

## Annotation Extensions

Rachael P. Huntley and Ruth C. Lovering

### Abstract

The specificity of knowledge that Gene Ontology (GO) annotations currently can represent is still restricted by the legacy format of the GO annotation file, a format intentionally designed for simplicity to keep the barriers to entry low and thus encourage initial adoption. Historically, the information that could be captured in a GO annotation was simply the role or location of a gene product, although genetically interacting or binding partners could be specified. While there was no mechanism within the original GO annotation format for capturing additional information about the context of a GO term, such as the target gene of an activity or the location of a molecular function, the long-term vision for the GO Consortium was to provide greater expressivity in its annotations to capture physiologically relevant information.

Thus, as a step forwards, the GO Consortium has introduced a new field into the annotation format, *annotation extensions*, which can be used to capture valuable contextual detail. This provides experimentally verified links between gene products and other physiological information that is crucial for accurate analysis of pathway and network data. This chapter will provide a simple overview of annotation extensions, illustrated with examples of their usage, and explain why they are useful for scientists and bioinformaticians alike.

**Key words** Gene Ontology, Annotation, Biocuration, Context, Pathway, Network, Analysis, Annotation extension

---

## 1 Introduction

Functional annotation of gene products using the GO has gone far in simplifying the task of finding functional roles of both individual and groups of gene products. It has enabled a multitude of analyses that were previously not possible. For example, GO annotations are invaluable for analyzing a list of genes that are identified as differentially expressed in a microarray experiment using one of the many freely available functional enrichment programs [1, 2] (*see also* Chap. 13 [3]).

The original simplistic GO annotation pairs a gene product with a GO term (one of *biological process*, *molecular function* or *cellular component*). Because these pair-wise associations are treated independently, vast amounts of correlated functional data are

omitted from the basic GO annotation and therefore inaccessible to network and pathway analyses. This contextual information is essential for understanding the physiological roles of gene products. Without contextual information bioinformatics analyses cannot identify gene products that perform a role only under certain conditions or in the presence of specific factors and therefore will present an incomplete view of the available data [4]. Specific gene products will often have different biological roles in different cells or tissues as these roles will be dependent on the available interacting partners; already tissue-specific network analyses are able to demonstrate the importance of the cellular environment. For example, Greene et al. [5] analyzed the GO and pathway annotations of the available interaction partners of the transcription factor, LEF1, in different tissue types. They demonstrated that LEF1 was significantly associated with biological processes that were relevant to each tissue type. For instance, in blood vessels the LEF1 interacting partners were associated with angiogenesis, whereas in hypothalamus they are associated with hypothalamus development.

Here we describe an incremental extension of the GO annotation format to allow more detailed statements about gene product function, which will benefit all types of functional analyses [5].

---

## 2 Extending the Core GO Annotation Model

In practical terms, the newly introduced *annotation extensions* field enables curators to provide appropriate experimentally evidenced contextual information for manually curated annotations (extant software pipelines for electronically inferred annotations (IEA) do not yet support population of this field).

Generating a comprehensive annotation, one that includes its context, involves refining the core pair-wise association with additional relationships to other ontology classes [5]. This dynamic approach is logically equivalent to creating a new term for the subtype in the ontology, but offers advantages in terms of both flexibility and efficiency.

In essence this approach allows curators to dynamically create “virtual” terms. It enables curators to combine all of the specific terms needed to fully describe a gene product in a way that can be reproducibly, computationally interpreted. For example, “core RNA polymerase binding transcription factor in hypothalamus”, associates a gene product with that activity occurring in that specific location. From the computer logic perspective this effectively has created a subclass of “core RNA polymerase binding transcription factor activity” (GO:0000990). The flexibility of expression thus supports the virtual creation of complex, compound child terms on an as-needed basis. Additionally, this approach to virtual

term creation is immediate. Because the parent can be automatically inferred from the primary term of the association, and because the additional relationships to other terms provide the refinements needed to create a more specific term, the result is that the previously independent processes of annotating gene products and creating ontology terms are now fully integrated. The use of annotation extensions means that curators can immediately make the biological statement required without having to return to the annotation to update it only after the term is available in the ontology, thus making the overall process more efficient. As these virtual terms are not consequently added to the ontology—although they could be if required—the extended annotations can be “folded” to create the logical equivalent of a GO term [5]. The GO Consortium (GOC) is in the process of incorporating these inferred annotations into the files it provides and so this contextual information will be included by default for use by anyone, or any analysis tool, that utilizes the annotation files.

