Capture of cardiovascular knowledge using Gene Ontology

Shirin C. C. Saverimuttu, Rachael P. Huntley, Saadullah Ahmad, Barbara Kramarz, Ruth C. Lovering

Functional Gene Annotation, Preclinical and Fundamental Science, Institute of Cardiovascular Science

r.lovering@ucl.ac.uk @UCLgene www.ucl.ac.uk/functional-gene-annotation

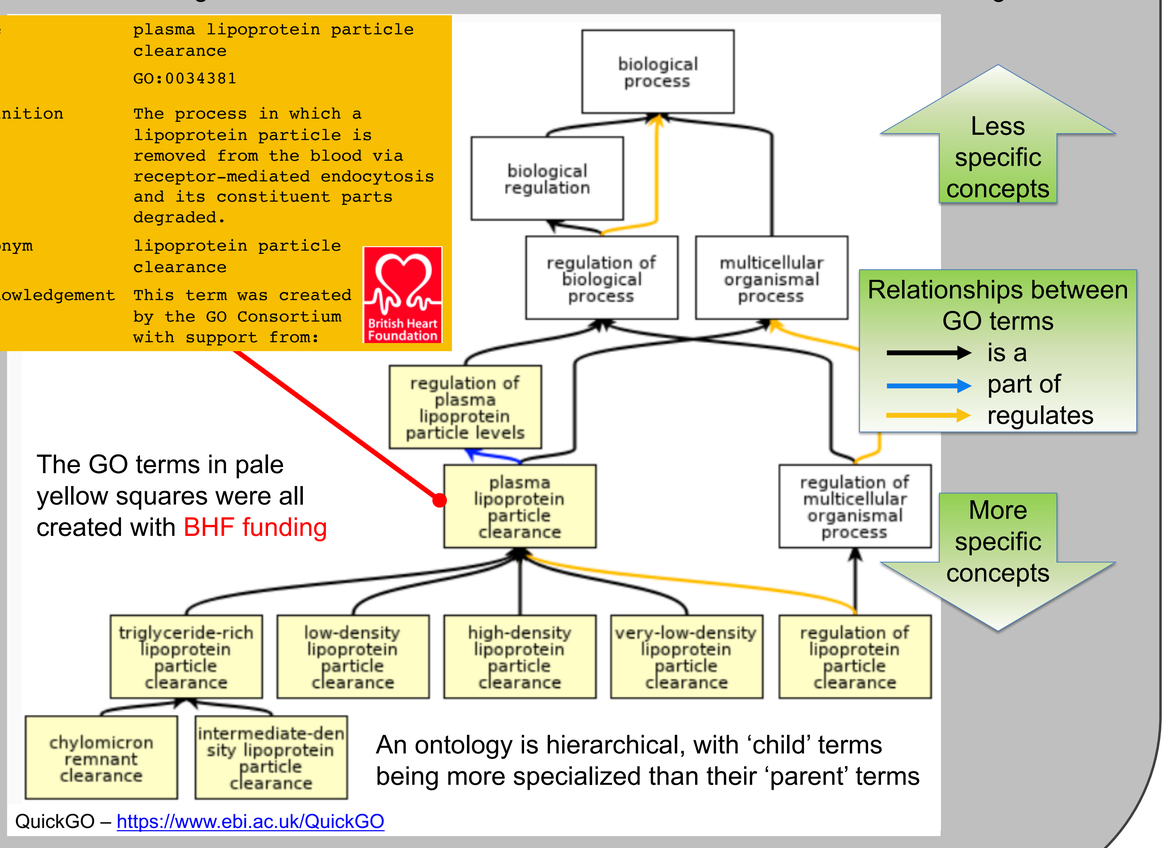
What is Gene Ontology?

