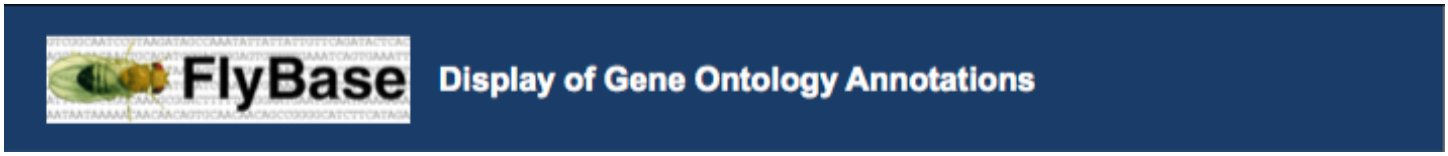


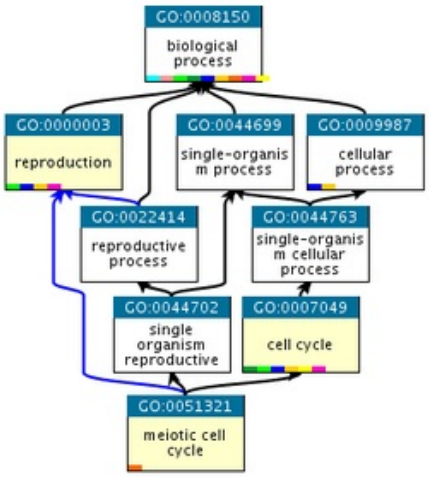
Survey questions

Display of Gene Ontology Annotations

Introduction

The Gene Ontology (GO) is a set of controlled terms used to describe the attributes of a gene product. The GO is divided into three aspects: Molecular Function (e.g. protein tyrosine kinase activity), Biological Process (e.g. phosphorylation) and Cellular Component (e.g. plasma membrane). The terms in the GO are structured in a hierarchy that formally describes the relationships between them. At its simplest level, for example, protein tyrosine kinase activity is a child of protein kinase activity, but more complex relationships exist. For example, meiotic cell cycle is related to the cell cycle and reproductive biological process branches of the GO (illustrated below).

FlyBase is examining ways to utilise the data we hold to help summarise a gene's function and to present this data in a more accessible manner. In this survey we are asking you to provide feedback on improving the presentation of Gene Ontology (GO) data.

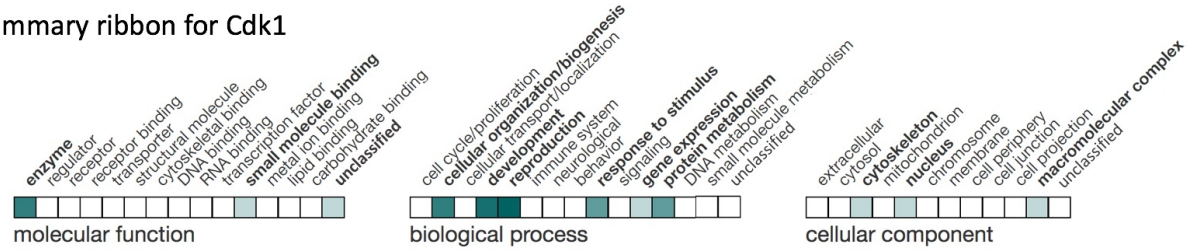


Display of Gene Ontology Annotations

GO summary ribbon displays in the gene report

In the new version of FlyBase, due for beta release later this year, we will introduce "ribbon" graphical summaries (developed by Mouse Genome Informatics, MGI) for the presentation of GO data. These summary ribbons use the hierarchical structure of the ontology to group terms under generalised, high-level categories. For example, the GO term "ATP binding" is grouped under "small molecule binding", as shown in the example for Cdk1 below.

GO summary ribbon for Cdk1



1. Where would you like to see the GO ribbons summaries displayed in the gene report?

Choose one option below.

