Gene Ontology: A Resource for Analysis and Interpretation of Alzheimer's Disease Data

Barbara Kramarz • Ruth C. Lovering

Functional Gene Annotation, Preclinical and Fundamental Science, UCL Institute of Cardiovascular Science, University College London, London, UK

Author for correspondence: Ruth Lovering, Functional Gene Annotation, Preclinical and Fundamental Science, UCL Institute of Cardiovascular Science, University College London, London, UK. Email: r.lovering@ucl.ac.uk.

Abstract: Gene Ontology (GO) is a universal resource for analyses and interpretation of high-throughput biological datasets. GO is developed and curated by several different groups, based at scientific institutions around the world, working together under the auspices of the GO Consortium. GO annotations capture biological functional knowledge by associating gene products with GO terms. GO term and gene product records all have computer-readable accession numbers; therefore, these annotations can be easily used for analyses of large datasets while retaining human-readable labels. The UCL Functional Gene Annotation group focuses on GO annotation of human gene products. Our group has led initiatives to systematically annotate proteins and microRNAs across specific biomedical fields, and our current biocuration effort, funded by the Alzheimer's Research UK foundation, is focused on dementia and Alzheimer's disease. Our group has also contributed to the development and revision of the ontology describing neurological domains of biology. Here we present an overview of GO and explain how our work, as well as the work of other members of the GO Consortium, is improving the neurological domains of the GO resource. These biocuration efforts will benefit the dementia and Alzheimer's research community by rendering GO more suitable for analyses of neurological datasets.

Keywords: annotation; biocuration; Gene Ontology; high-throughput analysis; neurobiology

INTRODUCTION

Several genes associated with monogenic Alzheimer's disease (AD) have been identified (1); however, the disease can also be caused by polygenic and environmental risk factors (1, 2). To understand the cellular processes and risk factors associated with AD, numerous transcriptomic, proteomic, and genome-wide association (GWA) studies have been conducted (3–5). Researchers are now turning to pathway-based GWA analysis and Next Generation Sequencing (NGS) to identify the genes contributing to the "missing heritability" (6, 7).

The process of finding gene variants that are causative, or modifiers, of disease is often time-consuming. Bioinformatics-based analyses can aid the identification of AD risk variants, based on the variant's association with a gene product implicated in neurobiological processes and pathways impaired in dementia. Such approaches are reliant on bioinformatics resources, including Gene Ontology (GO) (8, 9), KEGG (10), Reactome (11), and molecular interaction databases (12, 13). These resources provide connections between gene products and biological pathways or networks, which are relevant to AD. The end result of these analyses is the identification of both the risk variant and the candidate gene associated with the risk (14, 15). In addition, considerable research is now focused on the selection of biomarkers for AD (16), and the creation of biomarker panels is likely to be more successful if it is known what biological pathways the candidate biomarkers have in common.

GENE ONTOLOGY

The majority of analyses of high-throughput approaches rely on high-quality annotation data (4, 5) because these bridge the gap between data collation and data analysis (4, 17). Gene annotation datasets provide functional knowledge about gene products, such as proteins or microRNAs, in a computationally accessible format, thus these data can be exploited by systems biology investigators. The main resources used to identify significantly enriched pathways in "omics" studies are those provided by GO (8, 9), KEGG (10), Reactome (11), and protein interaction databases (12, 13). GO annotation data are frequently used because it can describe a gene product's role in a process or its location in a cell, even when the basic molecular activity of this gene product is still under investigation (Figure 1) (18). In contrast, Reactome and KEGG provide very specific information about the molecular function of a gene product within a pathway, with the "reaction" catalyzed or facilitated by each gene product clearly identified within a pathway diagram. Consequently, gene products whose role has not been fully elucidated cannot be included in these resources. Furthermore, although the human and mammalian phenotype ontologies (HPO, MP) (19) are being used to interpret NGS data, understanding how multiple genes contribute to a single

Gene Product	Symbol	GO Term		Evidence		Annotation Extension
UniProtKB:O95782	AP2A1	GO:0050750 (F)	low-density lipoprotein particle receptor binding	ECO:0000353	IPI	activated_by (CHEBI:64646) more...
UniProtKB:P30533	LRPAP1	GO:0150093 (P)	amyloid-beta clearance by transcytosis	ECO:0000315	IMP	has_input (UniProtKB:P05067:PRO_0000000093) more...
UniProtKB:Q00610	CLTC	GO:0072583 (P)	clathrin-dependent endocytosis	ECO:0000315	IMP	has_input (UniProtKB:Q07954) more...
UniProtKB:Q00610	CLTC	GO:0050750 (F)	low-density lipoprotein particle receptor binding	ECO:0000353	IPI	activated_by (CHEBI:64646) more...
UniProtKB:Q00610	CLTC	GO:0150093 (P)	amyloid-beta clearance by transcytosis	ECO:0000315	IMP	has_input (UniProtKB:P05067:PRO_0000000093) more...
UniProtKB:Q07954	LRP1	GO:0006898 (P)	receptor-mediated endocytosis	ECO:0000315	IMP	has_input (UniProtKB:P05067:PRO_0000000093) more...
UniProtKB:Q13492	PICALM	GO:0072583 (P)	clathrin-dependent endocytosis	ECO:0000315	IMP	has_input (UniProtKB:Q07954) more...
UniProtKB:Q13492	PICALM	GO:0050750 (F)	low-density lipoprotein particle receptor binding	ECO:0000353	IPI	activated_by (CHEBI:64646) more...
UniProtKB:Q15109	AGER	GO:0016324 (C)	apical plasma membrane	ECO:0000314	IDA	part_of (CL:0002144) more...

