[2] and Philippa Talmud[1]

[1]Centre for Cardiovascular Genetics, UCL Department of Medicine, Rayne Institute 5 University Street London WC1E 6JF.
[2]Gene Ontology Annotation (GOA) 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.

Gene Ontology (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 GOC provides a freely available key public annotation resource that can help bridge the gap between data collation and data analysis (www.geneontology.org).

The UCL-based GO annotation team works with bench scientists to improve the annotation of human proteins.

For more information about contributing to the annotation of the human genome contact GOannotation@UCL.ac.uk

TNF Homo sapiens P01375

GO Identifier	GO Term Name	Reference
GO:0002740	negative regulation of cytokine secretion involved in immune response	PMID:10443688
GO:0005615	extracellular space	PMID:18355445
GO:0045429	positive regulation of nitric oxide biosynthetic process	PMID:8383325
GO:0051044	positive regulation of membrane protein ectodomain proteolysis	PMID:18373975
GO:0005615	extracellular space	PMID:12594207
GO:0030730	sequestering of triglyceride	PMID:19032770
GO:0070265	necrotic cell death	PMID:11101870
GO:0032715	negative regulation of interleukin-6 production	PMID:10443688
GO:0032800	receptor biosynthetic process	PMID:10443688
GO:0051092	positive regulation of NF-kappaB transcription factor activity	PMID:10383454
GO:0043193	positive regulation of gene-specific transcription	PMID:17350185
GO:0050796	regulation of insulin secretion	PMID:8383325
GO:0033209	tumor necrosis factor-mediated signaling pathway	PMID:10748004
GO:0051092	positive regulation of NF-kappaB transcription factor activity	PMID:16280327
GO:0016564	transcription repressor activity	PMID:17350185
GO:0009615	response to virus	PMID:10490959
GO:0042346	positive regulation of NF-kappaB import into nucleus	PMID:17922812
GO:0045080	positive regulation of chemokine biosynthetic process	PMID:10490959
GO:0050995	negative regulation of lipid catabolic process	PMID:19032770
GO:0002439	chronic inflammatory response to antigenic stimulus	PMID:14512626
GO:0016481	negative regulation of transcription	PMID:16895791
GO:0010843	promoter binding	PMID:17350185
GO:0034116	positive regulation of heterotypic cell-cell adhesion	PMID:10604883
GO:0048661	positive regulation of smooth muscle cell proliferation	PMID:16518841
GO:0002740	negative regulation of cytokine secretion involved in immune response	PMID:10443688
GO:0005615	extracellular space	PMID:18355445
GO:0045429	positive regulation of nitric oxide biosynthetic process	PMID:8383325