- ○ In a GO summary ribbon section below the "Genome location" section (panel A)
- ○ In a GO summary ribbon section above the "Genome location" section (panel B)
- ○ In a GO summary ribbon section below the "Genome location" section and repeated in the corresponding GO section (panel A + C)
- ○ In a GO summary ribbon section above the "Genome location" section and repeated in the corresponding GO section (panel B + C)

A

General Information

Symbol	DmelCdk1	**Species**	*D. melanogaster*
Name	Cyclin-dependent kinase 1	**Annotation symbol**	CG5363
Feature type	protein_coding_gene	**FlyBase ID**	FBgn0004106
Gene Model Status	Current	**Stock availability**	24 publicly available
Also Known As	Cdc2, DmCdc2		
Gene Snapshot	Cyclin-dependent kinase (Cdk1) is a catalytic protein kinase subunit that can only become active after association with either CycA, CycB or CycB3. The protein kinase activities of these complexes (CycA-Cdk1, CycB-Cdk1, CycB3-Cdk1) control important aspects of progression through the cell cycle. Functionally, the different Cdk1 complexes are partially redundant. They phosphorylate hundreds of target proteins and are most important for progression into and through mitotic and meiotic M phases. [Date last reviewed: 2016-06-23]		

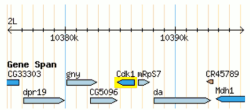
Genomic Location

Cytogenetic map	31D11-31D11	**Sequence location**	2L:10,384,739..10,386,262 [-]

Genomic Maps

GBrowse

JBrowse



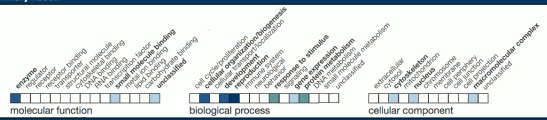
Decorated FASTA
Get genome region
Gene region
Get FastA

Other Genome Views

The following external sites may use different assemblies or annotations than FlyBase.

NCBI Genome Data Viewer UCSC Genome Browser

GO Summary ribbon



Families, Domains and Molecular Function

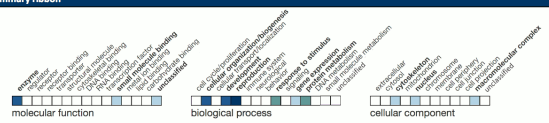
Gene Group Membership (FlyBase)	CYCLIN DEPENDENT KINASES
Protein Family (UniProt, Sequence Similarities)	Belongs to the protein kinase superfamily. CMGC Ser/Thr protein kinase family. CDC2/CDKX subfamily. (P23572)
Protein Domains/Motifs	**UniProt (Sequence Similarities)** Contains 1 protein kinase domain. (P23572) **InterPro** Protein kinase domain; Serine/threonine-protein kinase, active site; Protein kinase-like domain; Protein kinase, ATP binding site

B

General Information

Symbol	DmelCdk1	**Species**	*D. melanogaster*
Name	Cyclin-dependent kinase 1	**Annotation symbol**	CG5363
Feature type	protein_coding_gene	**FlyBase ID**	FBgn0004106
Gene Model Status	Current	**Stock availability**	24 publicly available
Also Known As	Cdc2, DmCdc2		
Gene Snapshot	Cyclin-dependent kinase (Cdk1) is a catalytic protein kinase subunit that can only become active after association with either CycA, CycB or CycB3. The protein kinase activities of these complexes (CycA-Cdk1, CycB-Cdk1, CycB3-Cdk1) control important aspects of progression through the cell cycle. Functionally, the different Cdk1 complexes are partially redundant. They phosphorylate hundreds of target proteins and are most important for progression into and through mitotic and meiotic M phases. [Date last reviewed: 2016-06-23]		

GO Summary ribbon



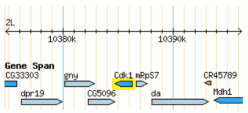
Genomic Location

Cytogenetic map	31D11-31D11	**Sequence location**	2L:10,384,739..10,386,262 [-]

Genomic Maps

GBrowse

JBrowse



Decorated FASTA
Get genome region
Gene region
Get FastA

Other Genome Views

The following external sites may use different assemblies or annotations than FlyBase.