- The Gene Ontology (GO) Consortium aims to provide a comprehensive, computational model of biological systems, ranging from the molecular to the organism level
- The GO knowledgebase is the world's largest source of functional information about gene products, such as proteins or non-coding RNAs, which are all referred to as 'genes' throughout this poster
- The biological knowledge is captured in both human-readable and machine-readable formats
- The freely-available descriptions of genes provided by GO form a foundation for computational analysis of large-scale cellular and genetic biomedical research data
- The association of a descriptive GO term with a gene results in a gene 'annotation'
- The GO terms are standardised 'phrases' that describe either the function of a gene product, its biological role or its location in the cell

What is an ontology?

- There are currently 45,000 GO terms describing a wide range of biological concepts to differing levels of specificity.
- The terms are organised as a unidirectional network, each term is a node and each edge a relation

name	plasma lipoprotein particle clearance
ID	GO:0034381
definition	The process in which a lipoprotein particle is removed from the blood via receptor-mediated endocytosis and its constituent parts degraded.
synonym	lipoprotein particle clearance
Acknowledgement	This term was created by the GO Consortium with support from: British Heart Foundation



The GO terms in pale yellow squares were all created with BHF funding

An ontology is hierarchical, with 'child' terms being more specialized than their 'parent' terms

QuickGO – https://www.ebi.ac.uk/QuickGO

What is an annotation?

- The association of a GO term with a gene record is known as a GO annotation
- The BHF-contributed 'heart valve development' term, or one of its child terms, has been associated with 68 human genes. Almost all of these annotations were created by BHF-funded biocurators
- The majority of these annotations use more specific 'heart valve development' terms, such as 'aortic valve morphogenesis' or 'mitral valve formation'

GO:0003170 (P) JSON

heart valve development

Biological Process

Definition (GO:0003170 GONUTS page)

The progression of a heart valve over time, from its formation to the mature structure. A heart valve is a structure that restricts the flow of blood to different regions of the heart and forms from an endocardial cushion.

Acknowledgement

This term was created by the GO Consortium with support from:

British Heart Foundation

GO terms are associated with a gene in a GO annotation

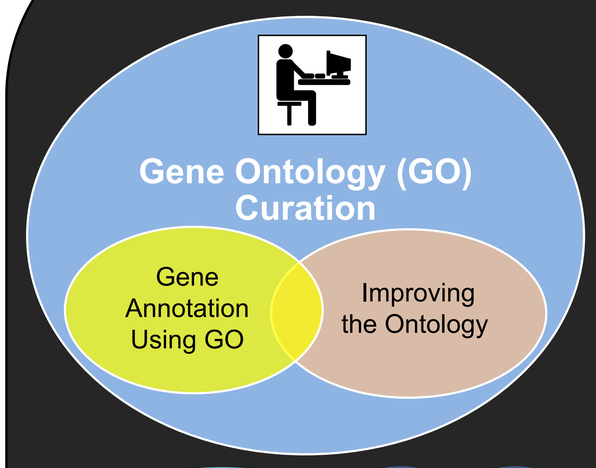
Gene	GO Term Name	Evidence	Reference	Assigned By
GATA4	aortic valve morphogenesis	IMP	PMID:29325903	BHF-UCL
GJA5	mitral valve development	IMP	PMID:20804746	BHF-UCL
GJA5	pulmonary valve morphogenesis	IMP	PMID:20804746	BHF-UCL
JAG1	pulmonary valve morphogenesis	IMP	PMID:20437614	BHF-UCL
NOTCH1	aortic valve morphogenesis	IMP	PMID:18593716	BHF-UCL
NOTCH1	mitral valve formation	IMP	PMID:16025100	BHF-UCL
NOTCH1	pulmonary valve morphogenesis	IMP	PMID:16025100	BHF-UCL
TGFB1	aortic valve morphogenesis	IMP	PMID:26634652	BHF-UCL
TWIST1	cell proliferation involved in heart valve development	IMP	PMID:20804746	BHF-UCL

The evidence code give an indication of the underlying assay supporting the annotation **IMP** = inferred from mutant phenotype

Each annotation is attached to a reference for traceability

- All of the 'heart valve development' child terms were created as part of a BHF-funded project to expand the ontology describing heart development from 12 terms to over 200

How are GO annotations created at UCL?



