

Neurological Gene Ontology development and protein annotation facilitate more informative data analysis and interpretation

B. Kramarz¹, P. Roncaglia², R. Bandopadhyay³, D. Brough⁴, M. J. Martin², N. M. Hooper⁴, H. Parkinson², R. C. Lovering¹

¹UCL Institute of Cardiovascular Science, Rayne Building, 5 University Street, London WC1E 6JF, UK; ²European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK; ³UCL Institute of Neurology and Reta Lila Weston Institute of Neurological Studies, 1 Wakefield Street, London WC1N 1PJ, UK; ⁴Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, AV Hill Building, Oxford Road, Manchester, M13 9PT, UK.



Project overview and objectives

The goal of this project, funded by Alzheimer's Research UK (ARUK), is to use Gene Ontology (GO, geneontology.org) to annotate biological roles of proteins interacting with amyloid-beta and/or tau. GO annotations comprise a major resource used for analyses of findings from high-throughput (HTP) experiments, such as transcriptomic, proteomic and genome wide association (GWA) studies. Annotations of proteins implicated in neurological disorders were previously underrepresented in GO, which posed a significant obstacle to interpretation of HTP datasets in the field of dementia research, hence our aim to bridge this gap.

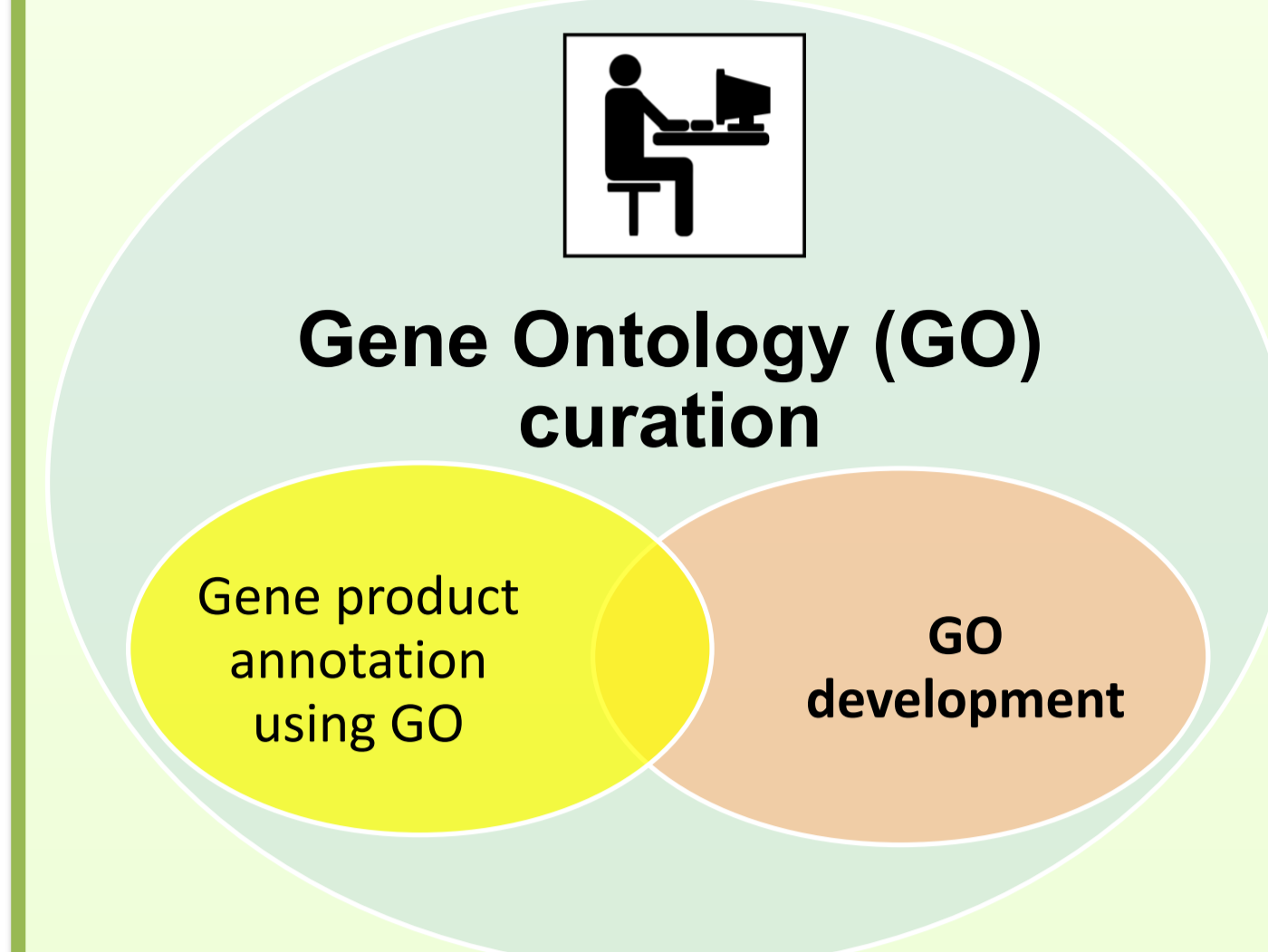
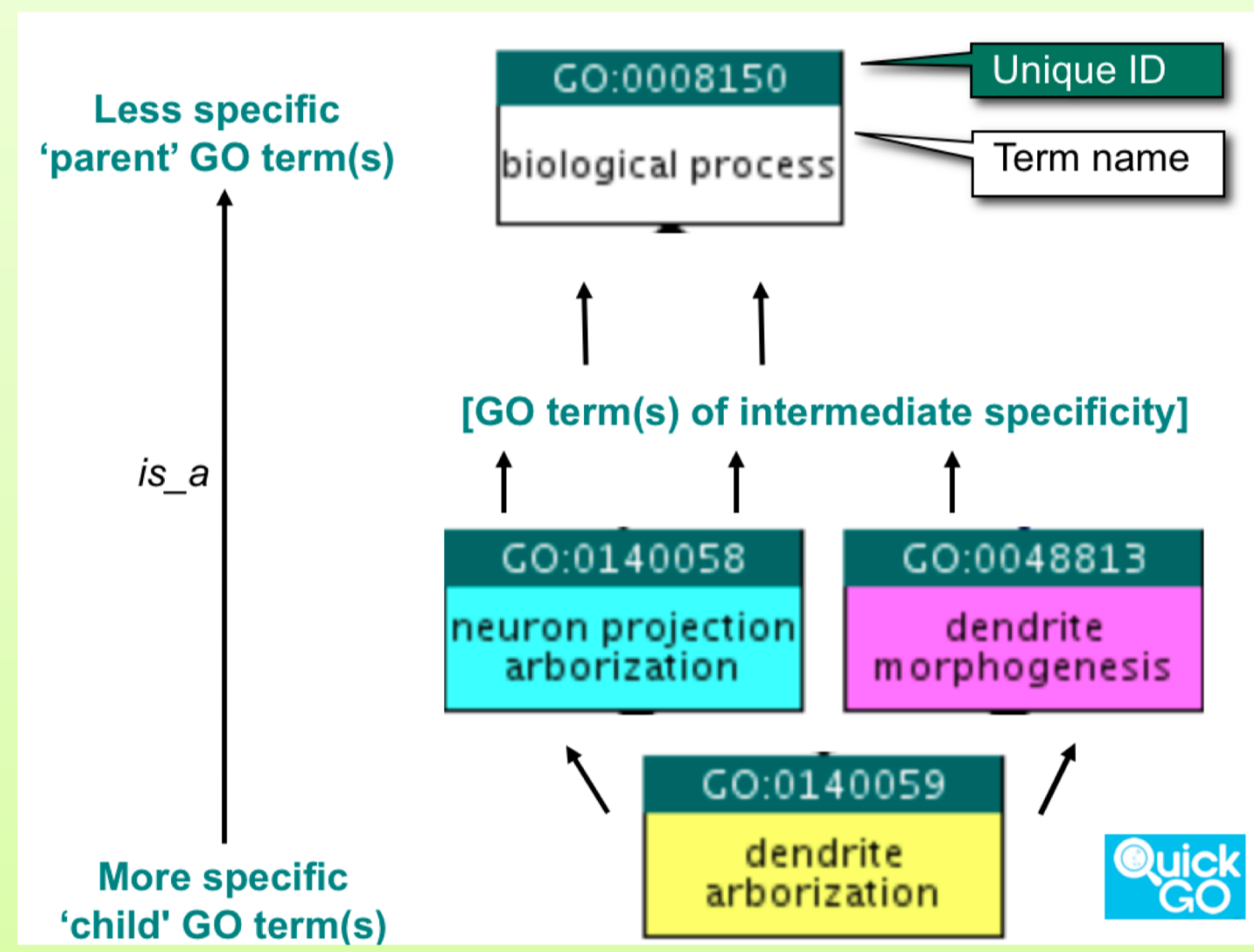
Gene Ontology: a dictionary for biology

What is an ontology?

