Check for updates

METHOD ARTICLE

REVISED Introducing high school students to the Gene Ontology classification system [version 2; peer review: 1 approved, 1 approved with reservations]

Mehek Dedhia¹, Kenneth Kohetuk ², Wim E. Crusio ^{3,4}, Anna Delprato ¹

¹BioScience Project, Wakefield, MA, 01880, USA

²Saint Dominic Savio Catholic High School, Austin, TX, 78717, USA

³Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (UMR 5287), Pessac, France

⁴University of Bordeaux (UMR 5287), Pessac, France

v2 First published: 01 Mar 2019, 8:241
<https://doi.org/10.12688/f1000research.18061.1>
 Second version: 15 Apr 2019, 8:241
<https://doi.org/10.12688/f1000research.18061.2>
 Third version: 24 May 2019, 8:241
<https://doi.org/10.12688/f1000research.18061.3>
 Latest published: 06 Aug 2019, 8:241
<https://doi.org/10.12688/f1000research.18061.4>

Abstract

We present a tutorial that introduces high school students to the Gene Ontology classification system which is widely used in genomics and systems biology studies to characterize large sets of genes based on functional and structural information. This classification system is a valuable and standardized method used to identify genes that act in similar processes and pathways and also provides insight into the overall architecture and distribution of genes and gene families associated with a particular tissue or disease. By means of this tutorial, students learn how the classification system works through analyzing a gene set using DAVID the Database for Annotation, Visualization and Integrated Discovery that incorporates the Gene Ontology system into its suite of analysis tools. This method of profiling genes is used by our high school student interns to categorize gene expression data related to behavioral neuroscience. Students will get a feel for working with genes and gene sets, acquire vocabulary, obtain an understanding of how a database is structured and gain an awareness of the vast amount of information that is known about genes as well as the online analysis tools to manage this information that is nowadays available. Based on survey responses, students intellectually benefit from learning about the Gene Ontology System and using the DAVID tools, they are better prepared for future database use and they also find it enjoyable.

Keywords

gene ontology, high school students, genomics

Open Peer Review

Approval Status

	1	2	3	4
version 4				
(revision)				view
06 Aug 2019				
version 3				
(revision)			view	view
24 May 2019				
version 2				
(revision)	view	view		
15 Apr 2019				
version 1				
01 Mar 2019	view	view		

1. **William Grisham**, Department of Psychology, Los Angeles, USA

2. **Jennifer A. Ufnar**, Vanderbilt University, Nashville, USA

3. **Ruth Lovering** , University College London, London, UK

4. **Pascale Gaudet** , SIB Swiss Institute of Bioinformatics, Geneva, Switzerland

This article is included in the **Teaching and communicating science in a digital age** collection.

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Anna Delprato (adelprato@bioscienceproject.org)

Author roles: **Dedhia M:** Resources, Visualization, Writing – Review & Editing; **Kohetuk K:** Methodology, Writing – Original Draft Preparation; **Crusio WE:** Methodology, Writing – Review & Editing; **Delprato A:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2019 Dedhia M *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Dedhia M, Kohetuk K, Crusio WE and Delprato A. **Introducing high school students to the Gene Ontology classification system [version 2; peer review: 1 approved, 1 approved with reservations]** F1000Research 2019, 8:241 <https://doi.org/10.12688/f1000research.18061.2>

First published: 01 Mar 2019, 8:241 <https://doi.org/10.12688/f1000research.18061.1>

REVISED Amendments from Version 1

This revised version of the article contains the results from a student survey which was conducted to address both Reviewers' comments. The survey responses are included in 3 tables and the results of the survey are mentioned in the abstract, results and conclusion sections. The abstract has also been revised to sound less like a proposal (Reviewer 2 comment). We have added a few sentences to relate the tutorial back to behavioral neuroscience and we provide a suggestion for expanding the scope of this project in response to Reviewer 2. Finally, we have added several Entrance and Exit Tickets that are more challenging in response to Reviewer 1.

See referee reports

Introduction

Genomics is the branch of biology concerned with the study of genes and their functions (see the [National Institutes of Health Frequently Asked Questions about Genetic and Genomic Science](#)). Genomics arose from the acceleration of genetic research which was fueled by the development of rapid and affordable DNA sequencing technologies ([Shendure et al., 2017](#)). This opened the door to the sequencing of entire genomes. Presently, the DNA codes for thousands of genomes from diverse species have been sequenced and studied (see the [National Center for Biotechnology Information Genome database](#)).

The goals in genomics research are to address all genes and their inter-relationships in order to understand the combined influence on the function of an organism. With this newfound knowledge of the staggering number of genes that make up an organism, the [Gene Ontology](#) (GO) classification system was created to organize genes by their similarities and differences (see [Gene Ontology Consortium 'About' page](#)). "Ontology" is not a commonly encountered term and there are several definitions that are related to philosophical concepts.

In the context of information science, as described here, "ontology" is concerned with the representation, formal naming and classification system with the purpose of describing the relationship categories and properties of the data.

This classification system provides the scientific community with a structured vocabulary for defining genes ([Ashburner et al., 2000](#); [du Plessis et al., 2011](#); [Hastings, 2017](#); [Thomas, 2017](#)). GO terms are commonly used in most, if not all, databases and analysis tools relevant to bioinformatics, systems biology ([Wanijek, 2011](#)), and genomics studies ([du Plessis et al., 2011](#)). GO terms are species specific and are continuously revised and expanded as biological knowledge is obtained ([Gaudet et al., 2017](#)).