---

### 3 Annotation Extension Format

Annotation extensions refine the GO term used in the basic annotation by adding one or more relational expressions (extensions). Each extension is written as *Relation(Entity)*, where *Relation* is a label describing the relationship between the GO term and the entity, and *Entity* is an identifier for a database object or ontology term, for example *part\_of(GO:0005634)*, where GO:0005634 is the Gene Ontology identifier for “nucleus”.

Relations can be one of two types: “molecular relations” that are used with entities such as a gene, gene product, complex, or chemical and “contextual relations” that are used with entities such as a cell type, anatomy term, developmental stage, or a GO term.

In order to clearly define the semantics of the extensions, rules have been implemented defining what types of entity identifiers may be used with each relation. Generally, curators may only use contextual relations (e.g., where and when) with terms from the Cell Type Ontology (CL) [6], Uber Anatomy Ontology (Uberon) [7], Plant Ontology (PO) [8], nematode life stages (WBs) [9] and certain GO terms, and molecular target relations may only apply to a physical entity such as a gene product (e.g., UniProtKB [10] or PomBase [11]), a macromolecular complex (e.g., Intact Complex Portal [12]), or a chemical using a ChEBI [13] identifier. Curation tools can incorporate these rules to prevent invalid annotations from being created. Table 1 shows the most commonly used relations with examples of their usage.

**Table 1**  
**Most commonly used relationships for annotation extension statements and examples of their usage**

Contextual relationships	Example (gene product; primary GO term; annotation extension)
part_of	<i>C. elegans</i> psf-1; nucleus; part_of(WBbt:0006804 <i>body wall muscle cell</i> )
occurs_in	Mouse opsin-4; G-protein coupled photoreceptor activity; occurs_in(CL:0000740 <i>retinal ganglion cell</i> )
happens_during	<i>S. pombe</i> wis4; stress-activated MAPK cascade; happens_during(GO:0071470 <i>cellular response to osmotic stress</i> )
Molecular relationships	Example (gene product; primary GO term; annotation extension)
has_regulation_target	Human suppressor of fused homolog SUFU; negative regulation of transcription factor import into nucleus; has_regulation_target(UniProtKB:P08151 <i>zinc finger protein GLII</i> )
has_input	<i>S. pombe</i> rlf2; protein localization to nucleus; has_input(PomBase:SPAC26H5.0 <i>pcf2</i> )
has_direct_input	Human WNK4; chloride channel inhibitor activity; has_direct_input(UniProtKB:Q7LBE3 <i>Solute carrier family 26 member 9</i> )
Molecular relations take an entity such as a gene, gene product, complex, or chemical as an argument; contextual relations take an entity such as a cell type, anatomy term, development stage, or a GO term as an argument. Entity names in italics are shown for clarity and are not part of the annotation extension format.	

Reproduced from Huntley et al. [5]. Open access licence <http://creativecommons.org/licenses/by/4.0/>

## 4 Improved Expressiveness of GO Annotations: Examples

### 4.1 Targets of an Enzyme

One means of adding value to a GO annotation, using annotation extensions, is by specifying the molecular target of an enzyme activity. The inability to add effector–target relationships has been a major limitation of the core GO annotation model, with this addition we can now begin to provide directional information that can be used for network and pathway analyses. Take as an example the annotation of human mitogen-activated protein kinase-activated protein kinase 2 (MAPKAP-K2), which was shown to phosphorylate the CapZ-interacting protein (CapZIP) [14]. A basic GO annotation would describe MAPKAP-K2 as a protein serine/threonine kinase:

<b>Gene product:</b>	UniProtKB:P49137 ( <i>human MAPKAP-K2</i> )
<b>GO term:</b>	GO:0004674 ( <i>protein serine/threonine kinase activity</i> )

Using an annotation extension, a curator can add more detail as follows:

<b>Gene product:</b>	UniProtKB:P49137 ( <i>human MAPKAP-K2</i> )
<b>GO term:</b>	GO:0004674 ( <i>protein serine/threonine kinase activity</i> )
<b>Extension:</b>	has_direct_input(UniProtKB:Q6JBY9) ( <i>human CapZIP</i> )

N.B. phrases in italics are not part of the syntax but are added for better interpretation by the reader.

The extended GO annotation describes MAPKAP-K2 as a protein serine/threonine kinase that can phosphorylate CapZIP. This is vital information that can be utilized for linking together processes and pathways that MAPKAP-K2 and CapZIP, and any further targets of these proteins, are involved in. The rules of usage for `has_direct_input` are that the primary GO term used should be a Biological Process or Molecular Function and in this example the term used is a Molecular Function, additionally the entity used in the extension should be a gene product, macromolecular complex, or chemical and in this example it is a gene product, i.e., a protein. Note that `has_direct_input` was used here instead of `has_input` because there was evidence in the paper that MAPKAP-K2 acted directly on the substrate CapZIP, if there was a possibility of an intermediate molecule in this reaction, `has_input` would have been used.

#### 4.2 Anatomical Location of a Gene Product's Function

An annotation can be extended to specify the locational context in which a gene product performs its roles. It is important to note that we intend only to capture those locations that are physiologically relevant to the organism and not the experimental detail in which the observation was made.

The rat protein dihydrofolate reductase (*Dhfr*) was shown to reduce dihydrofolic acid to tetrahydrofolic acid in rat neurons [15]. From this evidence a basic GO annotation could be made as follows:

<b>Gene product:</b>	UniProtKB:Q920D2 ( <i>rat Dhfr</i> )
<b>GO term:</b>	GO:0004146 ( <i>dihydrofolate reductase activity</i> )

By extending the annotation the curator can also specify in which cell type this activity occurs:

<b>Gene product:</b>	UniProtKB:Q920D2 ( <i>rat Dhfr</i> )
<b>GO term:</b>	GO:0004146 ( <i>dihydrofolate reductase activity</i> )
<b>Extension:</b>	<code>occurs_in(CL:0000540)</code> ( <i>neuron</i> )

This annotation now provides the physiologically relevant information that *Dhfr* is active in neurons. The rules for `occurs_in` are that the primary GO term used must be a Biological Process or Molecular Function (in this example it is a Molecular Function); additionally the entity in the extension must be a cell type, anatomical feature, or GO Cellular Component (in this example it is an identifier from the Cell Type Ontology).

#### 4.3 Timing-Specific Location of a Gene Product

A gene product's annotation may be made more specific by including the appropriate developmental stage. An example is the location of the *C. elegans* PAXT-1 protein, which is located in the nucleus during the embryo stage [16]. Using the basic GO annotation format, a curator might indicate that PAXT-1 is located in the nucleus:

<b>Gene product:</b>	UniProtKB:Q21738 ( <i>C. elegans</i> PAXT-1)
<b>GO term:</b>	GO:0005634 ( <i>nucleus</i> )

By extending the annotation the curator can also specify when this localization occurs:

<b>Gene product:</b>	UniProtKB:Q21738 ( <i>C. elegans</i> PAXT-1)
<b>GO term:</b>	GO:0005634 ( <i>nucleus</i> )
<b>Extension:</b>	exists_during(WBIs:0000003) ( <i>embryo</i> )

This annotation means that PAXT-1 is located in the nucleus during the *C. elegans* embryo stage. The rules for exists\_during are that the primary GO term used should be a Cellular Component and the entity in the extension should be a developmental stage or a GO Biological Process, in this case the entity is from the *C. elegans* life stage ontology.