- An organisation of a knowledge domain in a *written (term name)* and *computable (unique ID)* form
- A controlled vocabulary of terms with definitions (like a dictionary and thesaurus)
- Hierarchical: defined logical relationships between terms

Gene Ontology (GO)

- The gold standard used to describe / tag / label (annotate) normal physiological aspects of gene products, e.g. **proteins**, or non-coding RNAs
- Universal across species and biological fields / standardised
- Dynamic / frequently updated
- Free, publicly available



If an area of biology has been under-represented in GO, then during gene product annotation curators sometimes identify a lack of suitable GO terms. In order to ensure high accuracy and precision of GO annotations, this is addressed by expanding relevant branches of GO and adding more descriptive GO terms. The resulting new GO terms can be accessed in GO browsers (QuickGO and [AmiGO 2](http://AmiGO2)), and they can be used for annotation and analysis by any curator and/or researcher.

GO:0048813

dendrite morphogenesis

Biological Process

Definition (GO:0048813 GONUTS page)
The process in which the anatomical structures of a dendrite are generated and organized. PMID:22683681

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

3,659 annotations

[QuickGO, Accessed 28-Feb-2018]

GO:0140058

neuron projection arborization

Biological Process

Definition (GO:0140058 GONUTS page)
The process in which the anatomical structures of a neuron projection are generated and organized into branches. A neuron projection is any process extending from a neural cell, such as axons or dendrites. PMID:17114044 PMID:23270857 PMID:23764288

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

63 annotations

[QuickGO, Accessed 28-Feb-2018]

GO:0140059

dendrite arborization

Biological Process

Definition (GO:0140059 GONUTS page)
The process in which the anatomical structures of a dendrite are generated and organized into dendritic branches. PMID:23270857