The importance of the GO term system becomes apparent when analyzing the organization of genomes and coding regions, the distribution of genes involved in specific processes and the conservation of genes across species ([Gaudet et al., 2017](#)). This classification system is also quite powerful when analyzing data from large scale gene expression studies ([du Plessis et al., 2011](#))

that consider co-expression data from specific tissues obtained under defined circumstances such as treatment with pharmaceutical agents, or with neurodevelopmental disorders, cancer, or diabetes as examples. GO terms are instrumental for understanding the functions of these genes.

Introducing GO terms and the gene classification system to high school students will bring them up to speed on a commonly used research tool in current genomics methods and expose them to the vast amounts of data that have been derived from genomics and systems biology studies.

In the subsequent sections we show an example of how to extract information about a gene from its associated GO terms and then provide instruction for a practical exercise which will enable students to profile a list of genes using GO terms in the bioinformatics resource [DAVID](#), The [Database for Annotation, Visualization and Integrated Discovery](#). This is a protocol that we teach to our high school student interns when they are evaluating gene expression data for their summer projects ([Crusio et al., 2017](#), see [BioScience Project student posters](#)). The student research internship projects are in the context of behavioral neuroscience. Students typically work with gene expression data associated with a specific brain region or brain disorder. As an example for projects related to learning and memory, gene expression data for the hippocampus would be used. Concerning a neurodevelopmental disorder like Schizophrenia, gene expression data for the prefrontal cortex would be considered. There are many online databases that have freely available gene expression data and this could be a way to expand the scope of this tutorial. We use the Allen Brain Atlas for our primary source of gene expression data in the student internship projects.

Procedure

Making sense of a gene

The overall structure of GO is hierarchical and is based on parent-child terms where the parent term is broader and child term is more specialized.

GO terms group genes according to 3 categories, each of which are considered a distinct ontology: Molecular Function (MF, molecular-level activities performed by gene products), Biological Process (BP, the larger processes, or biological programs accomplished by multiple molecular activities), and Cellular Component (CC, the locations relative to cellular structures in which a gene product performs a function).

As an example, consider the GO term classification for the *RAB5A* gene ([Figure 1](#)). *RAB5A* belongs to a family of genes called Rab GTPases that are key regulators of intracellular membrane trafficking. Rabs are involved in the formation of transport vesicles and their fusion with membranes. They are enzymes and mediate their function by cycling between a GDP bound inactive and a GTP bound active state. Because of their fundamental and ubiquitous role, this family of genes are associated with many biological processes and diseases.

Rab5a	RAB5A, member RAS oncogene family(RAB5A)	Related Genes	Homo sapiens
BIOCARTA	The role of FYVE-finger proteins in vesicle transport, Phosphoinositides and their downstream targets., Rab GTPases Mark Targets In The Endocytotic Machinery,		
GOTERM_BP_DIRECT	endocytosis, phagocytosis, small GTPase mediated signal transduction, blood coagulation, protein transport, regulation of endocytosis, synaptic vesicle recycling, viral RNA genome replication, early endosome to late endosome transport, positive regulation of exocytosis, regulation of endosome size, regulation of filopodium assembly, receptor internalization involved in canonical Wnt signaling pathway, regulation of synaptic vesicle exocytosis, regulation of autophagosome assembly,		
GOTERM_CC_DIRECT	ruffle, intracellular, cytoplasm, endosome, early endosome, cytosol, plasma membrane, synaptic vesicle, endosome membrane, actin cytoskeleton, endocytic vesicle, axon, dendrite, phagocytic vesicle membrane, somatodendritic compartment, melanosome, neuronal cell body, terminal bouton, axon terminus, membrane raft, phagocytic vesicle, extracellular exosome, cytoplasmic side of early endosome membrane,		
GOTERM_MF_DIRECT	GTPase activity, protein binding, GTP binding, GDP binding,		
INTERPRO	Small GTPase superfamily, Small GTP-binding protein domain, P-loop containing nucleoside triphosphate hydrolase,		
KEGG_PATHWAY	Ras signaling pathway, Endocytosis, Phagosome, Vasopressin-regulated water reabsorption, Amyotrophic lateral sclerosis (ALS), Amoebiasis, Tuberculosis,		
UP_KEYWORDS	3D-structure, Alternative splicing, Cell membrane, Cell projection, Complete proteome, Cytoplasm, Cytoplasmic vesicle, Endocytosis, Endosome, GTP-binding, Lipoprotein, Membrane, Nucleotide-binding, Phagocytosis, Prenylation, Protein transport, Proteomics identification, Reference proteome, Transport,		
UP_SEQ_FEATURE	chain:Ras-related protein Rab-5A, helix, lipid moiety-binding region:S-geranylgeranyl cysteine, modified residue, mutagenesis site, nucleotide phosphate-binding region:GTP, sequence conflict, short sequence motif:Effector region, strand, turn,		

Figure 1. DAVID output for the *RAB5A* gene. Screenshot of the DAVID results for the *RAB5A* gene. Top left (blue bar): Gene Symbol identifier. Center (blue bar): full gene name. Labels (left): GO Term BP, GO Term CC, and GO Term MF descriptors. Note that these descriptors are clickable.