Figure 1 A selection of Gene Ontology annotations. This list of Gene Ontology annotations was downloaded from the QuickGO browser (37). All of these annotations, based on the experimental data presented by Zhao et al. (20), were created by the UCL Functional Gene Annotation group. The annotations were filtered by 'PMID:26005850'. The columns, in order from left to right, are as follows: Symbol, HGNC-approved gene symbol; GO term, GO term identifier and name; Evidence, one of the many Evidence and Conclusion Ontology (ECO) codes (38) associated with each GO annotation to indicate the type of experiments that support the annotation (IDA, Inferred from Direct Assay; IMP, Inferred from Mutant Phenotype; IPI, Inferred from Physical Interaction); Annotation Extension, additional information about the annotations, for example, the location of the function (occurs_in CL:0002144, capillary endothelial cell), or the entity that activates the function (activated_by CHEBI:64646, amyloid-beta polypeptide 40).

disease or phenotype will require resources, such as GO, that describe the cellular roles of these genes.

The GO resource (8, 9) is maintained, curated, and made available through the concerted efforts of the GO Consortium, whose aim is to provide both an ontology of terms and gene product annotations. Consequently, the GO Consortium includes skilled biocuration scientists, ontology editors, and software engineers. The ontology enables the description of attributes of gene products, including proteins, macromolecular complexes, and noncoding RNAs, in three key domains: *molecular function*, *biological process*, and *cellular component*. Fully defined computer-readable GO terms are used by the GO Consortium annotation groups, including our Functional Gene Annotation group at UCL, to create links (annotations) between GO terms and gene products across many species, based on published scientific findings, providing a computable and traceable summary of individual experiments. GO terms are used to describe gene products by their molecular functions (e.g., *scavenger receptor activity*), the biological processes they contribute toward (e.g., *microtubule cytoskeleton organization*), and their subcellular locations (e.g., *extracellular region*). For instance, GO curators have contributed 46 GO annotations based on experimental evidence presented by Zhao et al. (20), of which a selection is presented in Figure 1.

The gene product annotations contributed by GO biocurators are regularly submitted to the GO knowledgebase, where the most current and complete collection of GO terms and annotations is publicly available to all users (9).

Providers of bioinformatics tools, such as g:Profiler (21), Cytoscape (22), or DAVID (23), import GO data into their tools for use in enrichment analyses of large datasets. Therefore, the association of GO terms with gene product records (to create annotations) and the use of GO annotation data in analysis tools together enable groups of similarly annotated gene products, within an "omics" dataset, to be identified as significantly enriched (18, 24, 25). Thus, dysregulated pathways, functions, and macromolecular complexes can be identified within high-throughput datasets. However, GO annotation is a continuously ongoing initiative with certain biological aspects annotated more thoroughly than others. Insufficient annotation of key biological processes and pathways relevant to dementia can hinder the interpretation of outcomes from GWA studies, microarray, and proteomic approaches to dissect AD and other AD-relevant diseases (26). Consequently, these analyses may identify partial protein networks or only general GO terms as enriched in the dataset, for example *mitochondrion* (27) and *calcium-mediated signaling* (4). Having recognized this deficit, the Functional Gene Annotation group at UCL have, for the last 5 years, focused on the annotation of gene products relevant to Parkinson's and Alzheimer's diseases (26, 28, 29). This has led to substantial improvements in the representation of processes such as mitophagy, amyloid precursor protein processing, oxidative stress, and tau-associated processes.

Improving the GO to represent dementia-relevant processes

The GO is structured as directed graphs, with each GO term having a unique term name, for example, *phosphatidylcholine-sterol O-acyltransferase activity*, *proteasomal protein catabolic process*, or *high-density lipoprotein particle*, and a definition (Figure 2) as well as a computer-readable numerical identifier. In addition, the ontology is a dynamic resource, with the ontology itself continually being expanded and refined to capture current knowledge. Although GO terms exist which describe most gene products' processes, functions, and locations, many of these terms are very general and are not specific enough to fully describe the role of AD-associated gene products. The UCL Functional Gene Annotation group has begun to address this issue through the development of the ontology to provide more specific and descriptive GO terms, by improving the existing term definitions and by revising the existing ontology structure (26, 28, 29). The association of these more specific GO terms prevents the loss of valuable descriptions of gene products, based on experimental information, that would have been unavailable if the more general GO term had been applied. For example, we have improved the ontology domains describing the unfolded protein response (UPR) (28), autophagy (29), and neuron projection development (26). These improvements have led to an expansion of the number of GO terms describing these processes, as well as revision of relationships between terms within the ontology. All of these biological processes have relevance to AD as well as Parkinson's disease and other neurological conditions.