NCBI Genome Data Viewer UCSC Genome Browser

Families, Domains and Molecular Function

Gene Group Membership (FlyBase)	CYCLIN DEPENDENT KINASES
Protein Family (UniProt, Sequence Similarities)	Belongs to the protein kinase superfamily. CMGC Ser/Thr protein kinase family. CDC2/CDKX subfamily. (P23572)
Protein Domains/Motifs	**UniProt (Sequence Similarities)** Contains 1 protein kinase domain. (P23572) **InterPro** Protein kinase domain; Serine/threonine-protein kinase, active site; Protein kinase-like domain; Protein kinase, ATP binding site

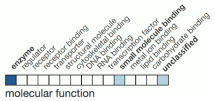
C

Families, Domains and Molecular Function

Gene Group Membership (FlyBase)	CYCLIN DEPENDENT KINASES
Protein Family (UniProt, Sequence Similarities)	Belongs to the protein kinase superfamily. CMGC Ser/Thr protein kinase family. CDC2/CDKX subfamily. (P23572)
Protein Domains/Motifs	**UniProt (Sequence Similarities)** Contains 1 protein kinase domain. (P23572) **InterPro** Protein kinase domain; Serine/threonine-protein kinase, active site; Protein kinase-like domain; Protein kinase, ATP binding site

Gene Ontology (GO): (23 terms)

Molecular Function (5 terms)



Terms Based on Experimental Evidence (3 terms)

CV term	Evidence	References
cyclin-dependent protein serine/threonine kinase activity	inferred from genetic interaction with Saccharomyces CDC28	(*Lehner and O'Farrell, 1990*)
protein binding	inferred from physical interaction with CycB3 (assigned by UniProt)	(*Jacobs et al., 1998*)
protein kinase activity	inferred from direct assay	(*Ayeni et al., 2014*)

Terms Based on Predictions or Assertions (3 terms)

CV term	Evidence	References
ATP binding	inferred from electronic annotation with InterPro:IPR000719, InterPro:IPR002290, InterPro:IPR017441	(*FlyBase Curators et al., 2004-*)
cyclin-dependent protein serine/threonine kinase activity	inferred from sequence or structural similarity with Saccharomyces CDC28	(*Lehner and O'Farrell, 1990*)
protein serine/threonine kinase activity	non-traceable author statement	(*Morrison et al., 2000*)

Display of Gene Ontology Annotations

Display of GO annotations in the gene report

In the gene report the GO annotations are split into the three aspects: Molecular Function, Biological Process and Cellular Component. These are further subdivided into annotations that have been inferred from experimental observations and those that have been inferred from predictions or assertions made by curators or automated pipelines. The following questions address the display of data in this section.

2. The GO terms are displayed in alphabetical order and not based on the hierarchy of the ontology. What changes can we make to the ordering of terms that will aid understanding?
The option examples illustrated below are based on experimentally inferred biological process annotations for Cdk1.

Choose one option.

- ◯ Do not change, I find the alphabetical listing easy to understand (panel A)
- ◯ Sort the GO data using the ribbon categories, using **only** populated ribbon categories as section headers (panel B)
- ◯ Sort the GO data using the ribbon categories, using **all** ribbon categories as section headers, indicating how many terms are under each category (panel C)
- ◯ Sort the GO data using the hierarchical structure of the ontology, but do not split into categories (panel D)
- ◯ None of the above - own comment

A

Biological Process (17 terms)

Terms Based on Experimental Evidence (14 terms)

CV term	Evidence	References
asymmetric neuroblast division	inferred from mutant phenotype	(*Tio et al., 2001*)
cellular response to DNA damage stimulus	inferred from mutant phenotype	(*Ravi et al., 2009*)
G1/S transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(*Lehner and O'Farrell, 1990*)
G2/M transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(*Lehner and O'Farrell, 1990*)
	inferred from mutant phenotype	(*Stern et al., 1993*)
germarium-derived cystoblast division	inferred from mutant phenotype	(*Jin et al., 2005*)
male meiosis	inferred from mutant phenotype	(*Jin et al., 2005*)
mitotic cell cycle	inferred from mutant phenotype	(*Kiger et al., 2003*)
mitotic G2 DNA damage checkpoint	inferred from mutant phenotype	(*Ayeni et al., 2014*)
mitotic G2/M transition checkpoint	inferred from mutant phenotype	(*Ayeni et al., 2014*)
neurogenesis	inferred from mutant phenotype	(*Neumüller et al., 2011*)
ovarian follicle cell development	inferred from mutant phenotype	(*Jia et al., 2015*)
protein phosphorylation	inferred from direct assay	(*Du et al., 1996*)