The GO term classification for the *RAB5A* gene gives:

GOTERM_BP: endocytosis, phagocytosis, small GTPase mediated signal transduction, blood coagulation, protein transport, regulation of endocytosis, synaptic vesicle recycling, viral RNA genome replication, early endosome to late endosome transport, positive regulation of exocytosis, regulation of endosome size, regulation of filopodium assembly, receptor internalization involved in canonical Wnt signaling pathway, regulation of synaptic vesicle exocytosis, regulation of autophagosome assembly

GOTERM_CC: ruffle, intracellular, cytoplasm, endosome, early endosome, cytosol, plasma membrane, synaptic vesicle, endosome membrane, actin cytoskeleton, endocytic vesicle, axon, dendrite, phagocytic vesicle membrane, somatodendritic compartment, melanosome, neuronal cell body, terminal bouton, axon terminus, membrane raft, phagocytic vesicle, extracellular exosome, cytoplasmic side of early endosome membrane.

GOTERM_MF: GTPase activity, protein binding, GTP binding, GDP binding

From the *RAB5A* related GO terms, we get the overall impression that this gene encodes an enzyme that is involved in signaling, transport and vesicle dynamics and is associated with cell membranes. How do we arrive at this description?

In this example, the information obtained from the **MF** category is that the protein product of the *RAB5A* gene binds to guanine nucleotides: GTP and GDP (Guanosine tri and di phosphate, respectively) and that it is an enzyme. This is evident by the “GTPase activity” term. Whenever the suffix “ase” is used in the context of a gene or protein, it refers to an enzyme, something that catalyzes a chemical reaction. For the **BP** category, there are several terms associated with intracellular transport, signaling, and endocytosis. Finally, the terms associated

with **CC** include endosome and endosome-like organelles (melanosomes, synaptic vesicles, phagocytic vesicles), as well as membrane structures (ruffles, rafts).

Gene Profiling in DAVID

DAVID is primarily a clustering program that groups genes based on different criteria related to GO terms. DAVID links to other databases that contain complementary information like The Gene Ontology. In this exercise, students will use the sample gene lists (DEMOLIST1 or DEMOLIST2) that are accessible from the DAVID database to see how the Gene Ontology classification partitions a set of genes based on GO Terms. Screenshots and videos are provided for step by step instruction. We also provide a video to instruct students on profiling a gene list in DAVID obtained from a random gene list generator.

Protocol

Screenshot 1 (Figure 2). DAVID landing page. The start analysis link is accessed here and is circled in red in this image. (Video 1, Delprato *et al.*, 2019a)

Screenshot 2 (Figure 3). Submitting a gene list. Select either DEMOLIST 1 or DEMOLIST 2 (left panel). The identifier will come up automatically because this is a demonstration list. If you are submitting your own gene list then, the identifier will have to be specified from the dropdown menu (Video 2, Delprato *et al.*, 2019b). Typically the identifier is the “Official Gene Symbol”. Click “Gene List”, then “Submit List” (Video 1; (Delprato *et al.*, 2019a)).

Screenshot 3 (Figure 4). Species selection. You will see a notice: “Multiple Species, have been Detected”, Highlight “Homo Sapiens” in the window, Select “Homo Sapiens” below the window (Example - DEMOLIST 1: 149 genes, highlighted in grey, left panel). Next, you will see the message “Submission Successful” (Video 1; Delprato *et al.*, 2019a).

Figure 2. David landing page. Screenshot of the DAVID landing page containing a brief description of the site and links for available tools. The "start analysis" button is circled in red and is located at the top left side of the page. This is the first step in submitting a geneset for profiling with GO Terms in DAVID.

Figure 3. Submitting a geneset. Screenshot of the page where a geneset can be submitted "Step1: Enter Gene List", there are options to copy and paste a geneset or upload a file. There is also an option to use either of two sample lists provided by the DAVID site. Under the pull down menu, "Step 2 Select Identifier", there are many types of identification designations for the same gene. For the sample lists, the identifier will come up automatically. When submitting a gene list from the geneset generator, the identifier is "Official Gene Symbol" which is used in most cases. If the identifier that you choose does not match the identifiers in the submitted gene set, you will receive a message stating this and an option to convert to the correct identifier.

Figure 4. Species selection. Screenshot of a successfully submitted gene set. Here it is necessary to select the species. In this instance “Homo sapiens” is highlighted along with the number of genes that are recognized by the site and for which information is available. Once the species is highlighted, it is necessary to click the “Select Species” button in order to limit the output to just the desired species.

Screenshot 4 (Figure 5). Obtaining the results. Select “Functional Annotation Tool”, beneath the blue arrow. Next, select “Functional Annotation Table”, Bottom of the page (Video 1, Delprato *et al.*, 2019a).

Screenshot 5 (Figure 6). Reading the output. The gene ID and the full gene name are shown in the blue bars above each entry. The GO Term BP (Biological Process), GO Term CC (Cellular Component), and GO Term MF (Molecular Function), terms are clickable descriptors and link to the Gene Ontology website. See above for a complete description of the GO categories (Video 1, Delprato *et al.*, 2019a).

Screenshot 6 (Figure 7). Keyword search. When selecting terms for a keyword search, a more complete outcome is achieved if just a few letters are specified. For example, -“neur” will capture terms both starting with neuro and neural (Video 1, Delprato *et al.*, 2019a). DAVID output can be searched for genes related to other process and diseases as well. Have students evaluate the gene list based on their interest. They can identify genes related to a particular process. Students may work individually or in groups.