Although GO terms are categorized into three key domains, as introduced above, revisions in one domain are often done in conjunction with another domain describing the same biological niche. For instance, our work on *neuron projection development* (26), a biological process GO term, resulted, first, in
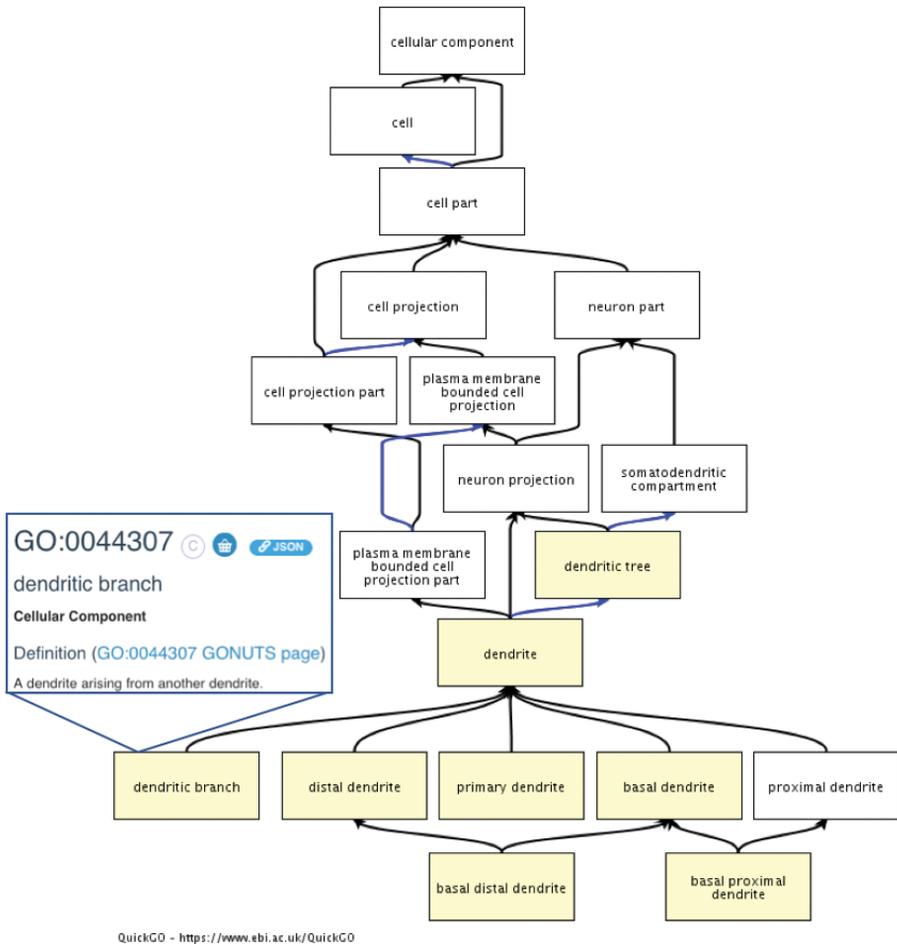
Figure 2 A selection of the dendrite Gene Ontology graph. This figure was generated by the QuickGO browser (37) and shows the *is_a* (black arrows) and *part_of* (blue arrows) hierarchy of just a small number of terms within the *dendrite* branch of the ontology; currently, the dendritic GO domain has 36 terms. The general term *dendrite* is used to group different types of dendrites, for example, *primary dendrite* and *distal dendrite* are both more descriptive child terms of *dendrite*. The definition (as displayed in QuickGO) of one of the GO terms, *dendritic branch*, is also included. The yellow highlighted terms were contributed by the UCL Functional Gene Annotation group.

contribution of new, more descriptive, GO terms, such as *neuron projection arborization*, *dendrite morphogenesis*, or *dendrite arborization*. Yet, simultaneously, we also improved the *dendrite* branch of the cellular component GO aspect, as shown in Figure 2. Similarly, curation of the autophagy (29) processes led to not only generation of highly specific *biological process* GO terms, but also resulted in revisions of related *cellular component* terms, such as *autophagosome*, *amphisome*, or *late endosome*. Thus, enhancing one ontology branch within a specific domain

of GO is often done in conjunction with improvements in other branches and domains, consequently enriching the ontology resource more broadly.

The neuroscience research community will also have benefited from curation work of the SYSCILIA research Consortium, which involved revisions and improvements to cilia-related biology in GO, resulting in contribution of 50 new GO terms (30). Among others, ciliary dysfunction has been shown to affect Sonic hedgehog signaling in the brain, a pathway with demonstrated implications in Alzheimer's (31), Parkinson's (32), and Huntington's (33) diseases. Consequently, revisions and new contributions to the ciliary niche will have improved the representation of cilia biology in GO and, therefore, resulted in more informative analyses of neurological datasets with changes in ciliary proteins.

Another ongoing biocuration initiative with direct relevance to elucidation of Alzheimer's data is the SynGO project and the associated synaptic GO portal (34). SynGO is a collaboration between the Stanley Center for Psychiatric Research at the Broad Institute (Cambridge, MA, USA), the Center for Neurogenomics and Cognitive Research at the Vrije Universiteit (Amsterdam, The Netherlands), and the GO Consortium, thus combining the efforts of experts in synapse biology and GO biocurators to generate the best possible representation of synapse biology in GO.