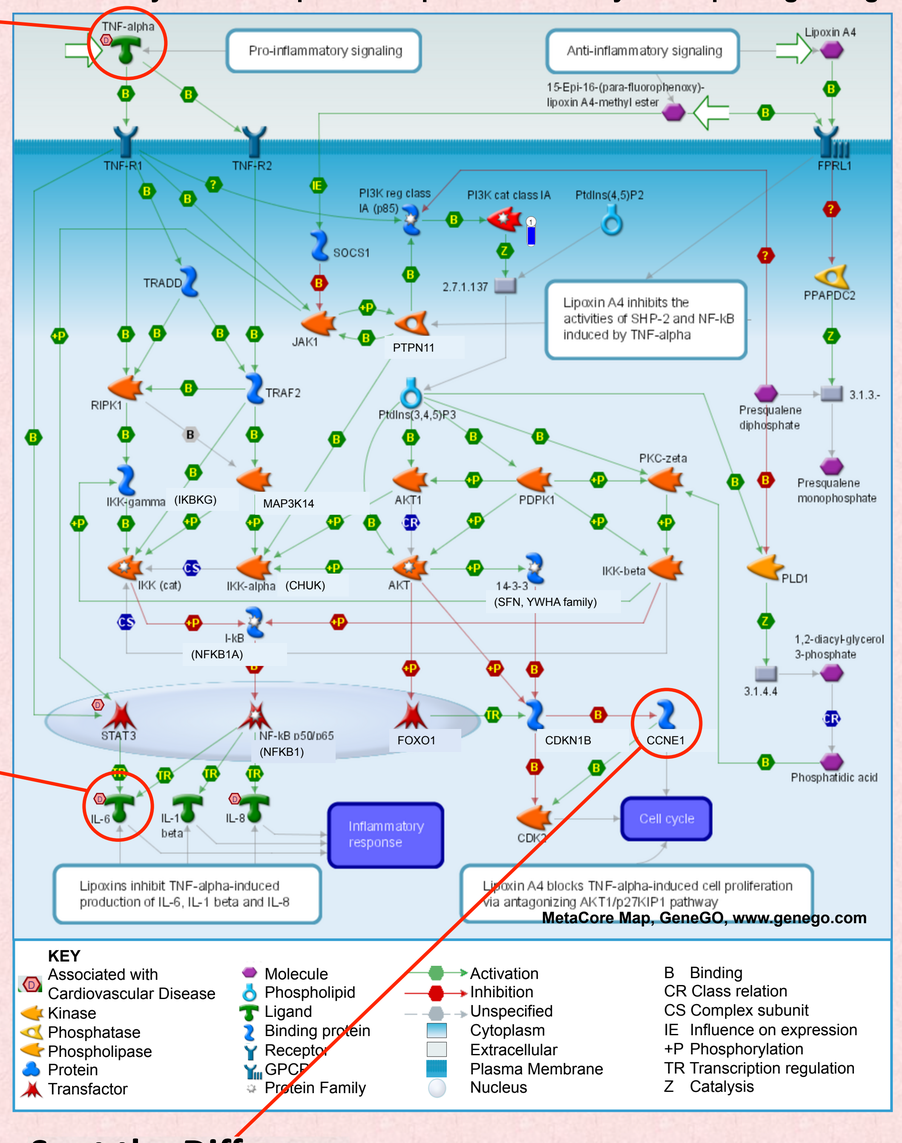
IL6 Homo sapiens P05231

GO Identifier	GO Term Name	Reference
GO:0002675	positive regulation of acute inflammatory response	PMID:2444978
GO:0032722	positive regulation of chemokine production	PMID:10510402
GO:0042981	regulation of apoptosis	PMID:9949178
GO:0042102	positive regulation of T cell proliferation	PMID:3130269
GO:0051384	response to glucocorticoid stimulus	PMID:10443688
GO:0050871	positive regulation of B cell activation	PMID:3491322
GO:0005138	interleukin-6 receptor binding	PMID:10925276
GO:0045079	negative regulation of chemokine biosynthetic process	GO_REF:0000024
GO:0051024	positive regulation of immunoglobulin secretion	PMID:3491322
GO:0006954	inflammatory response	PMID:10443688
GO:0002675	positive regulation of acute inflammatory response	PMID:2444978
GO:0032722	positive regulation of chemokine production	PMID:10510402
GO:0042981	regulation of apoptosis	PMID:9949178
GO:0042102	positive regulation of T cell proliferation	PMID:3130269
GO:0051384	response to glucocorticoid stimulus	PMID:10443688
GO:0046427	positive regulation of JAK-STAT cascade	PMID:12419823
GO:0005138	interleukin-6 receptor binding	PMID:12829785
GO:0048661	positive regulation of smooth muscle cell proliferation	PMID:10510402
GO:0006953	acute-phase response	PMID:12832423
GO:0051091	positive regulation of transcription factor activity	PMID:7749983
GO:0005138	interleukin-6 receptor binding	PMID:12829785
GO:0045669	positive regulation of osteoblast differentiation	PMID:12372336
GO:0050829	defense response to Gram-negative bacterium	PMID:16034137
GO:0005615	extracellular space	PMID:16034137
GO:0005896	interleukin-6 receptor complex	PMID:12829785
GO:0008083	growth factor activity	PMID:2261637
GO:0005138	interleukin-6 receptor binding	PMID:12829785
GO:0045669	positive regulation of osteoblast differentiation	PMID:12372336
GO:0050829	defense response to Gram-negative bacterium	PMID:16034137
GO:0005615	extracellular space	PMID:16034137
GO:0005896	interleukin-6 receptor complex	PMID:12829785
GO:0008083	growth factor activity	PMID:2261637
GO:0008083	growth factor activity	PMID:12643274
GO:0001781	neutrophil apoptosis	PMID:7595060