Optional exercise

Students may wish to try this with their own gene lists. This online gene list generator will enable students to generate a

random list of genes for evaluation. (See also Video 2 for instruction; Delprato *et al.*, 2019b)

Protocol

Step 1. Specify species: Human is the default

Step 2. Specify list length: 200-500 is a good representative number. Note that DAVID will not evaluate lists with more than 2000 genes. An error message stating this will be received.

Step 3. Select “Generate”

Step 4. Copy the gene list using the “Select All” option and paste the list directly into DAVID for evaluation as described above. Make sure to select “Official Gene Symbol” as the identifier when submitting the gene list.

Learning assessment

Student survey. We polled 12 student interns from our summer program for their feedback on learning about the Gene Ontology System and using the DAVID tools. The responses are shown in Table 1, Table 2, and Table 3. Table 1 contains the answers to direct yes or no type questions and Table 2 and Table 3 are based on short answer responses. To summarize the data, students believe that they benefited intellectually from this work and they enjoy this type of learning experience. They also

Secure | https://david.ncifcrf.gov/summary.jsp

Functional Annotation Tool
DAVID Bioinformatics Resources 6.8, NIAID/NIH

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

*** Welcome to DAVID 6.8 ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Upload List Background

Gene List Manager

Select to limit annotations by one or more species [Help](#)

- Use All Species -
Homo sapiens(149)
Unknown(15)

Select Species

List Manager [Help](#)

demolist1

Select List to:
Use Rename
Remove Combine
Show Gene List

[View Unmapped Ids](#)

Annotation Summary Results [Help and Tool Manual](#)

Current Gene List: demolist1
Current Background: Homo sapiens
145 DAVID IDs
Check Defaults ☒ Clear All

☒ Disease (1 selected)
☒ Functional_Categories (3 selected)
☒ Gene_Ontology (3 selected)
☐ General_Annotations (0 selected)
☐ Literature (0 selected)
☐ Main_Accessions (0 selected)
☒ Pathways (3 selected)
☒ Protein_Domains (3 selected)
☐ Protein_Interactions (0 selected)
☐ Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Figure 5. Obtaining the results. Screenshot of the different types of analysis options provided by DAVID for a given geneset. For the purpose of this tutorial, the relevant output is the "Functional Annotation Table" located at the bottom of the page.

Secure | https://david.ncifcrf.gov/annotationReport.jsp?annot=59,12,87,88,30,38,46,3,5,55,53,70,79¤tList=0

DAVID Bioinformatics Resources 6.8
Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

*** Welcome to DAVID 6.8 ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Functional Annotation Table [Help and Manual](#)

Current Gene List: demolist1
Current Background: Homo sapiens
145 DAVID IDs

139 record(s) [Download File](#)

37166_at	3-hydroxyanthranilate 3,4-dioxygenase(HAAO)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	tryptophan catabolic process, response to zinc ion, quinolinate biosynthetic process, 'de novo' NAD biosynthetic process from tryptophan, anthranilate metabolic process, response to cadmium ion, quinolinate metabolic process, oxidation-reduction process, neuron cellular homeostasis,		
GOTERM_CC_DIRECT	cytosol, extracellular exosome,		
GOTERM_MF_DIRECT	3-hydroxyanthranilate 3,4-dioxygenase activity, iron ion binding, protein binding, ferrous iron binding, electron carrier activity,		
INTERPRO	3-hydroxyanthranilic acid dioxygenase, RmlC-like cupin domain, RmlC-like jelly roll fold, 3-hydroxyanthranilate 3, 4-dioxygenase, metazoan,		
KEGG_PATHWAY	Tryptophan metabolism, Metabolic pathways,		
PIR_SUPERFAMILY	3-hydroxyanthranilate 3,4-dioxygenase, animal type,		
UP_KEYWORDS	3D-structure, Alternative splicing, Complete proteome, Cytoplasm, Dioxygenase, Iron, Metal-binding, Oxidoreductase, Polymorphism, Proteomics identification, Pyridine nucleotide biosynthesis, Reference proteome,		
UP_SEQ_FEATURE	binding site:Dioxygen, binding site:Substrate, chain:3-hydroxyanthranilate 3,4-dioxygenase, helix, metal ion-binding site:Iron; catalytic, sequence variant, splice variant, strand, turn,		
34467_g_at	5-hydroxytryptamine receptor 4(HTR4)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	G-protein coupled receptor signaling pathway, G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger, serotonin receptor signaling pathway, gamma-aminobutyric acid signaling pathway, chemical synaptic transmission, regulation of appetite,		
GOTERM_CC_DIRECT	cytoplasm, endosome, plasma membrane, integral component of plasma membrane, membrane,		
GOTERM_MF_DIRECT	G-protein coupled serotonin receptor activity,		
INTERPRO	G-protein-coupled receptor, rhodopsin-like, 5-Hydroxytryptamine 4 receptor, GPCR, rhodopsin-like, 7TM,		

Figure 6. Interpreting the output. Screenshot of the DAVID "Functional Annotation Table" results for a geneset. Top left: (blue bar): Gene Symbol identifier. Center (blue bar): full gene name. Labels (left): GO Term BP, GO Term CC, and GO Term MF descriptors. Note that these descriptors are clickable.