The UniProt Knowledgebase (EMBL-EBI, Cambridge, UK) has also been improving the representation of Alzheimer's data in their resource, an initiative which includes GO annotation as well as IntAct (13) curation of protein–protein interaction and/or curation of disease variants, as a part of a project funded by the National Institutes of Health (USA). The ultimate goal of this project is to create an online AD portal with thoroughly annotated and easily searchable information on the disease and biological pathways impaired in dementia (35). Importantly, all biocuration scientists, aiming to improve the representation of dementia-relevant biology in GO, work together under the auspices of the GO Consortium, thus ensuring GO annotation consistency and quality.

The creation of GO annotations

There are two major approaches that rely on concerted efforts of skilled biologists and software engineers (36), which result in high-quality GO annotations: manual techniques that depend on the knowledge and expertise of biocuration scientists and computational methods that generate annotations, for instance, based on sequence similarity algorithms. Every annotation is attributed to an identified reference, often a publication identifier, such as PMID, and each annotation must indicate what kind of evidence supports the association between the gene product and the GO term (Figure 1).

The computational annotation approach is a high-throughput and efficient method of associating high-level terms to a large number of gene products across all genomes. These annotations are often assigned, based on specific protein domains with known functions or cellular locations, or based on orthology to a manually curated gene product. However, to provide more specific annotations, GO biocurators read the published scientific literature and use the published data to manually associate highly descriptive GO terms to gene products. Consequently, complete, highly detailed annotation of the processes and networks

that a single gene product is involved in may take a considerable time. Depending on the number of published papers describing the gene product, a curator will annotate an average of 1–3 experimental papers per day.

Furthermore, as there is no limit to the number of GO annotations that can be assigned to a gene product record, it is possible to describe the many different roles that the gene product may have, depending on the cell type it is expressed in, the developmental stage of the organism, and the environmental stimuli the cell is responding to. The UCL Functional Gene Annotation group takes an unusual approach to annotation, in that we usually focus on annotation of a specific process involving a number of gene products, such as amyloid precursor protein processing, rather than working through an unrelated set of gene products. This enables us to develop a better understanding of the biology and apply a consistent annotation approach to all gene products involved in the process, thus providing depth to the annotations. In addition, at UCL we annotate full papers, whereas some groups will curate only the information in a paper that is relevant to a specific prioritized gene. This approach enables us to provide annotations to a large number of relevant gene products, involved in a specific process, which may not be included in the list of annotation priorities. For example, after completing the annotation of 84 proteins and protein complexes, prioritized for annotation as part of the amyloid-beta or tau projects, we had, in total, annotated 526 proteins and complexes (26).

Furthermore, in response to the research community's needs (39), at UCL our annotation procedure involves inclusion of annotation extensions (40) to capture information about the cell and tissue types in which a particular gene product is active, as well as the specific target of a protein or a microRNA. These detailed annotations provide critical knowledge for biomarkers, diagnostics, and drug discovery and will be of considerable value to the research community and allow users of GO to query a variety of data. For example, a GO user could investigate all targets of a particular protein ubiquitin ligase, or, more specifically, search for all proteins involved in catabolic pathways in microglial cells. Unfortunately, although biocurators have been contributing the annotation extension data for over 6 years, there are no tools that are using this data, and only a few browsers display it (9, 41). In the near future, the annotation extension information will be ported to Gene Ontology Causal Activity Modeling (GO-CAM) (42).

Gene products annotated using GO

Historically, GO was used specifically for annotation of proteins. Recently, the GO Consortium has extended the range of gene products that are annotated; rather than only annotating proteins some members of the GO Consortium are now annotating protein complexes (43) and microRNAs (26, 44, 45). To curate these entities, it has been necessary to create new identifiers (43, 46) and develop strict guidelines to ensure that a consistent annotation approach is applied. For example, there are many papers describing the coregulation of a microRNA or a set of microRNAs with the transcription of a panel of mRNAs and implying that these microRNAs therefore regulate the coregulated mRNAs. Such data do not comply with quality standards implemented by the GO Consortium and are not being captured as GO annotations (44). Instead microRNA GO annotations are

contributed based on more precise low-throughput functional experiments, involving microRNA mimics or knockdown, followed by an assessment of the expression of a panel of specific mRNAs. In addition, reporter assay data, confirming a direct interaction between a microRNA and an mRNA, are being captured using specific GO terms (e.g., *mRNA binding involved in posttranscriptional gene silencing*). Furthermore, in these cases the annotation extension will be used to capture the identifier of the targeted mRNA. The resulting interaction data are not only available in the GO annotation files, but also within the EBI-GOA-miRNA dataset from the PSICQUIC web server (45).

The impact of improving the GO resource on data interpretation

By creating an open access dataset of high-quality annotations, which describe the cellular role of those proteins and microRNAs that contribute to pathways dysregulated in AD, the GO provides an invaluable resource for researchers. GO annotations are incorporated into over 50 functional analysis tools, the majority of which are freely available, such as g:Profiler (21), PANTHER (47), Cytoscape (22), and DAVID (23), but others are subscription based, such as Ingenuity Pathway Analysis (QIAGEN Bioinformatics) (48) and MetaCore (Clarivate Analytics) (49). These tools, and many other functional analysis tools, are used by researchers to analyze a variety of high-throughput data, including transcriptomic (4, 50–53), proteomic (5, 54, 55), and GWA (6, 14, 56) data. In addition, existing pipelines ensure that the GO annotations are included in widely used public resources such as UniProt (57), NCBI Gene (58), Ensembl (59), RNAcentral (46), and even Wikipedia. GO annotations associated with individual protein, RNA, or macromolecular complex records are used by researchers to extract a synopsis of the cellular role of a gene product. These gene summaries have many uses in research, for example, they can help guide researchers to the most likely candidate gene associated with a risk locus (14, 18). However, it is the use of GO for the interpretation of data from high-throughput analyses where this resource can be exploited to the full.