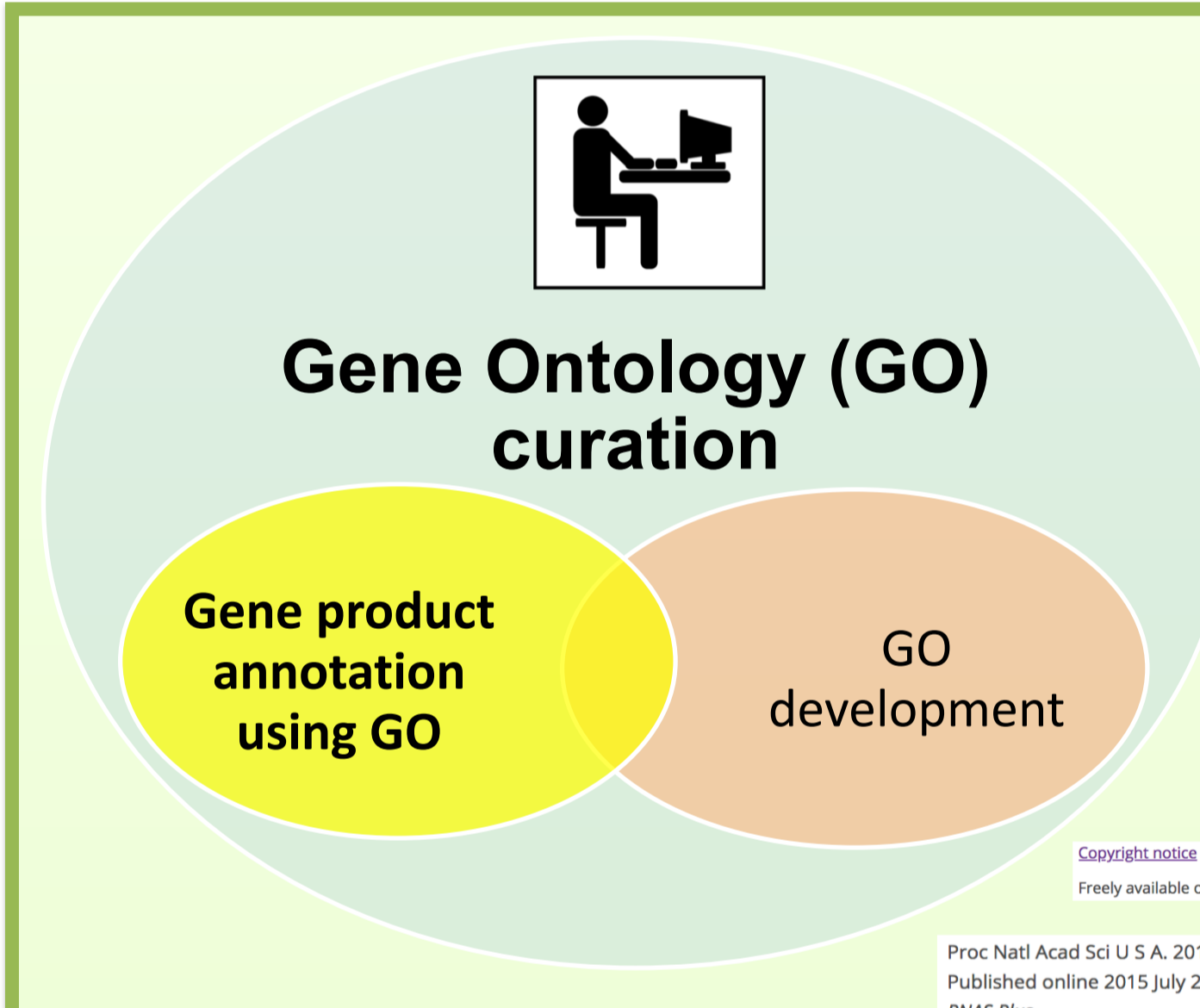
Acknowledgement
This term was created by the GO Consortium with support from:

148 annotations

[QuickGO, Accessed 28-Feb-2018]

GO development

Figure from Romand *et al.*, 2011 (PMID: 21369363); published by Frontiers in Neuroanatomy. Copyright: © 2011 Romand, Wang, Toledo-Rodriguez and Markram. This is an open-access article (...) permits unrestricted use, distribution, and reproduction in any medium (...).



GO curators search 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](http://AmiGO2)), and they are next exported to other biological databases, such as UniProt, Ensembl, or NCBI Gene.

| Gene Product | Symbol | Qualifier | GO Term | Evidence | Reference | Assigned By | Annotation Extension |
|------------------|--------|-------------|--|------------------|---------------|-------------|-------------------------------|
| UniprotKB:P13637 | ATP1A3 | involved_in | GO:0000316 neuron projection maintenance | ECO:0000316 (IG) | PMID:29224839 | ARUK-UCL | occurs_in (ICL:0002008) |
| UniprotKB:P13637 | ATP1A3 | part_of | GO:0030424 axon | ECO:0000314 (ID) | PMID:29224839 | ARUK-UCL | part_of (ICL:0000908) more... |
| UniprotKB:P13637 | ATP1A3 | part_of | GO:0040205 neuronal cell body | ECO:0000314 (ID) | PMID:29224839 | ARUK-UCL | part_of (ICL:0000908) more... |

Gene product annotation using GO

Location: 19,41,966,582-41,997,497 Gene: ATP1A3

Official Symbol: ATP1A3 (accessed by UniProt)

Official Full Name: ATP1A3 (accessed by UniProt)