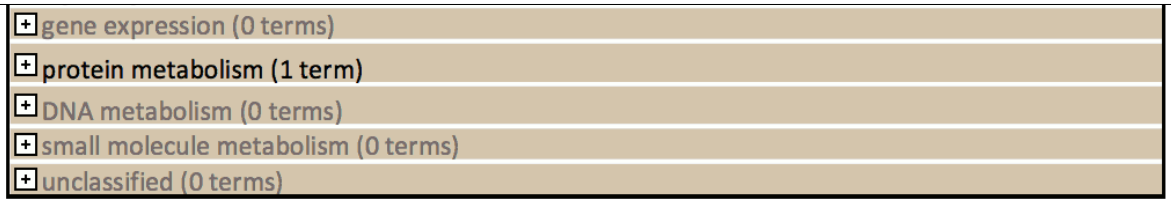
regulation of protein localization	inferred from mutant phenotype	(*Royou et al., 2002*)
spermatogonial cell division	inferred from mutant phenotype	(*Jin et al., 2005*)
Terms Based on Predictions or Assertions (5 terms)		
CV term	Evidence	References

B

Biological Process (17 terms)		
Terms Based on Experimental Evidence (14 terms)		
cell cycle/proliferation		
G1/S transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(Lehner and O'Farrell, 1990)
G2/M transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(Lehner and O'Farrell, 1990)
	inferred from mutant phenotype	(Stern et al., 1993)
asymmetric neuroblast division	inferred from mutant phenotype	(Tio et al., 2001)
germarium-derived cystoblast division	inferred from mutant phenotype	(Jin et al, 2005)
male meiosis	inferred from mutant phenotype	(Jin et al, 2005)
mitotic G2 DNA damage checkpoint	inferred from mutant phenotype	(Ayeni et al., 2014)
mitotic G2/M transition checkpoint	inferred from mutant phenotype	(Ayeni et al., 2014)
mitotic cell cycle	inferred from mutant phenotype	(Kiger et al, 2003)
cellular organization/biogenesis		
male meiosis	inferred from mutant phenotype	(Jin et al, 2005)
cellular transport/localization		
regulation of protein localization	inferred from mutant phenotype	(Royou et al, 2002)
developmental process		
germarium-derived cystoblast division	inferred from mutant phenotype	(Jin et al, 2005)
neurogenesis	inferred from mutant phenotype	(Neumüller et al, 2011)
ovarian follicle cell development	inferred from mutant phenotype	(Jia et al, 2015)
reproduction		
germarium-derived cystoblast division	inferred from mutant phenotype	(Jin et al, 2005)
male meiosis	inferred from mutant phenotype	(Jin et al, 2005)
ovarian follicle cell development	inferred from mutant phenotype	(Jia et al, 2015)
spermatogonial cell division	inferred from mutant phenotype	(Jia et al, 2015)
response to stimulus		
cellular response to DNA damage stimulus	inferred from mutant phenotype	(Ravi et al, 2009)
mitotic G2 DNA damage checkpoint	inferred from mutant phenotype	(Ayeni et al., 2014)
protein metabolic process		
protein phosphorylation	inferred from direct assay	(Du et al, 1996)

C

Biological Process (17 terms)		
Terms Based on Experimental Evidence (14 terms)		
⊟ cell cycle/proliferation (8 terms)		
G1/S transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(Lehner and O'Farrell, 1990)
G2/M transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(Lehner and O'Farrell, 1990)
	inferred from mutant phenotype	(Stern et al., 1993)
asymmetric neuroblast division	inferred from mutant phenotype	(Tio et al., 2001)
germarium-derived cystoblast division	inferred from mutant phenotype	(Jin et al, 2005)
male meiosis	inferred from mutant phenotype	(Jin et al, 2005)
mitotic G2 DNA damage checkpoint	inferred from mutant phenotype	(Ayeni et al., 2014)
mitotic G2/M transition checkpoint	inferred from mutant phenotype	(Ayeni et al., 2014)
mitotic cell cycle	inferred from mutant phenotype	(Kiger et al, 2003)
⊞ cellular organization/biogenesis (1 term)		
⊞ cellular transport/localization (1 term)		
⊞ developmental process (3 terms)		
⊞ reproduction (4 terms)		
⊞ immunity (0 terms)		
⊞ neurological (0 terms)		
⊞ response to stimulus (0 terms)		
⊞ signaling (0 terms)		