The quality of the GO annotations used in the analysis of large biological datasets will determine how informative the outcomes of this analysis will be. Without highly descriptive annotations the analysis can only identify GO functions, processes, or location that are not very specific, such as *site of polarized growth*, *wound healing*, and *cell migration* (54). The identification of more informative enriched terms is dependent not only on the presence of highly descriptive GO terms describing biological knowledge, but also on the association of these terms with a sufficient number of gene products to enable the term to be detected as significantly enriched. A recent meta-analysis of late-onset AD, that included over 94,000 individuals, identified over 100 new risk loci, associated with amyloid-beta and tau processes, as well as immune response pathways and lipid processing (14). This meta-analysis took a wide range of approaches to identify new risk loci, one of which was the use of the pathway analysis software, MAGMA (60), and GO annotation files (36). The GO terms *plasma lipoprotein particle assembly*, *reverse cholesterol transport*, *regulation of amyloid precursor catabolic process*, and *activation of immune response* were identified as processes with relevance to AD. The first three of these GO terms provide a good description of the processes involved, whereas the last term *activation of immune*

response is too general to really give an indication of the mechanism involved. This is likely to reflect the considerable investment in annotation of cardiovascular (18, 61, 62) and nervous system genes (26, 28), and the lack of focused annotation of the immune system. In addition, papers describing the immune system are often highly detailed and more challenging for biocurators without a background in immunology to fully annotate (63). Thus, the annotation of immune-associated pathways does not reflect the volume of literature and knowledge in this domain.

Another study aimed to elucidate protein expression in different brain regions in Alzheimer's cases relative to controls to provide a broader understanding of molecular pathways impaired in dementia (64). In this study, GO analysis was used to identify the *biological processes* that had the largest numbers of differentially expressed proteins associated with them. A wide variety of processes were identified, in this way, including *regulation of apoptosis* associated with the hippocampus and *protein transport* associated with the cerebellum and cingulate gyrus, therefore allowing researchers to identify new routes for potential therapeutic interventions.

GO term enrichment analysis has also been implemented in pilot studies aiming to identify biomarkers associated with dementia, which can be detected using noninvasive methods in easily accessible bodily fluids, such as blood (65) and urine (66). For instance, Chouliaras et al. (65) used GO enrichment together with KEGG pathway analysis to demonstrate the relationships between the identified blood biomarkers with neurological processes and neuronal components. Significantly enriched GO terms included *regulation of amyloid-beta formation* and *amyloid-beta binding*, *main axon*, and *ion channel complex*, whereas KEGG pathways included *glutamatergic synapse* and *Alzheimer's disease*, thus confirming their relevance to cognitive impairments.

Similarly, Watanabe et al. (66) used GO term enrichment and KEGG pathway analyses to delineate the roles of proteins differentially expressed in urine of Alzheimer's patients relative to healthy controls to identify a urine biomarker signature, which could be used for noninvasive diagnostic purposes. *Lipoprotein metabolism*, *heat shock protein 90 signaling pathway*, and *matrix metalloproteinase signaling pathway* as well as *redox regulation by thioredoxin* were among the molecular pathways with the highest enrichment scores, providing evidence for impairment of vascular processes key to the development of dementia (66). In addition, Watanabe et al. (66) also supplemented their functional GO and KEGG analyses with an *interrelation* network analysis to determine, which of the proteins differentially expressed in the Alzheimer's urine samples interact with each other (either directly or via an intermediate). This network analysis of molecular relationships enabled these researchers to further elucidate which GO biological processes and KEGG pathways should be prioritized in future studies and whether they correlate with and confirm other findings. An alternative approach to using GO annotations is to visualize them on an interaction network. This provides the researcher with an overview of the contribution that a network, or part of the network, makes to a particular process or the cellular location of the interacting entities, as shown in the example in Figure 3. Thus, the use of multiple, interoperable, annotation resources provides the opportunity to fully exploit and interrogate individual datasets.
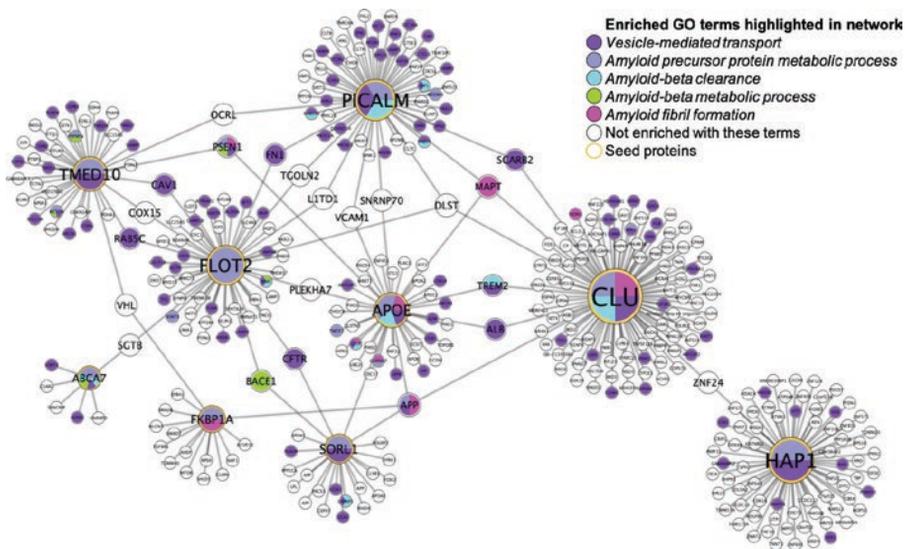
Figure 3 Network of proteins identified in Alzheimer's disease meta-analysis. Nine proteins identified in an Alzheimer's disease meta-analysis (14) due to their association with the GO term "regulation of amyloid precursor protein catabolic process" were used to seed an interaction network using Cytoscape (22) and five files available on the PSICQUIC web server (67) (IntAct, BHF-UCL, UniProt, MINT, and EBI-GOA-non-IntAct). The seed proteins are outlined in yellow. The network was analyzed using Golorize (68), BiNGO (69), and GO ontology and annotation files (36), as described in Denny et al. (29) (downloaded March 29, 2019). The proteins associated with a selection of the enriched GO terms (or one of their child terms, including regulation child terms) are shown in the network.