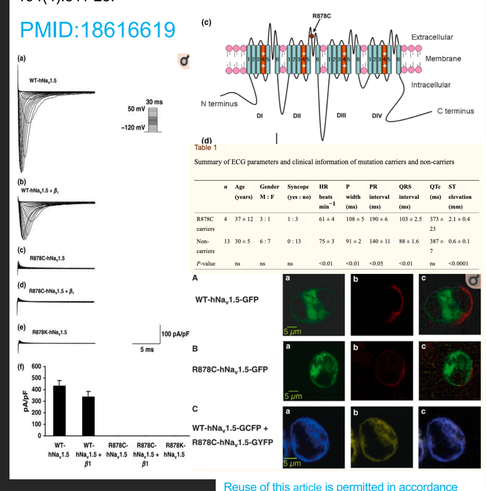
GO biocurators search the literature for experimental data and capture information about biological roles using the universal GO terms. The resulting GO annotations are regularly submitted to GO browsers (QuickGO and AmiGO 2), and exported to other biological databases, such as UniProt, Ensembl and NCBI Gene, as well as to Wikipedia.

GO term F: Function; P: Process; C: Cell location | Evidence code | Source of evidence | Contributor

Symbol	Qualifier	GO Term	Evidence	Reference	Assigned By
SCN5A	enables	GO:0086063 (F) voltage-gated sodium channel activity involved in SA node cell action potential	ECO:0000315 IMP	PMID:18616619	BHF-UCL
SCN5A	involved_in	GO:0086046 (P) membrane depolarization during SA node cell action potential	ECO:0000315 IMP	PMID:18616619	BHF-UCL
SCN5A	enables	GO:0086060 (F) voltage-gated sodium channel activity involved in AV node cell action potential	ECO:0000315 IMP	PMID:18616619	BHF-UCL
SCN5A	involved_in	GO:0086045 (P) membrane depolarization during AV node cell action potential	ECO:0000315 IMP	PMID:18616619	BHF-UCL

Zhang Y, *et al.* Correlations between clinical and physiological consequences of the novel mutation R878C in a highly conserved pore residue in the cardiac Na+ channel. *Acta Physiol (Oxf).* 2008 194(4):311-23.

PMID:18616619



Reuse of this article is permitted in accordance with the Creative Commons License, Attribution 2.5, which does not permit commercial exploitation

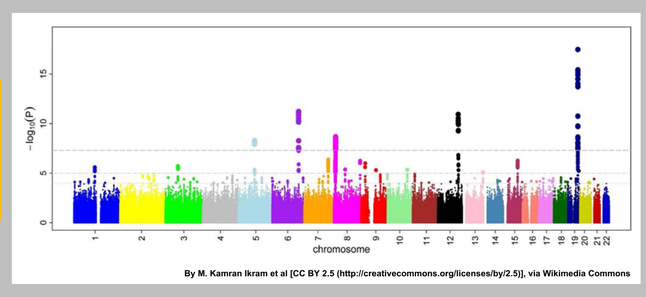
A subset of the 8 human protein annotations created following the curation of experimental data presented in Zhang *et al.* 2008. All of the annotations above include GO terms from the 'cardiac conduction' branch of the ontology. Adapted from in the EBI GO browser QuickGO (www.ebi.ac.uk/QuickGO).