D

Biological Process (17 terms)

Terms Based on Experimental Evidence (14 terms)

CV term	Evidence	References
mitotic cell cycle	inferred from mutant phenotype	(Kiger et al, 2003)
G2/M transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(Lehner and O'Farrell, 1990)
	inferred from mutant phenotype	(Stern et al., 1993)
mitotic G2 DNA damage checkpoint	inferred from mutant phenotype	(Ayeni et al., 2014)
mitotic G2/M transition checkpoint	inferred from mutant phenotype	(Ayeni et al., 2014)
G1/S transition of mitotic cell cycle	inferred from genetic interaction with Saccharomyces CDC28	(Lehner and O'Farrell, 1990)
male meiosis	inferred from mutant phenotype	(Jin et al, 2005)
spermatogonial cell division	inferred from mutant phenotype	(Jia et al, 2015)
germarium-derived cystoblast division	inferred from mutant phenotype	(Jin et al, 2005)
ovarian follicle cell development	inferred from mutant phenotype	(Jia et al, 2015)
cellular response to DNA damage stimulus	inferred from mutant phenotype	(Ravi et al, 2009)
asymmetric neuroblast division	inferred from mutant phenotype	(Tio et al., 2001)
neurogenesis	inferred from mutant phenotype	(Neumüller et al, 2011)
regulation of protein localization	inferred from mutant phenotype	(Royou et al, 2002)
protein phosphorylation	inferred from direct assay	(Du et al, 1996)

3. Should we continue to display GO annotations as two separate sections - those inferred from experimental data and those based on predictions/assertions, as in the example for Cdk1 molecular function below? Or would you prefer to have GO annotations from both experimental and non-experimental evidence displayed together?

Choose one option.

- ◯ Continue to separate based on evidence, as this is useful to me
- ◯ Do not separate based on evidence, I find the evidence statement sufficient to differentiate
- ◯ Do not separate based on evidence, but highlight where the evidence is experimental (e.g. by an icon)
- ◯ Don't know

Gene Ontology (GO): (23 terms)		
⊟ Molecular Function (5 terms)		
Terms Based on Experimental Evidence (3 terms)		
CV term	Evidence	References
cyclin-dependent protein serine/threonine kinase activity	inferred from genetic interaction with Saccharomyces CDC28	(*Lehner and O'Farrell, 1990*)
protein binding	inferred from physical interaction with CycB3 (assigned by UniProt)	(*Jacobs et al., 1998*)
protein kinase activity	inferred from direct assay	(*Ayeni et al., 2014*)
Terms Based on Predictions or Assertions (3 terms)		
CV term	Evidence	References
ATP binding	inferred from electronic annotation with InterPro:IPR000719, InterPro:IPR002290, InterPro:IPR017441	(*FlyBase Curators et al., 2004-*)
cyclin-dependent protein serine/threonine kinase activity	inferred from sequence or structural similarity with Saccharomyces CDC28	(*Lehner and O'Farrell, 1990*)
protein serine/threonine kinase activity	non-traceable author statement	(*Morrison et al., 2000*)

Future developments for ribbon summaries

FlyBase would like to introduce ribbon summaries for other data types. The questions below will help direct future development.

4. In order of preference, rank the following data types you would like to see summarized in ribbon displays, where 1 is most valuable and 3 is the least.