The above examples demonstrate how continuous, systematic, and consistent improvements to the GO resource, including contribution of new descriptive GO terms, and their association with gene products in the form of GO annotations, impact more informative outcomes of analyses of high-throughput datasets. The functional analyses relying on GO allow researchers to, first, plan and design further studies leading to a better understanding of the molecular mechanisms underlying dementia, and, second, to develop noninvasive diagnostic methods, which collectively will help to improve the management and treatment of AD.

CONCLUSION

The GO resource (8, 9, 70) adds value to published experimental data by creating computer-readable annotations that describe specific functions of a gene product, such as protein, complex, or noncoding RNA, and the biological processes and pathways it contributes to. This benefits all biological areas, including the AD field, on which the UCL Functional Gene Annotation team has recently been focusing their biocuration efforts. Gene annotations using GO terms enable groups of gene products, with similar cellular roles or locations in the cell, to be

easily identified within a dataset, such as a list of differentially expressed genes from AD cases. Thus, dysregulated pathways, functions, and macromolecular complexes can be identified within high-throughput datasets using GO annotation data and functional enrichment tools. GO annotation data are therefore needed for pathway construction, enrichment analyses and interpretation of large-scale datasets (3–5) and to inform biomarker selection decisions (27), and can also be used to identify novel drug targets or novel repurposing of drugs. Furthermore, the AD-focused and comprehensive efforts of the UCL Functional Gene Annotation team have improved and continue to improve the GO resource, enhancing its applicability to this neurobiological research domain and facilitating analyses and interpretation of AD *big data*.

REFERENCES

1. Carmona S, Hardy J, Guerreiro R. The genetic landscape of Alzheimer disease. Handb Clin Neurol. 2018;395–408.
2. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol. 2011 Sep;10(9):819–28.
3. Patel H, Dobson RJB, Newhouse SJ. A meta-analysis of Alzheimer's disease brain transcriptomic data. J Alzheimer's Dis. 2019 Apr 23;68(4):1635–56.
4. Cooper-Knock J, Kirby J, Ferraiuolo L, Heath PR, Rattray M, Shaw PJ. Gene expression profiling in human neurodegenerative disease. Nat Rev Neurol. 2012;8(9):518–30.
5. Kang M-G, Byun K, Kim JH, Park NH, Heinsen H, Ravid R, et al. Proteogenomics of the human hippocampus: The road ahead. Biochim Biophys Acta. 2015 Jul;1854(7):788–97.
6. Guio-Vega GP, Forero DA. Functional genomics of candidate genes derived from genome-wide association studies for five common neurological diseases. Int J Neurosci. 2017 Feb 17;127(2):118–23.
7. Ebbert MTW, Ridge PG, Kauwe JSK. Bridging the gap between statistical and biological epistasis in Alzheimer's disease. Biomed Res Int. 2015;2015:1–7.
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.

9. The Gene Ontology Consortium. The Gene Ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 2019 Jan 8;47(D1):D330–8.

10. Kotera M, Hirakawa M, Tokimatsu T, Goto S, Kanehisa M. The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. Methods Mol Biol. 2012;802:19–39.

11. Haw R, Hermjakob H, D'Eustachio P, Stein L. Reactome pathway analysis to enrich biological discovery in proteomics data sets. Proteomics. 2011;11(18):3598–613.

12. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. Nucleic Acids Res. 2012;40(Database issue):D841–6.

13. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project – IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014;42(Database issue):D358–63.

14. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019 Mar 28;51(3):414–30.

15. Holmans P, Moskvina V, Jones L, Sharma M, Vedernikov A, Buchel F, et al. A pathway-based analysis provides additional support for an immune-related genetic susceptibility to Parkinson's disease. Hum Mol Genet. 2013;22(5):1039–49.