GO annotations are used to interpret 'big data'

- GO annotations group genes based on their function, biological role or cellular location
- GO annotations are incorporated into > 50 functional analysis tools
- Researchers around the world use GO annotations to analyse high-throughput datasets

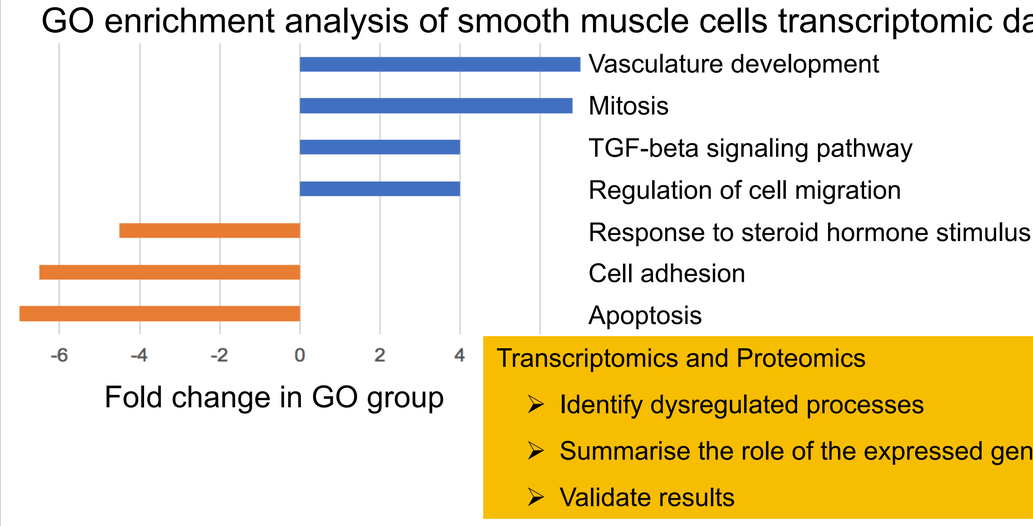
Genome-wide association studies

- Inform choice of candidate genes
- Detect new risk genes



By M. Kamran Ikram et al [CC BY 2.5 (http://creativecommons.org/licenses/by/2.5)], via Wikimedia Commons

GO enrichment analysis of smooth muscle cells transcriptomic data



Transcriptomics and Proteomics

- Identify dysregulated processes
- Summarise the role of the expressed genes
- Validate results

- Smooth muscle cell migration is associated with coronary artery disease
- 500 significantly dysregulated genes in human coronary artery smooth muscle cells following *TCF21* overexpression

Kim JB, *et al.* (2017) *PLOS Genetics* 13(5): e1006750

Further reading

- Khodiyar VK, Hill DP, Howe D, Berardini TZ, Tweedie S, Talmud PJ, Breckenridge R, Bhattarcharya S, Riley P, Scambler P, Lovering RC. **The Representation of Heart Development in the Gene Ontology.** *Dev Biol.* 2011 354(1):9-17. PMID: 21419760
- Lovering RC, Roncaglia P, Howe DG, Laulederkind SJF, Khodiyar VK, Berardini TZ, Tweedie S, Foulger RE, Osumi-Sutherland D, Campbell NH, Huntley RP, Talmud PJ, Blake JA, Breckenridge R, Riley PR, Lambiase PD, Elliott PM, Clapp L, Tinker A, Hill DP. **Improving Interpretation of Cardiac Phenotypes and Enhancing Discovery With Expanded Knowledge in the Gene Ontology.** *Circ Genom Precis Med*. 2018 11(2):e001813. PMID: 29440116
- The Gene Ontology Consortium. **The Gene Ontology Resource: 20 years and still GOing strong.** *Nucleic Acids Res*. 2019 47(D1):D330-D338. PMID: 30395331

Email us for more information or to receive our quarterly newsletter, or follow us on Twitter:

r.lovering@ucl.ac.uk

@UCLgene

British Heart Foundation

Alzheimer's Research UK

The Functional Gene Annotation team is a member of the GO Consortium and supported by Alzheimer's Research UK (grant ARUK-NAS2017A-1) and the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

EMBL-EBI UniProt RNAcentral GENEONTOLOGY Unifying Biology