#### 4.4 Multiple Relational Expressions

If several contextual statements can be made for the gene product, it is possible to combine relational expressions to make even more complex statements. Relational expressions can be separated by commas “,” (meaning AND) or by pipes, “|” (meaning OR), depending on whether the conditions in the statement are co-occurring (AND) or independent (OR).

The human microRNA miR-145 provides an example of the application of multiple annotation extensions. MiR-145 was shown to directly bind and silence the POU5F1 transcription factor, among others, causing inhibition of embryonic stem cell division [17]. This evidence could therefore be represented by two basic GO annotations as follows:

<b>Gene product:</b>	RNACentral:URS0000527F89_9606 ( <i>human miR-145</i> )
<b>GO term:</b>	GO:1903231 ( <i>mRNA binding involved in posttranscriptional gene silencing</i> )
<b>Gene product:</b>	RNACentral:URS0000527F89_9606 ( <i>human miR-145</i> )
<b>GO term:</b>	GO:1904676 ( <i>negative regulation of somatic stem cell division</i> )

Using relational expressions, separated by commas, we can make one extended annotation as follows:

<b>Gene product:</b>	RNACentral:URS0000527F89_9606 ( <i>human miR-145</i> )
<b>GO term:</b>	GO:1903231 ( <i>mRNA binding involved in posttranscriptional gene silencing</i> )
<b>Extension:</b>	has_direct_input(Ensembl:ENSG00000204531), occurs_in(CL:0002322), part_of(GO:1904676) ( <i>human POU5F1, embryonic stem cell, negative regulation of somatic stem cell division</i> )

The extended annotation signifies that miR-145 directly binds and silences POU5F1 mRNA expression as part of the inhibition of somatic stem cell division of embryonic stem cells. Again, this contextual information will be essential information when analyzing the physiological relevance of the role of a gene product in a pathway.

Although the use of a pipe (|) to indicate independent contextual statements does not provide any additional expressivity to the statements already made, it allows a curator to capture several statements from the same evidence within a paper. An example is when specifying the multiple substrates of an enzyme—the enzyme may act on each of the substrates independently, but not all at the same time; therefore, the substrates can be listed in the extension separated by pipe symbols:

<b>Gene product:</b>	UniProtKB:O14522 ( <i>human PTPRT</i> )
<b>GO term:</b>	GO:0004725 ( <i>protein tyrosine phosphatase activity</i> )
<b>Extension:</b>	has_direct_input(UniProtKB:P12830)   has_direct_input(UniProtKB:O60716) ( <i>E-cadherin</i>   <i>CTNND1</i> )

This annotation indicates that the receptor protein tyrosine phosphatase rho (PTPRT) dephosphorylates E-cadherin and CTNND1, but not necessarily both simultaneously. It would be equally correct to create two separate annotations each with a single substrate in the extension.

---

## 5 Practical Use of Extended Annotations

There are likely to be many use cases for extended annotations—even some we have not yet envisioned. Users will be able to perform more advanced queries with the available functional data; such as filtering on the subcellular, cellular or anatomical locations in which a gene product performs its roles, or which genes a transcription factor regulates in a specified cell type. Annotation extensions can also help create functional networks through the use of directional relationships such as `has_input` and `has_direct_input`, which allow specification of the target of an effector, for example in a signaling pathway or the substrates of a metabolic enzyme activity.

Without contextual detail, bioinformatics analyses of gene products involved in a specified process cannot distinguish, for example, between those gene products that are active only in a particular cell type and those that are inactive or absent from that cell type, therefore creating a bias in the interpretation of the data. With extended annotations any differences in the active components of a process or pathway between various cell or tissue types can be determined.