16. Hrubešová K, Fousková M, Habartová L, Fišar Z, Jirák R, Raboch J, et al. Search for biomarkers of Alzheimer's disease: Recent insights, current challenges and future prospects. Clin Biochem. 2019;72:39–51.

17. Lovering RC, Dimmer EC, Talmud PJ. Improvements to cardiovascular Gene Ontology. Atherosclerosis. 2009;205(1):9–14.

18. Lovering RC, Roncaglia P, Howe DG, Laulederkind SJF, Khodiyar VK, Berardini TZ, et al. Improving interpretation of cardiac phenotypes and enhancing discovery with expanded knowledge in the Gene Ontology. Circ Genom Precis Med. 2018;11(2):e001813.

19. Bone WP, Washington NL, Buske OJ, Adams DR, Davis J, Draper D, et al. Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. Genet Med. 2016 Jun 12;18(6):608–17.

20. Zhao Z, Sagare AP, Ma Q, Halliday MR, Kong P, Kisler K, et al. Central role for PICALM in amyloid-β blood-brain barrier transcytosis and clearance. Nat Neurosci. 2015 Jul 25;18(7):978–87.

21. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016;44(W1):W83–9.

22. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics. 2011;27(3):431–2.

23. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

24. Manzoni C, Denny P, Lovering RC, Lewis PA. Computational analysis of the LRRK2 interactome. PeerJ. 2015;3:e778.

25. Evans IM, Kennedy SA, Paliashvili K, Santra T, Yamaji M, Lovering RC, et al. Vascular endothelial growth factor (VEGF) promotes assembly of the p130Cas interactome to drive endothelial chemotactic signaling and angiogenesis. Mol Cell Proteomics. 2017;16(2):168–80.

26. Kramarz B, Roncaglia P, Meldal BHM, Huntley RP, Martin MJ, Orchard S, et al. Improving the Gene Ontology resource to facilitate more informative analysis and interpretation of Alzheimer's disease data. Genes (Basel). 2018;9(12):593.

27. Levine AJ, Miller JA, Shapshak P, Gelman B, Singer EJ, Hinkin CH, et al. Systems analysis of human brain gene expression: mechanisms for HIV-associated neurocognitive impairment and common pathways with Alzheimer's disease. BMC Med Genomics. 2013 Feb 13;6(1):4.

28. Foulger RE, Denny P, Hardy J, Martin MJ, Sawford T, Lovering RC. Using the Gene Ontology to annotate key players in Parkinson's disease. Neuroinformatics. 2016;14(3):297–304.

29. Denny P, Feuermann M, Hill DP, Lovering RC, Plun-Favreau H, Roncaglia P. Exploring autophagy with Gene Ontology. Autophagy. 2018 Mar 04;14(3):419–36.

30. Roncaglia P, van Dam TJP, Christie KR, Nacheva L, Toedt G, Huynen MA, et al. The Gene Ontology of eukaryotic cilia and flagella. Cilia. 2017 Dec 16;6(1):10.

31. Vorobyeva AG, Saunders AJ. Amyloid-β interrupts canonical sonic hedgehog signaling by distorting primary cilia structure. Cilia. 2018 Dec 17;7(1):5.

32. Dhekne HS, Yanatori I, Gomez RC, Tonelli F, Diez F, Schüle B, et al. A pathway for Parkinson's disease LRRK2 kinase to block primary cilia and sonic hedgehog signaling in the brain. Elife. 2018 Nov 6;7.

33. Kaliszewski M, Knott AB, Bossy-Wetzel E. Primary cilia and autophagic dysfunction in Huntington's disease. Cell Death Differ. 2015 Sep 10;22(9):1413–24.

34. Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijsouw T, Coba MP, et al. SynGO: An evidence-based, expert-curated knowledge base for the synapse. Neuron. 2019;103(2):217–234.e4.

35. Lussi YC, Breuza L, Casas CC, Georghiou G, Magrane M, Masson P, et al. UniProtKB and Alzheimer's disease: Linking molecular defects to disease phenotype. 2019 Mar 21 [accessed 2019 Jun 5];8. Available from: https://f1000research.com/posters/8-316.

36. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, et al. The GOA database: Gene Ontology annotation updates for 2015. Nucleic Acids Res. 2015;43(Database issue):D1057–63.

37. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: A web-based tool for Gene Ontology searching. Bioinformatics. 2009;25(22):3045–6.

38. Chibucos MC, Siegele DA, Hu JC, Giglio M. The Evidence and Conclusion Ontology (ECO): Supporting GO annotations. Methods Mol Biol. 2017;1446:245–59.

39. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: Current approaches and outstanding challenges. PLoS Comput Biol. 2012;8(2):e1002375.

40. Huntley RP, Harris MA, Alam-Faruque Y, Blake JA, Carbon S, Dietze H, et al. A method for increasing expressivity of Gene Ontology annotations using a compositional approach. BMC Bioinformatics. 2014;15:155.

41. Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. The GOA database in 2009 – An integrated Gene Ontology annotation resource. Nucleic Acids Res. 2009;37(Database issue):D396–403.

42. Thomas PD, Hill DP, Mi H, Osumi-Sutherland D, Van Auken K, Carbon S, et al. Gene Ontology causal activity modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. Nat Genet. 2019;51(10):1429–1433.