- Expression
- Phenotype
- Human Disease connection*

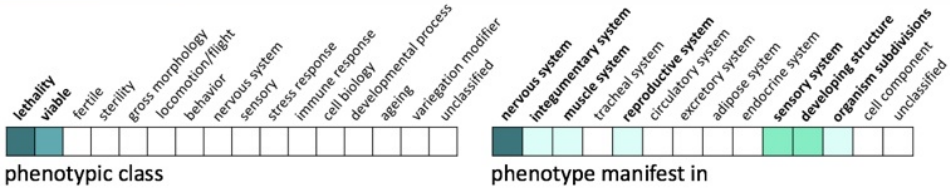
*for example, summarizing the data on fly genes used to model human disease and/or fly orthologs of human genes linked to disease

5. The allele reports contain associated phenotype data. Along with gene-level ribbon summaries in gene reports, should we also build ribbons specific for alleles to be included in the individual allele reports? An example of how phenotype data could be displayed for an allele of Cdk1 is shown below.

Choose one option.

- ( ) Yes, a ribbon display of this data in allele reports would be useful to me
- ( ) No. The summaries should be only on the gene reports
- ( ) Don't know

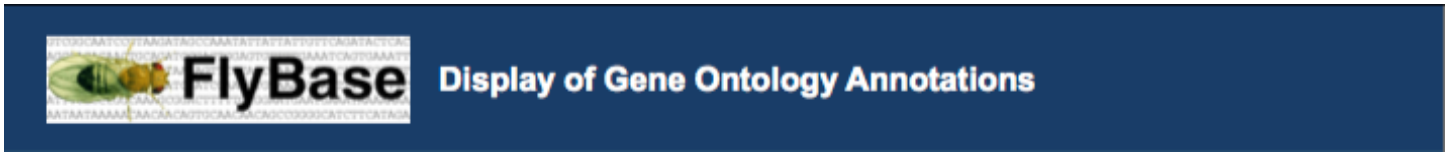
General Information			
Symbol	Dmel\Cdk1^{B47}	**Species**	*D. melanogaster*
Name		**FlyBase ID**	FBal0030731
Feature type	allele	**Associated gene**	Dmel\Cdk1
Associated Insertion(s)		**Carried in Construct**	
Also Known As	cdc2^{B47}, Dmcdc2^{B47}		



Phenotypic Data	
Phenotypic Class	
lethal (with Cdk1$^{GT-000294}$)	(*Roote, 2004.11.9*)
lethal | recessive	(*Stern et al., 1993*, *Tio et al., 2001*)
lethal | recessive | heat sensitive	(*Hayashi, 1996*)
lethal - all die before end of pupal stage | recessive	(*Clegg et al., 1993*)
lethal - all die during pupal stage | heat sensitive (with Cdk1^{E1-24})	(*Ayeni et al., 2014*)
some die during pupal stage	(*Clegg et al., 1993*)
Phenotype Manifest In	
A1-7 dorsal acute muscle 1 | ectopic, with Cdk1^{A171T}	(*Tio et al., 2001*)
central nervous system	(*Hayashi, 1996*)
embryonic/larval brain	(*Clegg et al., 1993*)
embryonic/larval optic lobe | heat sensitive (with Cdk1^{E1-24})	(*Hayashi and Yamaguchi, 1999*)
embryonic/larval salivary gland | embryonic stage	(*Hayashi, 1996*)
embryonic neuroblast, with Cdk1^{A171T}	(*Tio et al., 2001*)
histoblast	(*Hayashi, 1996*)
histoblast & nucleus | conditional ts (with Cdk1^{E1-24})	(*Hayashi and Yamaguchi, 1999*)
imaginal disc	(*Stern et al., 1993*, *Hayashi, 1996*)
imaginal disc | heat sensitive (with Cdk1^{E1-24})	(*Hayashi and Yamaguchi, 1999*)
imaginal disc | larval stage	(*Clegg et al., 1993*)
oocyte | temperature conditional (with Cdk1^{E1-24})	(*Von Stetina et al., 2008*)
optic lobe	(*Hayashi, 1996*)
RP2 motor neuron | ectopic, with Cdk1^{A171T}	(*Tio et al., 2001*)
tormogen cell | heat sensitive (with Cdk1^{E1-24})	(*Fichelson and Gho, 2004*)
trichogen cell | heat sensitive (with Cdk1^{E1-24})	(*Fichelson and Gho, 2004*)
wing hair | ectopic | somatic clone	(*Adler et al., 2000*)
wing hair | somatic clone	(*Adler et al., 2000*)