## 5.1 Access

Extended annotations are available for download in the current GO annotation files, both in the GAF2.0 format (column 16; [http://www.geneontology.org/GO.format.gaf-2\\_0.shtml](http://www.geneontology.org/GO.format.gaf-2_0.shtml)) and in the Gene Product Association Data format (GPAD column 11; <http://www.geneontology.org/GO.format.gpad.shtml>). These files can be accessed from the GOC website (<http://geneontology.org/GO.downloads.annotations.shtml>) and the GOA website (<http://www.ebi.ac.uk/GOA/downloads>).

Extended annotations can be accessed on the web via the GO browsers QuickGO ([18]; [www.ebi.ac.uk/QuickGO-Beta](http://www.ebi.ac.uk/QuickGO-Beta)) and AmiGO 2 ([19]; <http://amigo.geneontology.org/amigo/>). Both browsers allow users to filter annotation sets based on the contents of the annotation extension. The display of extended annotations may be different depending on the resource (Fig. 1), but the GO annotation files display the plain text extension since this is more compatible for computational analysis (*see* also Chap. 11 [20]). Any questions on how to access or use extended annotations should be directed to the GOC helpdesk (<http://geneontology.org/form/contact-go>).

## 5.2 Exercise

The addition of extended annotations to Gene Ontology datasets enables users to perform sophisticated queries. This exercise will demonstrate how to build such a query in the GOC browser

**a**

Gene Product ID	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Reference	With	Taxon	Assigned By	Annotation Extension
Q5S007	LRRK2	enables	GO:000149	SNARE binding	F	ECO:0000353 (IPI)	PMID:21307259	UniProtKB:P46460	9606	ParkinsonsUK-UCL	occurs_in UBERON:0000955...
Q5S007	LRRK2	enables	GO:0001948	glycoprotein binding	F	ECO:0000353 (IPI)	PMID:21307259	UniProtKB:Q9J155	9606	ParkinsonsUK-UCL	occurs_in UBERON:0000955...

**b**

Gene/product	Gene/product name	Qualifier	Direct annotation	Annotation extension	Assigned by	Taxon	Evidence	Evidence with	PANTHER family	Isoform	Reference	Date
<input type="checkbox"/>	LRRK2		Leucine-rich repeat serine/threonine-protein kinase 2	SNARE binding	occurs in brain	ParkinsonsUK-UCL	Homo sapiens	IPI	UniProtKB:P46460	leucine-rich repeat-containing protein pthr23155	PMID:21307259	20140513
<input type="checkbox"/>	LRRK2		Leucine-rich repeat serine/threonine-protein kinase 2	glycoprotein binding	occurs in brain	ParkinsonsUK-UCL	Homo sapiens	IPI	UniProtKB:Q9J155	leucine-rich repeat-containing protein pthr23155	PMID:21307259	20140513

**c**

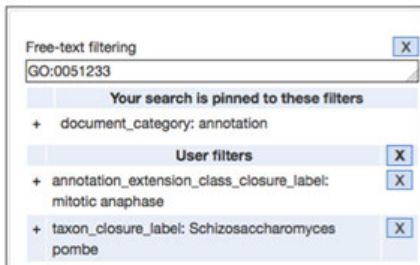
**Fig. 1** Display of extended annotations in (a) the beta version of the EBI GO browser QuickGO, (b) AmiGO 2, and (c) PomBase (<http://www.pombase.org/>)



AmiGO 2, namely, to provide all of the gene products from *S. pombe* that are located in the spindle midzone during mitotic anaphase.

1. Open the AmiGO 2 browser (<http://amigo.geneontology.org/amigo/>).
2. Click on the Advanced Search button and select “Annotations” from the drop-down list.
3. In the free-text filtering box on the left (Fig. 2a) type in GO:0051233, the GO identifier for the Cellular Component term “spindle midzone”.
4. Now open the Taxon menu on the left and click on the “more” button at the bottom. A pop-up menu will open, in the top filter box start typing “pombe”--“Schizosaccharomyces pombe” should be the only option that appears. Click on the + next to the species name to add this to the filter.
5. Now open the Annotation Extension menu on the left and click on the + button next to the term “mitotic anaphase” to add this to the filter.
6. AmiGO 2 will display all of the annotations that use the “mitotic anaphase” term (or one of its child terms) in the annotation extension of a primary annotation to “spindle midzone” (or one of its child terms) (Fig. 2b).