43. Meldal BH, Forner-Martinez O, Costanzo MC, Dana J, Demeter J, Dumousseau M, et al. The Complex Portal – An encyclopaedia of macromolecular complexes. Nucleic Acids Res. 2015;43(Database issue):D479–84.

44. Huntley RP, Sitnikov D, Orlic-Milacic M, Balakrishnan R, D'Eustachio P, Gillespie ME, et al. Guidelines for the functional annotation of microRNAs using the Gene Ontology. RNA. 2016;22(5):667–76.

45. Huntley RP, Kramarz B, Sawford T, Umrao Z, Kalea A, Acquaah V, et al. Expanding the horizons of microRNA bioinformatics. RNA. 2018;24(8):1005–17.

46. RNAcentral Consortium. RNAcentral: An international database of ncRNA sequences. Nucleic Acids Res. 2015;43(Database issue):D123–9.

47. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: Expanded protein families and functions, and analysis tools. Nucleic Acids Res. 2016;44(D1):D336–42.

48. Kramer A, Green J, Pollard J Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523–30.

49. Dubovenko A, Nikolsky Y, Rakhmatulin E, Nikolskaya T. Functional analysis of OMICs data and small molecule compounds in an integrated "knowledge-based" platform. Methods Mol Biol. 2017;1613:101–24.

50. Avramopoulos D, Szymanski M, Wang R, Bassett S. Gene expression reveals overlap between normal aging and Alzheimer's disease genes. Neurobiol Aging. 2011;32(12):2319 e27–34.

51. Brooks LRK, Mias GI. Data-driven analysis of age, sex, and tissue effects on gene expression variability in Alzheimer's disease. Front Neurosci. 2019 Apr 24;13:392.

52. Newman M, Hin N, Pederson S, Lardelli M. Brain transcriptome analysis of a familial Alzheimer's disease-like mutation in the zebrafish presenilin 1 gene implies effects on energy production. Mol Brain. 2019 Dec 3;12(1):43.

53. Carbajosa G, Malki K, Lawless N, Wang H, Ryder JW, Wozniak E, et al. Loss of Trem2 in microglia leads to widespread disruption of cell coexpression networks in mouse brain. Neurobiol Aging. 2018 Sep;69:151–66.

54. Perreau VM, Orchard S, Adlard PA, Bellingham SA, Cappai R, Ciccotosto GD, et al. A domain level interaction network of amyloid precursor protein and Abeta of Alzheimer's disease. Proteomics. 2010;10(12):2377–95.

55. Karri V, Ramos D, Martinez JB, Odena A, Oliveira E, Coort SL, et al. Differential protein expression of hippocampal cells associated with heavy metals (Pb, As, and MeHg) neurotoxicity: Deepening into the molecular mechanism of neurodegenerative diseases. J Proteomics. 2018 Sep 15;187:106–25.

56. Shahbaba B, Shachaf CM, Yu Z. A pathway analysis method for genome-wide association studies. Stat Med. 2012;31(10):988–1000.

57. The UniProt Consortium UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158–69.

58. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, et al. Gene: A gene-centered information resource at NCBI. Nucleic Acids Res. 2015 Jan 28;43(D1):D36–42.

59. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2012. Nucleic Acids Res. 2012;40(Database issue):D84–90.

60. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.

61. Alam-Faruque Y, Huntley RP, Khodiyar VK, Camon EB, Dimmer EC, Sawford T, et al. The impact of focused Gene Ontology curation of specific mammalian systems. PLoS One. 2011;6(12):e27541.

62. Khodiyar VK, Hill DP, Howe D, Berardini TZ, Tweedie S, Talmud PJ, et al. The representation of heart development in the Gene Ontology. Dev Biol. 2011;354(1):9–17.

63. Lovering RC, Camon EB, Blake JA, Diehl AD. Access to immunology through the Gene Ontology. Immunology. 2008;125(2):154–60.

64. Xu J, Patassini S, Rustogi N, Riba-Garcia I, Hale BD, Phillips AM, et al. Regional protein expression in human Alzheimer's brain correlates with disease severity. Commun Biol. 2019 Dec 4;2(1):43.

65. Chouliaras L, Pishva E, Haapakoski R, Zsoldos E, Mahmood A, Filippini N, et al. Peripheral DNA methylation, cognitive decline and brain aging: pilot findings from the Whitehall II imaging study. Epigenomics. 2018 May;10(5):585–95.

66. Watanabe Y, Hirao Y, Kasuga K, Tokutake T, Semizu Y, Kitamura K, et al. Molecular network analysis of the urinary proteome of Alzheimer's disease patients. Dement Geriatr Cogn Dis Extra. 2019 Feb 8;9(1):53–65.

67. Orchard S, Albar JP, Deutsch EW, Eisenacher M, Binz PA, Martinez-Bartolome S, et al. From proteomics data representation to public data flow: A report on the HUPO-PSI workshop September 2011, Geneva, Switzerland. Proteomics. 2012;12(3):351–5.

68. Garcia O, Saveanu C, Cline M, Fromont-Racine M, Jacquier A, Schwikowski B, et al. GOlorize: A Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring. Bioinformatics. 2007;23(3):394–6.

69. Maere S, Heymans K, Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics. 2005;21(16):3448–9.

70. Gene Ontology Consortium. Gene Ontology Consortium: Going forward. Nucleic Acids Res. 2015;43(Database issue):D1049–56.