Comments and Suggestions

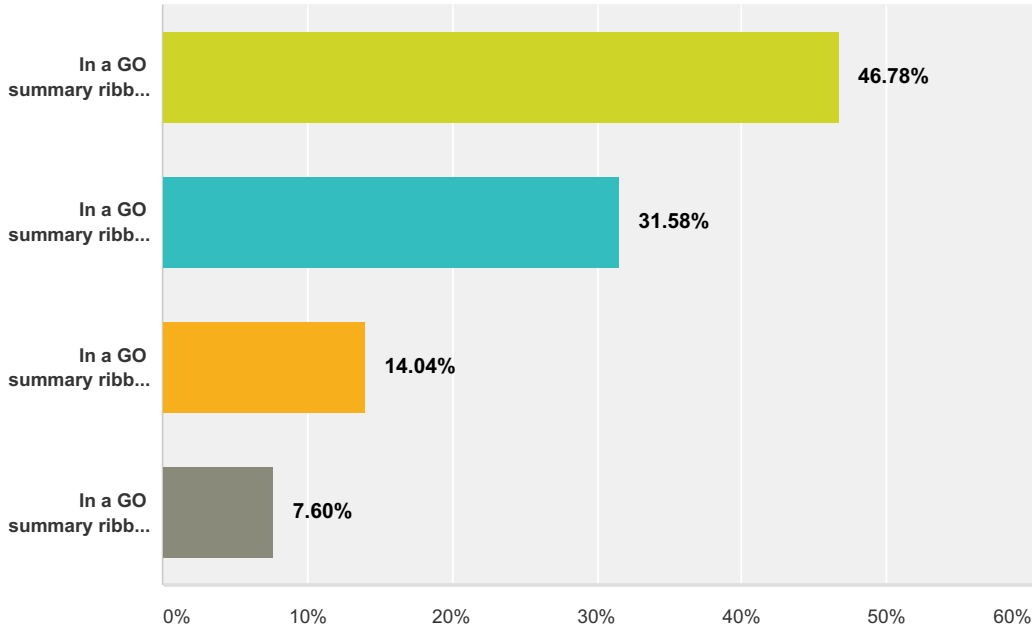
6. Thank you for taking the time to complete this survey. If you have any comments or suggestions about ribbon summaries or the display of GO data, please add them below.



Survey answers

Q1 Where would you like to see the GO ribbons summaries displayed in the gene report? Choose one option below.