**a**



**b**

Found entities

Total: 7; showing 1-7 Results count 10

<input type="checkbox"/>	Gene/product	Gene/product name	Qualifier	Direct annotation	Annotation extension	Assigned by	Taxon	Evidence	Evidence with	PANTHER family	Isoform	Reference	Date
<input type="checkbox"/>	ark1	aurora-B kinase Ark1		spindle midzone	exists_during mitotic anaphase	PomBase	Schizosaccharomyces pombe	IDA				PMID:11792803	20041203
<input type="checkbox"/>	kip9	kinesin-like protein Kip9		spindle midzone	exists_during mitotic anaphase B	PomBase	Schizosaccharomyces pombe	IDA				PMID:19686686	20090820
<input type="checkbox"/>	peg1	CLASP family microtubule-associated protein		spindle midzone	exists_during mitotic anaphase	PomBase	Schizosaccharomyces pombe	IDA				PMID:16951255	20070205
<input type="checkbox"/>	bir1	survivin, Bir1		spindle midzone	exists_during mitotic anaphase A	PomBase	Schizosaccharomyces pombe	IDA				PMID:11861551	20041203
<input type="checkbox"/>	bir1	survivin, Bir1		spindle midzone	exists_during mitotic anaphase B	PomBase	Schizosaccharomyces pombe	IDA				PMID:16824200	20140513
<input type="checkbox"/>	nb11	Borealin homolog Nbl1		spindle midzone	exists_during mitotic anaphase	PomBase	Schizosaccharomyces pombe	IDA				PMID:19570910	20090707
<input type="checkbox"/>	cut7	kinesin-like protein Cut7		mitotic spindle midzone	exists_during mitotic anaphase B	PomBase	Schizosaccharomyces pombe	IDA				PMID:1538784	20130125

**Fig. 2** Finding annotations in AmiGO 2 based on annotation extension data. **(a)** Filters applied in the AmiGO 2 browser: GO:ID (GO:0051233 “spindle midzone”), annotation extension (mitotic anaphase), taxon (Schizosaccharomyces pombe). **(b)** Results of the search using the filters applied in **(a)**. Six unique gene products are located to the spindle midzone during mitotic anaphase

## 6 Summary

The Gene Ontology has proven a vital resource for researchers, enabling them to easily find and use functional data. GO is continually evolving to reflect both accumulating biological knowledge and the computational techniques that researchers need for analysis of a list of gene products. One of the major limitations of using the original simple GO functional annotation has been the lack of contextual information linking together gene products and the roles and pathways they are involved in [4]. Inclusion of this type of data within GO annotations can advance pathway and network analyses substantially, allowing more sophisticated queries and analyses to be performed.

As with all other aspects of GO, annotation extensions continue to evolve—through discussion involving all GOC members and the community—to allow representation and ultimately simple access to a wide variety of contextual data.

**Funding** Open Access charges were funded by the University College London Library, the Swiss Institute of Bioinformatics, the Agassiz Foundation, and the Foundation for the University of Lausanne.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

## References

1. Khatri P, Drăghici S (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* 21:3587–3595. doi:[10.1093/bioinformatics/bti565](https://doi.org/10.1093/bioinformatics/bti565)
2. Schmidt A, Forne I, Imhof A (2014) Bioinformatic analysis of proteomics data. *BMC Syst Biol* 8(Suppl 2):S3. doi:[10.1186/1752-0509-8-S2-S3](https://doi.org/10.1186/1752-0509-8-S2-S3)
3. Bauer S (2016) Gene-category analysis. In: Dessimoz C, Škunca N (eds) *The gene ontology handbook. Methods in molecular biology*, vol 1446. Humana Press. Chapter 13
4. Khatri P, Sirota M, Butte AJ (2012) Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 8:e1002375. doi:[10.1371/journal.pcbi.1002375](https://doi.org/10.1371/journal.pcbi.1002375)