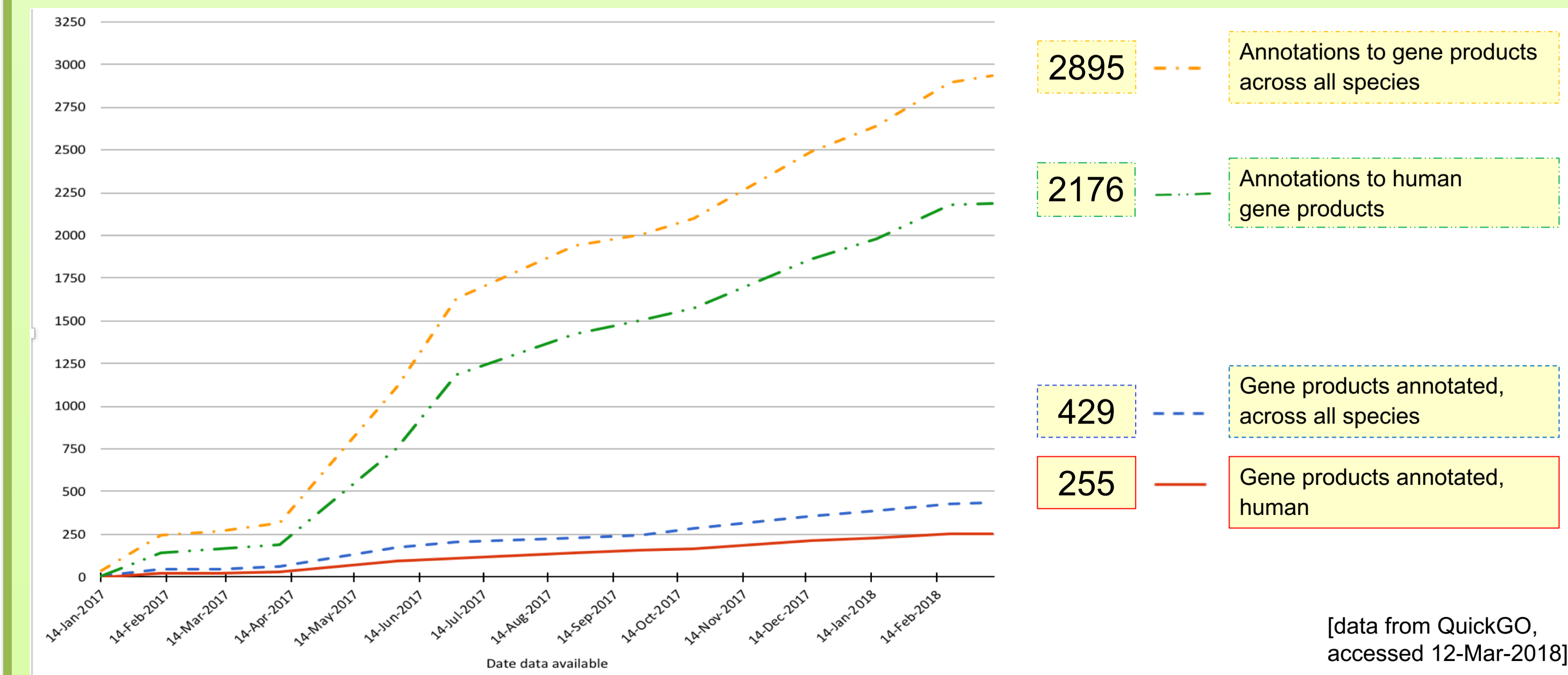
Primary Source: UniProtKB/Swiss-Prot

Secondary Source: UniProtKB/TrEMBL

UniProtKB - P13637 (AT1A3_HUMAN)

Function: **memory** Source: Ensembl; **neuron projection maintenance** Source: ARUK-UCL; **potassium ion import** Source: BHF-UCL; **regulation of cardiac conduction** Source: Reactome

ARUK project – GO annotation progress summary



ARUK project – GO development progress summary

- ❖ 79 New GO terms have been added to the GO database as a result of this project
- ❖ [data from AmiGO2, accessed 12-Mar-2018]

The new ARUK-contributed 'neuron projection arborization' term has been used for annotation 63 times. But only 3 of these annotations have resulted from this ARUK project.

→ The other 60 'neuron projection arborization' annotations have been contributed by other curation groups.

Discussion – how can GO enhance your research?

→ E.g. GO enrichment analysis aiming to assess the biological relevance of disease-associated genes identified through GWAS and SNPs-to-genes analyses.