Secure | <https://david.ncicrf.gov/annotationReport.jsp?annot=59,12,87,88,30,38,46,3,5,55,53,70,79¤tList=0>

KEGG_PATHWAY	ABC transporters, Lysosome,		
SMART	AAA,		
UP_KEYWORDS	Alternative splicing, ATP-binding, Complete proteome, Lysosome, Membrane, Nucleotide-binding, Peptide transport, Polymorphism, Protein transport, Proteomics identification, Reference proteome, Transmembrane, Transmembrane helix, Transport,		
UP_SEQ_FEATURE	chain:ATP-binding cassette sub-family B member 9, compositionally biased region:Poly-Glu, domain:ABC transmembrane type-1, domain:ABC transporter, modified residue, nucleotide phosphate-binding region:ATP, sequence variant, signal peptide, splice variant, transmembrane region,		
32439_at	ATPase H ⁺ /K ⁺ transporting beta subunit(ATP4B)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	potassium ion transport, sodium ion transport, cell adhesion, response to organonitrogen compound, response to lipopolysaccharide, ion transmembrane transport, hydrogen ion transmembrane transport,		
GOTERM_CC_DIRECT	plasma membrane, sodium:potassium-exchanging ATPase complex,		
GOTERM_MF_DIRECT	protein binding, hydrogen:potassium-exchanging ATPase activity,		
INTERPRO	Sodium/potassium-transporting ATPase subunit beta,		
KEGG_PATHWAY	Oxidative phosphorylation, Collecting duct acid secretion, Gastric acid secretion,		
UP_KEYWORDS	Cell adhesion, Cell membrane, Complete proteome, Disulfide bond, Glycoprotein, Hydrogen ion transport, Ion transport, Membrane, Potassium, Potassium transport, Reference proteome, Signal-anchor, Transmembrane, Transmembrane helix, Transport,		
UP_SEQ_FEATURE	chain:Potassium-transporting ATPase subunit beta, disulfide bond, glycosylation site:N-linked (GlcNAc...), topological domain:Cytoplasmic, topological domain:Extracellular, transmembrane region,		
39469_s_at	ATPase phospholipid transporting 8A2(ATP8A2)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	involuntary skeletal muscle contraction, axonogenesis, aging, negative regulation of cell proliferation, retina layer formation, positive regulation of neuron projection development, response to auditory stimulus, positive regulation of multicellular organism growth, inner ear morphogenesis, eating behavior, skin development, phospholipid translocation, Golgi vesicle budding, neuron development, neuromuscular process controlling posture, detection of light stimulus involved in visual perception, neurofilament cytoskeleton organization, positive regulation of phospholipid translocation,		
GOTERM_CC_DIRECT	Golgi membrane, photoreceptor outer segment, endosome, Golgi apparatus, trans-Golgi network, plasma membrane, integral component of membrane,		
GOTERM_MF_DIRECT	magnesium ion binding, phospholipid-translocating ATPase activity, ATP binding,		
INTERPRO	Cation-transporting P-type ATPase, Phospholipid-transporting P-type ATPase, subfamily IV, P-type ATPase, A domain, P-type ATPase, phosphorylation site, HAD-like domain, P-type ATPase, cytoplasmic domain N,		
OMIM_DISEASE	Cerebellar ataxia, mental retardation, and dysequilibrium syndrome 4,		
UP_KEYWORDS	Alternative splicing, ATP-binding, Cell membrane, Cell projection, Chromosomal rearrangement, Complete proteome, Disease mutation, Endosome, Golgi apparatus, Hydrolase, Lipid transport, Magnesium, Membrane, Mental retardation, Metal-binding, Nucleotide-binding, Phosphoprotein, Polymorphism, Proteomics identification, Reference proteome, Transmembrane, Transmembrane helix, Transport,		
UP_SEQ_FEATURE	active site:4-aspartylphosphate intermediate, chain:Probable phospholipid-transporting ATPase IB, metal ion-binding site:Magnesium, modified residue, sequence variant, splice variant, topological domain:Cytoplasmic, topological domain:Extracellular, transmembrane region,		
33685_at	ATPase phospholipid transporting 8B3(ATP8B3)	Related Genes	Homo sapiens
GOTERM_BP_DIRECT	Golgi organization, binding of sperm to zona pellucida, phospholipid translocation,		
GOTERM_CC_DIRECT	acrosomal membrane, endoplasmic reticulum membrane, Golgi apparatus, plasma membrane, integral component of membrane,		
GOTERM_MF_DIRECT	magnesium ion binding, phospholipid-translocating ATPase activity, ATP binding,		
INTERPRO	Cation-transporting P-type ATPase, Phospholipid-transporting P-type ATPase, subfamily IV, P-type ATPase, A domain, P-type ATPase,		

Figure 7. Keyword search. Screenshot of a keyword search of the DAVID output to identify genes that are relevant to a given biological function. In this case the search term is “neur” (highlighted in yellow), to identify genes related to neurological processes. The keyword search is done via the general search function on your computer. For larger gene sets, a computer program may be used.

Table 1. DAVID and GO Student Survey.

Questions	Responses		
	Yes	No	Other
1. Have you ever heard of or did you have any experience working with the DAVID database prior to the summer internship with BioScience Project?	0	12	0
2. Have you ever heard of the Gene Ontology Classification (GO Terms) System prior to the summer internship with BioScience Project?	2	10	0
3. Did you benefit intellectually from working with the DAVID tools and learning about the Gene Ontology System?	12	0	0
4. Did you enjoy working with the DAVID tools and learning about gene profiling with the Gene Ontology System?	12	0	0
5. If applicable, are you better able to navigate other genomic databases as a result of working in DAVID?	11	0	1

Table 2. Student Short Answer Responses to question 3a. (Question 3: Did you benefit intellectually from working with the DAVID tools and learning about the Gene Ontology System?) 3a: If yes, please explain how?.