5. Huntley RP, Harris MA, Alam-Faruque Y et al (2014) A method for increasing expressivity of Gene Ontology annotations using a compositional approach. *BMC Bioinformatics* 15:155. doi:[10.1186/1471-2105-15-155](https://doi.org/10.1186/1471-2105-15-155)
6. Meehan TF, Masci AM, Abdulla A et al (2011) Logical development of the cell ontology. *BMC Bioinformatics* 12:6. doi:[10.1186/1471-2105-12-6](https://doi.org/10.1186/1471-2105-12-6)
7. Mungall CJ, Torniai C, Gkoutos GV et al (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol* 13:R5. doi:[10.1186/gb-2012-13-1-r5](https://doi.org/10.1186/gb-2012-13-1-r5)
8. Avraham S, Tung C-W, Ilic K et al (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res* 36:D449–D454. doi:[10.1093/nar/gkm908](https://doi.org/10.1093/nar/gkm908)
9. Lee RYN, Sternberg PW (2003) Building a cell and anatomy ontology of *Caenorhabditis elegans*. *Comp Funct Genomics* 4:121–126. doi:[10.1002/cfg.248](https://doi.org/10.1002/cfg.248)
10. The UniProt Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Res* 43:D204–D212. doi:[10.1093/nar/gku989](https://doi.org/10.1093/nar/gku989)
11. McDowall MD, Harris MA, Lock A et al (2015) PomBase 2015: updates to the fission yeast database. *Nucleic Acids Res* 43:D656–D661. doi:[10.1093/nar/gku1040](https://doi.org/10.1093/nar/gku1040)
12. Meldal BHM, Forner-Martinez O, Costanzo MC et al (2014) The complex portal—an encyclopaedia of macromolecular complexes. *Nucleic Acids Res*. doi:[10.1093/nar/gku975](https://doi.org/10.1093/nar/gku975)
13. Hastings J, de Matos P, Dekker A et al (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res* 41:D456–D463. doi:[10.1093/nar/gks1146](https://doi.org/10.1093/nar/gks1146)
14. Evers CE, McNeill H, Knebel A et al (2005) The phosphorylation of CapZ-interacting protein (CapZIP) by stress-activated protein kinases triggers its dissociation from CapZ. *Biochem J* 389:127–135. doi:[10.1042/BJ20050387](https://doi.org/10.1042/BJ20050387)
15. Iskandar BJ, Rizk E, Meier B et al (2010) Folate regulation of axonal regeneration in the rodent central nervous system through DNA methylation. *J Clin Invest* 120:1603–1616. doi:[10.1172/JCI40000](https://doi.org/10.1172/JCI40000)
16. Gloerich M, ten Klooster JP, Vliem MJ et al (2012) Rap2A links intestinal cell polarity to brush border formation. *Nat Cell Biol* 14:793–801. doi:[10.1038/ncb2537](https://doi.org/10.1038/ncb2537)
17. Xu N, Papagiannakopoulos T, Pan G et al (2009) MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* 137:647–658. doi:[10.1016/j.cell.2009.02.038](https://doi.org/10.1016/j.cell.2009.02.038)
18. Binns D, Dimmer E, Huntley R et al (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25:3045–3046. doi:[10.1093/bioinformatics/btp536](https://doi.org/10.1093/bioinformatics/btp536)
19. The Gene Ontology Consortium (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res* 38:D331–D335. doi:[10.1093/nar/gkp1018](https://doi.org/10.1093/nar/gkp1018)
20. Munoz-Torres M, Carbon S (2016) Get GO! retrieving GO data using AmiGO, QuickGO, API, files, and tools. In: Dessimoz C, Škunca N (eds) *The gene ontology handbook. Methods in molecular biology*, vol 1446. Humana Press. Chapter 11