CCNE1 Homo sapiens P24864

GO Identifier	GO Term Name	Reference
GO:0033197	response to vitamin E	Compara
GO:0050681	androgen receptor binding	PMID:15572661
GO:0031670	cellular response to nutrient	Compara
GO:0032570	response to progesterone stimulus	Compara
GO:0010243	response to organic nitrogen	Compara
GO:0001889	liver development	Compara
GO:0005813	centrosome	Compara
GO:0045597	positive regulation of cell differentiation	Compara
GO:0051301	cell division	Swiss-Prot Keywords2GO
GO:0005634	nucleus	Swiss-Prot Keywords2GO
GO:0014074	response to purine	Compara
GO:0034097	response to cytokine stimulus	Compara
GO:0005515	protein binding	PMID:17525332
GO:0005634	nucleus	PMID:16109376
GO:0005634	nucleus	InterPro2GO
GO:0001547	antral ovarian follicle growth	Compara
GO:0005515	protein binding	PMID:9840943
GO:0005634	nucleus	Subcellular Location2GO
GO:0032403	protein complex binding	Compara
GO:0005829	cytosol	PMID:7799941
GO:0014070	response to organic cyclic substance	Compara
GO:0005515	protein binding	PMID:12607005
GO:0031100	organ regeneration	Compara
GO:0005515	protein binding	PMID:8756624
GO:0045471	response to ethanol	Compara
GO:0005634	nucleus	GO_REF:0000029
GO:0042493	response to drug	Compara
GO:0005634	nucleus	PMID:18029348
GO:0048545	response to steroid hormone stimulus	Compara
GO:0051597	response to methylmercury	Compara
GO:0005737	cytoplasm	Compara
GO:0051412	response to corticosterone stimulus	Compara
GO:0030521	androgen receptor signaling pathway	PMID:15572661
GO:0000082	G1/S transition of mitotic cell cycle	PMID:8207080

Inhibitory action of lipoxins on pro-inflammatory TNF-alpha signalling



Gene Ontology provides a systematic language for the description of gene product attributes in three key domains

Biological Process | Molecular Function | Cellular Component

Annotation

GO terms are associated with gene products (proteins)

Distribution of Data

GO annotations are available through major biological databases and numerous high-throughput analysis GO tools

Large number of uses

- Biomarker discovery
- Enhancing annotation of any genome
- Validation of cell separation methodologies
- Identification of disease-associated processes
- Quick access to information about individual proteins
- Validation of automated ways of deriving gene information
- Drug therapies based on process variations between individuals
- Identification of predominant activities within a specific group of proteins
- Identification of common pathways targeted by different pathogens, proteins etc

Current uses of GO

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[1]. Furthermore, GO data is often used to support hypotheses to explain proteome-wide alterations in response to certain diseases, such as cardiac hypertrophy[2], or stress states, such as hypoxia[3]. 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 disease biomarkers[4,5]. GO data also provides an indispensable resource to indicate the success of subcellular enrichment strategies or large scale confocal microscopy analyses[6,7].

Spot the Difference

Completing the annotation of every gene product, using GO is a substantial undertaking, especially for highly investigated genes. Consequently, at present, there is a wide variation between the quality and quantity of annotations associated with different proteins.

QuickGO (www.ebi.ac.uk/QuickGO) views of the GO terms associated with TNF-alpha and IL-6 (left) illustrate the how detailed GO annotations can be. The number of GO annotations associated with each protein is not simply a reflection of the current knowledge about each protein. Thousands of publications describe TNF-alpha and IL-6 and yet there are over twice as many unique GO terms associated with TNF-alpha (111) as there are with IL-6 (55). This variation in the number of GO terms associated with human proteins is due to the time constraints facing GO curators. Providing comprehensive annotation of all proteins associated with cardiovascular processes is a key goal of the UCL-based GO annotation team.

References

1. Pasini, E.M., Kirkegaard, M., Mortensen, P., *et al*. In-depth anyalysis of the membrane and cytosolic proteome of red blood cells. Blood, 2006, 108, 791-801.
2. Pan, Y., Kislinger, T., Gramolini, A. O., *et al*. Identification of biochemical adaptations in hyper- or hypocontractile hearts from phospholamban mutant mice by expression proteomics, Proc Natl Acad Sci U S A, 2004, 101: 2241-2246.
3. Boraldi, F., Annovi, G., Carraro, F., *et al*. Hypoxia influences the cellular cross-talk of human dermal fibroblasts. A proteomic approach, Biochim Biophys Acta, 2007, 1774: 1402-1413.
4. Shi, M., Jin, J., Wang, Y., *et al*. Mortalin: a protein associated with progression of Parkinson disease?, J Neuropathol Exp Neurol, 2008, 67: 117-124.
5. Perco, P., Wilflingseder, J., Bernthaler, A., *et al*. Biomarker candidates for cardiovascular disease and bone metabolism disorders in chronic kidney disease: A systems biology perspective, J Cell Mol Med, 2008.
6. Kislinger, T., Rahman, K., Radulovic, D., *et al*. PRISM, a generic large scale proteomic investigation strategy for mammals, Mol Cell Proteomics, 2003, 2: 96-106.
7. Barbe, L., Lundberg, E., Oksvold, P., *et al*. Toward a confocal subcellular atlas of the human proteome, Mol Cell Proteomics, 2008, 7: 499-508.

Grant: SP/07/007/23671

Annotating cardiovascular genes

Annotation@EBI

BEATING HEART DISEASE TOGETHER

www.cardiovasculargeneontology.com