Connecting the Gene Ontology (GO) to Parkinson's disease

R. E. Foulger*[4], B. Kramarz[1], S. Saverimuttu[1], P. Denny*[5], M. J. Martin[3], J. Hardy[2] and R. C. Lovering[1]

1. Functional Gene Annotation, Institute of Cardiovascular Science, University College London (UCL), Rayne Building, 5 University Street, London
2. Department of Neurodegerative Diseases, Russell Square House, 10-12 Russell Square, London
3. European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton

*Previous Functional Gene Annotation biocurators, currently: 4. Genomics England, Queen Mary University of London; 5. HUGO Gene Nomenclature Committee, EMBL-EBI

What is Gene Ontology?

- The Gene Ontology (GO) Consortium is a collaborative effort to provide freely-available, consistent descriptions of gene products, such as proteins and non-coding RNAs, across all species
- The 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
- 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 analyses of large-scale cellular and genetic biomedical research data
- The association of a descriptive GO term with a gene results in a GO '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

The UCL Parkinson's GO annotation project

- The aim of the project was to provide high-quality GO annotations describing proteins and biological processes relevant to Parkinson's disease
- This was the first annotation effort to focus on Parkinson's and we established collaborations with local and international neurological researchers to guide our priorities
- We extracted data from primary papers and reviews to associate GO terms to Parkinson's-relevant proteins
- Our primary focus was to curate human proteins, but we also captured information from model organisms including fly, rat and mouse, when this provided additional relevant information
- We curated over 700 papers, which led to the annotation of 1,500 distinct proteins and the creation of 7,350 GO annotations, of which 4,500 annotations were associated with 800 human proteins

Annotation priorities

Products of genes associated with Parkinson's, including causative genes and risk genes

Included: 48 high priority genes identified through GWAS and genetic analyses

Cellular pathways known to be dysregulated in Parkinson's

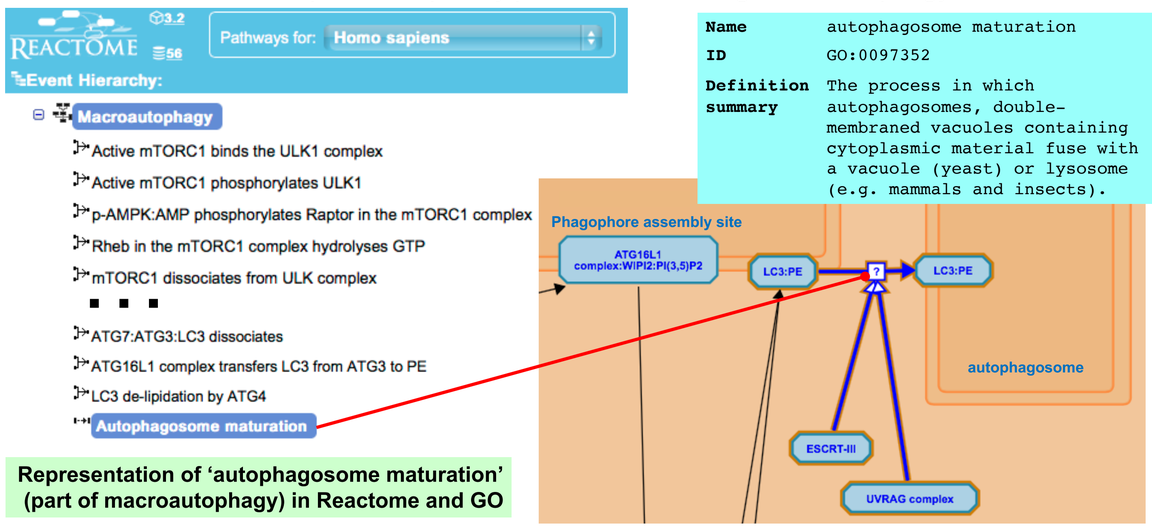
Included: **mitophagy**, **autophagy**, oxidative stress response, unfolded protein response, regulation of neuron death, **neuron differentiation**, synaptic vesicle transport