Through this internship using DAVID, I was exposed for the first time to the field of bioinformatics and the gene ontology system. I learned more about the role genes play in neurological disorders in humans. I also got a glimpse of what biomedical researchers and neuroscientists actually do and the type of resources that they work with.

I learned about current methods used by researchers and was able to apply those techniques to my own project. The exposure and experience I gained with bioinformatics gave me a better understanding of gene interactions and how researchers analyze them. The tools I used during the internship furthered my understanding of gene profiling and provided me with insight on the importance of research.

Rather than just learning about genetics as I had in biology classes in the past, I had the opportunity to partake in hands-on learning with the DAVID tools. Through using DAVID and the Gene Ontology system, I gained a broader understanding of the biological functions of genes by seeing and being exposed to such a large variety. Being able to apply what I learned from DAVID and the results I gathered to my own independent project helped to further my understanding of genetics while exposing me to the field of bioinformatics.

I learned different ways genes are classified and analyzed (kegg pathway, etc.). I can use this database for future research projects to get a sense of different pathways these genes are involved in.
I learnt a lot about genomics and how interconnected the different genes in our body are. It also helped me work on my project on Alzheimer's and the APOE 4 allele.
I was able to learn more about PTSD and identifying candidate genes for PTSD. I performed this project with DAVID and learned a lot about the Gene Ontology System.
The DAVID tools helped me cluster annotation terms based on keywords associated with neurological disorders. Also, the KEGG Pathway map allowed me to visualize genes and how they interact.
This was the first time that I worked with either of these tools and I think I benefited mostly because I got an idea of what working in the field of Biology/Neuroscience would be like.
I was able to learn more about the GO System, different databases such as the Allen Brain Atlas and StringDB, which I used in conjunction with DAVID and learned about how DAVID analyzes the data.
I found that working with the DAVID tools and Gene Ontology System I was exposed to a real world experience in science that gave me a better understanding of where we are in research now, and what is still to be done. The benefits I found in myself were a capital in scientific nomenclature, new skills in analysis of data, and a wholesome exposure into the field of genomics research.
I learned a lot about gene databases and how to look at specific sets of data while ignoring information that may be necessary.
I felt like working with the DAVID tools exposed me to an area of research that I wasn't previously familiar with in an engaging and fascinating way. Learning how to find specific information about gene interactions/gene functions through an online database is a valuable tool that I believe will benefit me in my future research.

Table 3. Student Short Answer Responses to question 4a. (Question 4. Did you enjoy working with the DAVID tools and learning about gene profiling with the Gene Ontology System?) 4a: If yes, please explain why?.

Working with DAVID and the other databases gave me a chance to learn about the field of bioinformatics and genetics, beyond our school curriculum. It was extremely interesting to learn about the hundreds of genes in the human genome, their various functions both at the molecular and biological level and how they affect the neurological characteristics of human beings. The databases were also very interactive and allowed me to explore the other parts of the database myself.
Although it was initially difficult to navigate the DAVID tools, I found the experience rewarding in the end. The process became easier as I persisted in using the database, and I enjoyed being able to explore the realm of gene profiling.
I enjoyed working with the DAVID tools and learning about gene profiling because from a biological point of view, DAVID is very good at finding relevant information, like other correlates, to the keyword I'm looking for.
I felt like what I was learning about through DAVID and the Gene Ontology system was applicable to my project and I was able to utilize and apply my knowledge of these tools effectively, making me more excited to use it.
I gained insight into gene analysis and gained knowledge that I can use in other situations.
I got practical hands-on experience working with a scientific database which was quite different from the textbook learning taught in schools. I found this quite refreshing and enjoyable.
Although it was pretty confusing for me at the time, I think what made it enjoyable was that I found using these tools interesting especially since this was all new to me! Also it was not too difficult or overwhelming since the provided instructions for the internship walked me through each step. I'm actually now studying Neuroscience at BU and I'm actually hoping to get back to relearning how to use these tools again now that I have a better understanding.
I enjoyed working with the DAVID tools and learning about gene profiling with the GO System because it was very interesting to see how different genes were connected to each other and how far reaching the effects of certain genes are.
Because the breadth and depth of information felt like a million different rabbit holes that I could fall into and learn something new from. However, these tools required navigational help and direction from the supervisor and fellow interns for me to truly reach this point of knowing how to immerse myself in it out of mere curiosity, because of how complex felt at first, and I still have so much more to learn, but overall enjoyed working with these tools once I was comfortable with them.
It was a really interesting and valuable experience to have, and I feel like I learned a lot about how different genes may be connected to each other and what is important to consider and look for in gene profiling.
I thought it was interesting to be able to visualize some of the molecular pathways through the diagrams provided. In completing my research, it was helpful to have all of the biological processes and molecular functions of certain genes all in one place
I really enjoyed working on a topic that was interesting to me. I could learn about PTSD while learning more about biology and gene profiling.

state that as a result of this experience, they are better prepared for future database use.

Entry and exit tickets

A basic entry and exit ticket method is suggested to determine what students know about genomics and genes before the lesson as well as what they have learned: main points, questions they may have and what they found most interesting. Sample questions are provided in what follows.