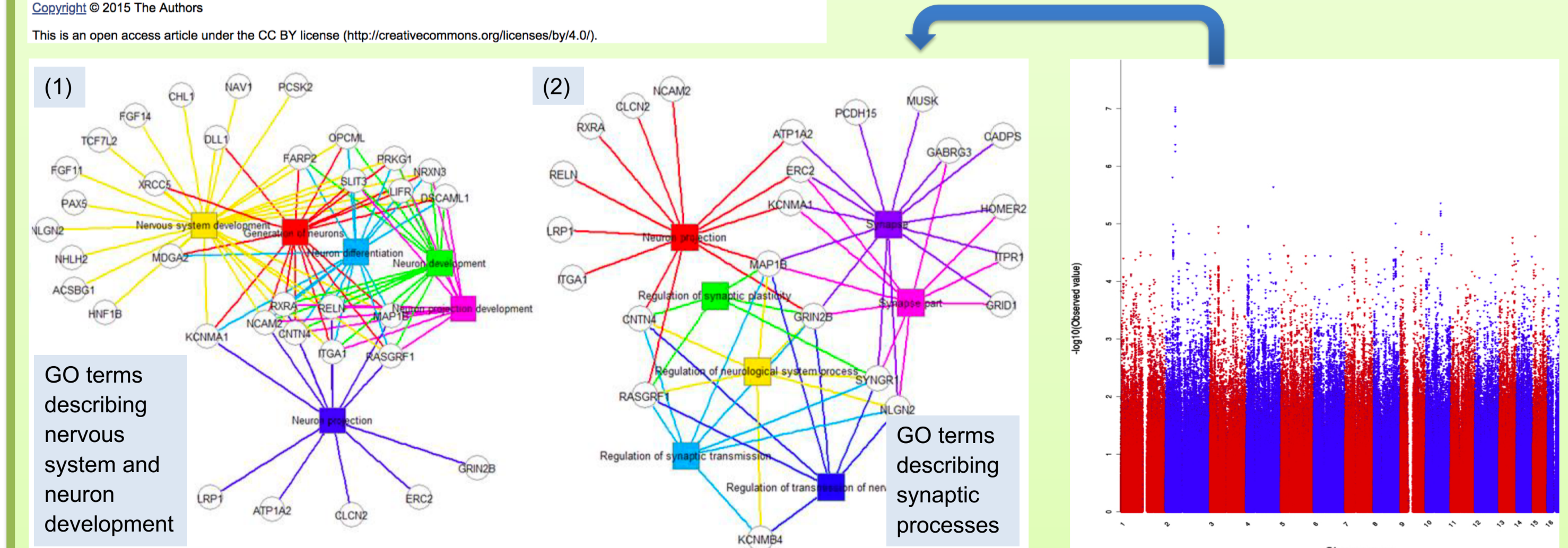
Neurobiol Aging. 2015 Oct; 36(10): 2904.e13–2904.e26. PMID: 2604706156
Published online 2015 Oct. doi: 10.1016/j.neurobiolaging.2015.06.005

A genome-wide screening and SNPs-to-genes approach to identify novel genetic risk factors associated with frontotemporal dementia

Raffaella Ferrari, A. A. V. Mario Grassi, Erika Salvi, Barbara Borroni, Fernando Palluzzi, Daniele Penco, Francesco D'Avila, Alessandro Padovani, Silvana Archetti, Innocenzo Rainone, Elisa Rubino, Lorenzo Pinessi, Luisa Benusa, Giuliano Binetti, Roberta Ghidoni, Daniela Galimberti, Elio Scarpini, Maria Sereneta, Giacomina Rossi, Giorgio Giaccone, Fabrizio Tagliavini, Benedetta Nacmias, Irene Paoletti, Silvia Baonni, Annalia C. Bruni, Raffaella G. Maletta, Livia Bernardi, Alfredo Postiglione, Graziella Milan, Massimo Franceschi, Annibale A. Puccia, Valeria Novelli, Cristina Barlassina, Nicola Giordano, Paolo Manunta, Andrew Singleton, Daniele Cusi, John Hardy, and Paraskei Momen

The GO analysis revealed that the disease-associated genes identified in this study cluster into two functional groups:

- (1) Nervous system and neuron development;
- (2) Synaptic plasticity and transmission.



Summary and future work

Jan - Jun 2017 Amyloid-beta receptors
Jul 2017 – Mar 2018 Tau biology
Apr 2018 – Jun 2020 Neuroinflammation