Parkinson's UK-funded research

Within each topic, and for each priority protein, we prioritised publications funded by Parkinson's UK

Aligning with Reactome

- Reactome (http://www.reactome.org) is a peer-reviewed pathway database and curates many cellular events that overlap with our priority processes, including macroautophagy:



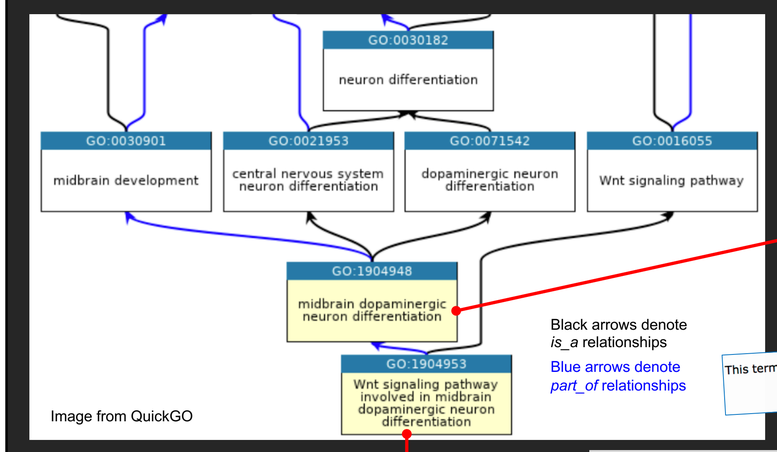
Name	autophagosome maturation
ID	GO:0097352
Definition summary	The process in which autophagosomes, double-membraned vacuoles containing cytoplasmic material fuse with a vacuole (yeast) or lysosome (e.g. mammals and insects).

Representation of 'autophagosome maturation' (part of macroautophagy) in Reactome and GO

- All proteins in the Reactome **Macroautophagy** pathway were downloaded and cross-checked against our GO-annotated proteins, and additional GO annotations added where required.
- We suggested new mappings between Reactome and GO, this led to 150 new GO annotations associating the GO term 'Wnt signaling pathway' (or one of it's child terms) to 140 human proteins.

GO annotation and ontologies

- GO contains three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated **biological processes**, **cellular locations** and **molecular functions**
- Originally developed in 1998, the ontologies have grown to include over 44,000 terms describing a wide range of biological concepts to differing levels of specificity
- GO is an essential resource for high-throughput data analysis, facilitating the grouping of genes to common pathways, functions and cellular locations



name	midbrain dopaminergic neuron differentiation
ID	GO:1904948
definition	The process in which a relatively unspecialized cell acquires the specialized features of a midbrain dopaminergic neuron.
synonym	midbrain DA neurogenesis

GO terms are associated with a gene product in a GO annotation

Evidence codes indicate the information supporting the annotation:
IDA = inferred from direct assay
ISS = inferred from sequence similarity
NAS/TAS = author statements

Each annotation is attached to a reference for traceability

Symbol	GO Term Name	Evidence	With	Reference	Assigned By
LRP6	Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation	TAS		PMID:24431302	ParkinsonsUK-UCL
WNT1	canonical Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation	TAS		PMID:22988876	ParkinsonsUK-UCL
WNT2	canonical Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation	IDA		PMID:20018874	ParkinsonsUK-UCL
RYK	Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation	ISS	Q6BC88	GO_REF:0000024	ParkinsonsUK-UCL
WNT3A	Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation	TAS		PMID:22988876	ParkinsonsUK-UCL
WNT5A	planar cell polarity pathway involved in midbrain dopaminergic neuron differentiation	TAS		PMID:22988876	ParkinsonsUK-UCL
FZD1	Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation	NAS		PMID:23461676	ParkinsonsUK-UCL
LMX1A	midbrain dopaminergic neuron differentiation	ISS	Q9JKU8	GO_REF:0000024	ParkinsonsUK-UCL
NR4A2	midbrain dopaminergic neuron differentiation	TAS		PMID:24431302	ParkinsonsUK-UCL

A subset of human proteins annotated with 'midbrain dopaminergic neuron differentiation' and descendants during this project. Adapted from www.ebi.ac.uk/QuickGO.