Entry ticket questions and answers

1. What is a gene?

A sequence of DNA or RNA which codes for a molecule that has a function. A gene is the basic physical and functional unit of heredity

2. What is genomics?

Study of the full set of the genes and DNA in an organism

3. How many protein coding genes does a human have?

~20,000

4. Do humans all have the same genes?

Yes, but people have different alleles. Alleles are the variation of a gene resulting from mutations. As an example consider eye color. We all have the gene for eye color but some of us have brown, blue or green eyes and there are different shades and hues within those categories.

5. Do genes work together?
6. If yes, provide an example
7. Have you ever worked with a biological database?

Exit ticket questions

1. What were the main points of the lesson?
2. Do you have any questions?
3. What aspect of this lesson did you find most interesting?
4. Make up a gene and describe it using Gene Ontology classifiers for the 3 categories Molecular Function (MF), Biological Process (BP), and Cellular Component (CC)
5. Why do some genes have many classifiers while others do not?

6. For the Biological Process – BP category, what are the classifiers based on, i.e., how are they derived?
7. How do you think you could use this database in a high school research project?

Conclusions

We describe a procedure for students to become acquainted with the Gene Ontology classification systems which is widely used in genomics and systems biology research to characterize gene function. Grouping genes with GO Terms and the DAVID database is based on a protocol that we use with our summer interns to profile gene expression data related to behavioral neuroscience studies (Crusio *et al.*, 2017). Grouping genes in this way identifies genes that function in like processes and also provides information about the overall distribution of a set of genes associated with a particular tissue or process. This tutorial will familiarize early stage students with a biological database and teach them how to mine and extract useful information from a sample list of genes. Based on survey responses, students benefit intellectually from learning about and using these tools, they are better prepared for future database use, and also find it enjoyable. Entry and exit ticket questions are also included as a formative assessment strategy.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required

Extended data

Extended data is available from figshare

Figshare: Extended data 1. Video 1: GeneSetProfiling Instructional video for using DAVID to obtain Gene Ontology classifiers for a sample geneset which is provided by the DAVID site <https://doi.org/10.6084/m9.figshare.7649225.v1> (Delprato *et al.*, 2019a)

Figshare: Extended data 2. Video 2. UploadGeneSet Instructional video for generating a random geneset and submitting this geneset to DAVID for Gene Ontology classification, <https://doi.org/10.6084/m9.figshare.7649231.v1> (Delprato *et al.*, 2019b)

Grant information

The author(s) declared that no grants were involved in supporting this work.

References

- Ashburner J, Ball CA, Blake JA, *et al.*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet.* 2000; 25(1): 25–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Crusio WE, Rubino C, Delprato A: **Engaging high school students in neuroscience research -through an e-internship program [version 2; referees: 3 approved].** *F1000Res.* 2017; 6: 20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Delprato A, Dedhia M, Crusio W, *et al.*: **Video_1 GeneSetProfiling.mp4.** *figshare.* Media. 2019a.
<http://www.doi.org/10.6084/m9.figshare.7649225.v1>

- Delprato A, Crusio W, Dedhia M, *et al.*: **Video_2.UploadGeneSist.** *figshare.* Media. 2019b.
<http://www.doi.org/10.6084/m9.figshare.7649231.v1>
- du Plessis L, Skunca N, Dessimoz C: **The what, where, how and why of gene ontology—a primer for bioinformaticians.** *Brief Bioinform.* 2011; 12(6): 723–35.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gaudet P, Škunca N, Hu JC, *et al.*: **Primer on the Gene Ontology.** In: Dessimoz C, Škunca N. (eds) *The Gene Ontology Handbook. Methods Mol Biol.* Humana Press, New York, NY. 2017; 1446: 25–37.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hastings J: **Primer on Ontologies**. In: Dessimoz C, Škunca N. (eds) *The Gene Ontology Handbook. Methods Mol Biol*. Humana Press, New York, NY. 2017; **1446**: 3–13.

[PubMed Abstract](#) | [Publisher Full Text](#)

Shendure J, Balasubramanian S, Church GM, *et al.*: **DNA sequencing at 40: past, present and future**. *Nature*. 2017; **550**(7676): 345–353.

[PubMed Abstract](#) | [Publisher Full Text](#)

Thomas PD: **The Gene Ontology and the Meaning of Biological Function**. In: Dessimoz C, Škunca N. (eds) *The Gene Ontology Handbook. Methods Mol Biol*. Humana Press, New York, NY. 2017; **1446**: 15–24.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wanjek C: **Systems Biology as Defined by NIH**. *The NIH Catalyst*. 2011; **19**(6). Date accessed Dec 27, 2018.

[Reference Source](#)

Open Peer Review

Current Peer Review Status:

Version 2

Reviewer Report 02 May 2019

<https://doi.org/10.5256/f1000research.20696.r47231>

© 2019 Ufnar J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jennifer A. Ufnar

Department of Teaching and Learning, Vanderbilt University, Nashville, TN, USA

The authors have provided some data in a table to show that the students liked the database method and felt that it was helpful in their learning. I feel that the authors need to dig more deeply into the data, though, to explain this to readers. The results and conclusions sections were very short, and did not explain the data in enough detail. As an educator or someone wanting to follow this new method, I would want more data to show that it does work and why/how it is working. Since the results are qualitative, more work with the data is needed to show that the qualitative data actually shows the reader that this method works (i.e. coding of open ended survey responses).

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular microbiology; STEM outreach.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 24 April 2019

<https://doi.org/10.5256/f1000research.20696.r47230>

© 2019 Grisham W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

William Grisham

University of California, Los Angeles (UCLA), Department of Psychology, Los Angeles, CA, USA

All of my concerns have been met, and the article seem fine as it now stands.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroscience and pedagogy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 26 March 2019

<https://doi.org/10.5256/f1000research.19752.r45178>

© 2019 Ufnar J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jennifer A. Ufnar

Department of Teaching and Learning, Vanderbilt University, Nashville, TN, USA

This article presents an interesting way to introduce gene ontologies and related vocabulary to high school students to enhance biological education. While the proposal describes the method in detail, there is no data presented to show that the method actually works. While this is an interesting protocol, it is not a fully fleshed out methods article. The authors will need to present data to show the viability of the method, or resubmit potentially as a protocol.

A few more specific comments about the article are listed below:

1. The abstract is worded as more of a proposal rather than a paper.
2. The protocol seems very much like following a set of directions (not very thought-provoking or inquiry-based), without relating it to the neuroscience. I think the paper would be much stronger if the authors were to relate it back to the behavioral neuroscience projects that are mentioned in the abstract.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

No

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular microbiology; STEM outreach

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 21 March 2019

<https://doi.org/10.5256/f1000research.19752.r45177>

© 2019 Grisham W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

William Grisham

University of California, Los Angeles (UCLA), Department of Psychology, Los Angeles, CA, USA

This article is aimed at describing an easily searchable database that can be used by students to search by gene ontologies. The point of the article is to provide a guide to use this tool in education at the high school level, but the article could provide a good starting point for college level instructors and even industry trainers. The article describes the underlying question well and chooses a gene, RAB5A, whose function should be readily understood by most biologists.

The submitted article is timely because the DAVID tool described is not very intuitive. Thus, an easy to follow guide such as the article provides is most welcome. Notably, database tools are becoming more elaborate, and their complexity has created a barrier for effectively use in teaching, particularly at lower levels. The article does a good job of describing this tool, DAVID, in a fashion that will be intelligible to high school and college level students. Lavish illustrations are provided as are links to videos to aid in instruction.

The use of Entry Ticket and Exit Ticket questions is an excellent pedagogical idea. The questions, however, seem a bit simplistic.

A bigger problem with the article is that there are no data showing how using the DAVID tool in teaching has produced changes in measurable ways. Do students do better on tests of content knowledge? Are students more sophisticated in their thinking as a function of using the DAVID

tool? Are they better able to navigate other genomic tools? Do they simply enjoy the lessons more than with traditional instruction? Many of these questions can be answered with a simple pre- and posttest quiz or even comparing current scores on tests/evaluations to historical scores. This addition would make this article of much more value to educators who could then see if incorporating this tool into their pedagogy is worthwhile.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Partly

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroscience and pedagogy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Comments on this article

Version 1

Author Response 07 Apr 2019

Anna Delprato, BioScience Project, USA

Response to Reviewers

We would like to thank the reviewers for reviewing our article and for providing insightful and constructive feedback. We have taken your recommendations into account and have revised our article accordingly. We believe that the revised version is much stronger and of more value to the

academic community than the originally submitted article.

Reviewer 1

1) The use of Entry Ticket and Exit Ticket questions is an excellent pedagogical idea. The questions, however, seem a bit simplistic.

Response: We agree with the reviewer's comment that the Entry and Exit Tickets were simplistic. To address this we have added additional questions that we think are more challenging.

2) A bigger problem with the article is that there are no data showing how using the DAVID tool in teaching has produced changes in measurable ways. Do students do better on tests of content knowledge? Are students more sophisticated in their thinking as a function of using the DAVID tool? Are they better able to navigate other genomic tools? Do they simply enjoy the lessons more than with traditional instruction? Many of these questions can be answered with a simple pre- and posttest quiz or even comparing current scores on tests/evaluations to historical scores. This addition would make this article of much more value to educators who could then see if incorporating this tool into their pedagogy is worthwhile.

Response: To address this point, we have conducted and provided the results of a student survey designed to determine the viability of the method. The survey questions and responses are provided in the newly added Tables, 1, 2, and 3. To summarize the results, students report that they benefited intellectually from this tutorial, they found it enjoyable and feel better prepared for future database work.

Reviewer 2

1) This article presents an interesting way to introduce gene ontologies and related vocabulary to high school students to enhance biological education. While the proposal describes the method in detail, there is no data presented to show that the method actually works. While this is an interesting protocol, it is not a fully fleshed out methods article. The authors will need to present data to show the viability of the method, or resubmit potentially as a protocol.

Response: This point has been addressed in our response to Reviewer 1.

A few more specific comments about the article are listed below:

1) The abstract is worded as more of a proposal rather than a paper.

Response: The abstract has been revised to make it sound less like a proposal and we have incorporated the information obtained from the survey responses.

2) The protocol seems very much like following a set of directions (not very thought-provoking or inquiry-based), without relating it to the neuroscience. I think the paper would be much stronger if the authors were to relate it back to the behavioral neuroscience projects that are mentioned in the abstract.

Response: The method presented in this article is intended to introduce students to an unfamiliar and sophisticated topic. This instruction may serve as a stepping stone for inquiry-based projects.

We expanded the set of Exit Ticket questions to stimulate ideas on how these databases and tools may be used in a research project.

*A paragraph to relate the tutorial back to the behavioral neuroscience has been added.
We also provide a suggestion for expanding the scope of the tutorial into an inquiry-based project.*

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research