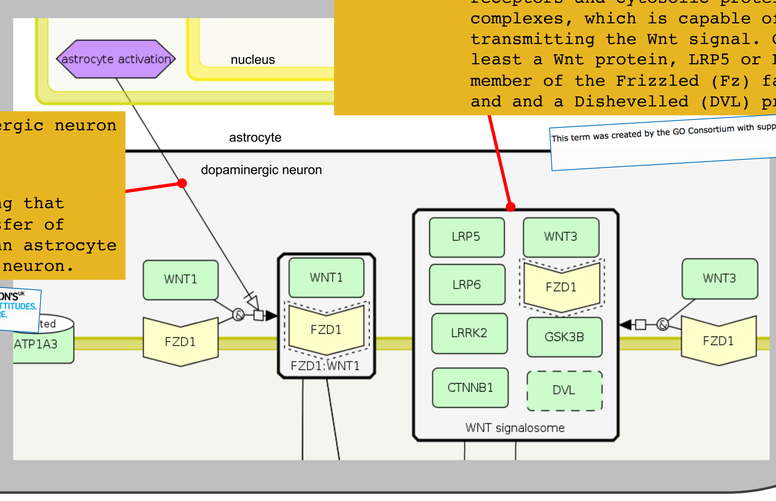
Annotation transferred from mouse Lmx1a

Group providing the GO annotations

Aligning with PD-map

- Parkinson's disease map (PD-map; http://minerva.uni.lu/MapViewer/) offers a pathway-view of Parkinson's disease. Our collaborations with PD-map led to improvements in the representation of cellular processes underlying Parkinson's in both PD-map and the GO annotations.
- We aligned our annotations with PD-map; for both **Wnt signaling** and **autophagy** all proteins and processes represented in PD-map were curated with GO terms, to ensure that we had annotated all key areas.
- This collaboration also led to changes to the Gene Ontology itself as we expand and improve areas of the ontology relevant to Parkinson's disease. We worked with GO editors to create >360 new GO terms relevant to Parkinson's, and many of our new terms were based on PD-map; for Wnt signaling, we created new terms including '**Wnt signalosome**' and '**astrocyte-dopaminergic neuron signaling**':

A subsection of the PD-map Wnt signaling pathway



Name	Wnt signalosome
ID	GO:1990909
Definition	A multiprotein protein complex containing membrane-localized Wnt receptors and cytosolic protein complexes, which is capable of transmitting the Wnt signal. Contains at least a Wnt protein, LRP5 or LRP6, a member of the Frizzled (Fz) family, Axin and and a Dishevelled (DVL) protein.

Name	astrocyte-dopaminergic neuron signaling
ID	GO:0036520
Definition	Cell-cell signaling that mediates the transfer of information from an astrocyte to a dopaminergic neuron.

Further reading

- **Exploring autophagy with Gene Ontology.** Denny P, Feuermann M, Hill DP, Lovering RC, Plun-Favreau H, Roncaglia P. Autophagy. 2018, 17:1-18. *PMID:29455577*
- **Using the Gene Ontology to annotate key players in Parkinson's disease**. Foulger RE, Denny P, Hardy J, Martin MJ, Sawford T, Lovering RC. Neuroinformatics 2016, 14(3):297-304. *PMID 26825309*
- **Computational analysis of the LRRK2 interactome.** Manzoni C, Denny P, Lovering RC, Lewis PA. PeerJ. 2015, 19;3:e778. doi: 10.7717/peerj.778. *PMID 25737818*
- **Ten quick tips for using the Gene Ontology**. Blake J.A. PLoS Comput Biol. 2013, 9(11):e1003343. *PMID 24244145*

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

@UCLgene

r.lovering@ucl.ac.uk

The Parkinson's UK annotation project was funded by a Parkinson's UK grant, G-1307.
UCL Functional Gene Annotation is a member of the GO Consortium
www.ucl.ac.uk/functional-gene-annotation

www.geneontology.org