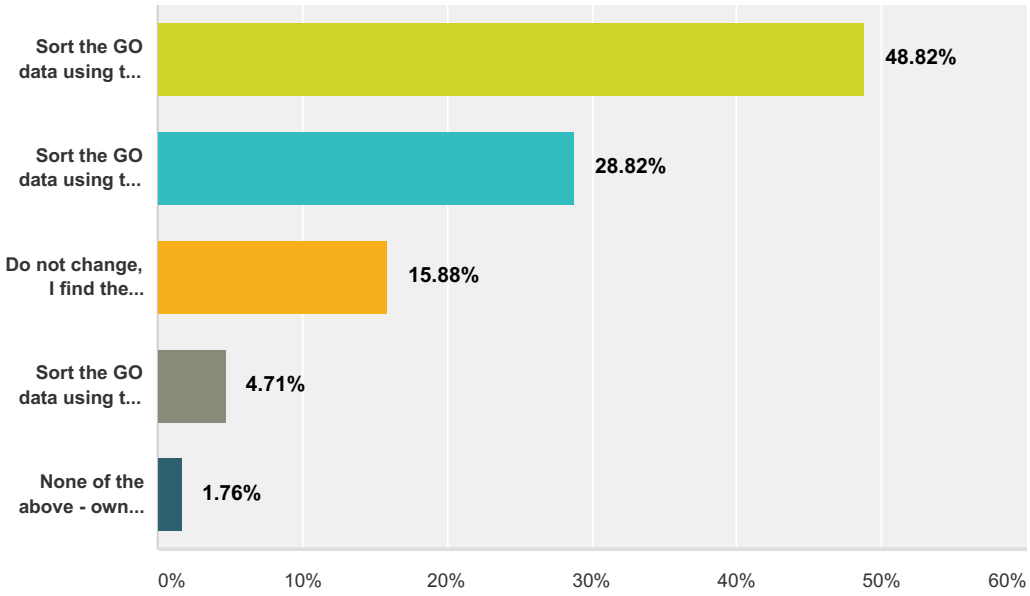
Answered: 171 Skipped: 1



Answer Choices	Responses	
In a GO summary ribbon section below the "Genome location" section (panel A)	46.78%	80
In a GO summary ribbon section below the "Genome location" section and repeated in the corresponding GO section (panel A + C)	31.58%	54
In a GO summary ribbon section above the "Genome location" section (panel B)	14.04%	24
In a GO summary ribbon section above the "Genome location" section and repeated in the corresponding GO section (panel B + C)	7.60%	13
Total		**171**

Q2 The GO terms are displayed in alphabetical order and not based on the hierarchy of the ontology. What changes can we make to the ordering of terms that will aid understanding?The option examples illustrated below are based on experimentally inferred biological process annotations for Cdk1.Choose one option.

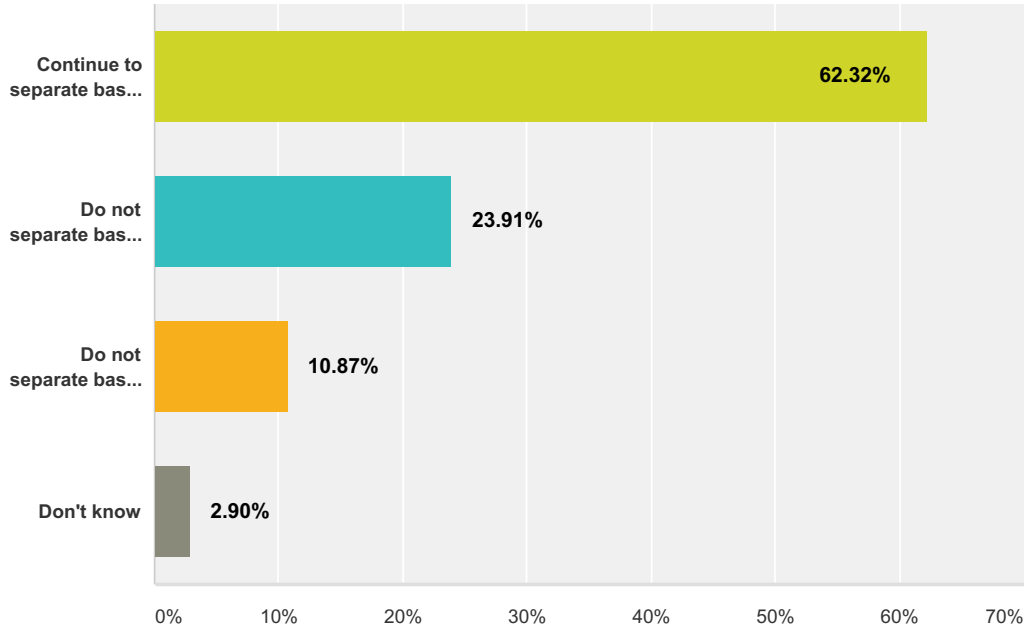
Answered: 170 Skipped: 2



Answer Choices	Responses
Sort the GO data using the ribbon categories, using only populated ribbon categories as section headers (panel B)	48.82% 83
Sort the GO data using the ribbon categories, using all ribbon categories as section headers, indicating how many terms are under each category (panel C)	28.82% 49
Do not change, I find the alphabetical listing easy to understand (panel A)	15.88% 27
Sort the GO data using the hierarchical structure of the ontology, but do not split into categories (panel D)	4.71% 8
None of the above - own comment	1.76% 3
Total	**170**

Q3 Should we continue to display GO annotations as two separate sections - those inferred from experimental data and those based on predictions/assertions, as in the example for Cdk1 molecular function below? Or would you prefer to have GO annotations from both experimental and non-experimental evidence displayed together?Choose one option.

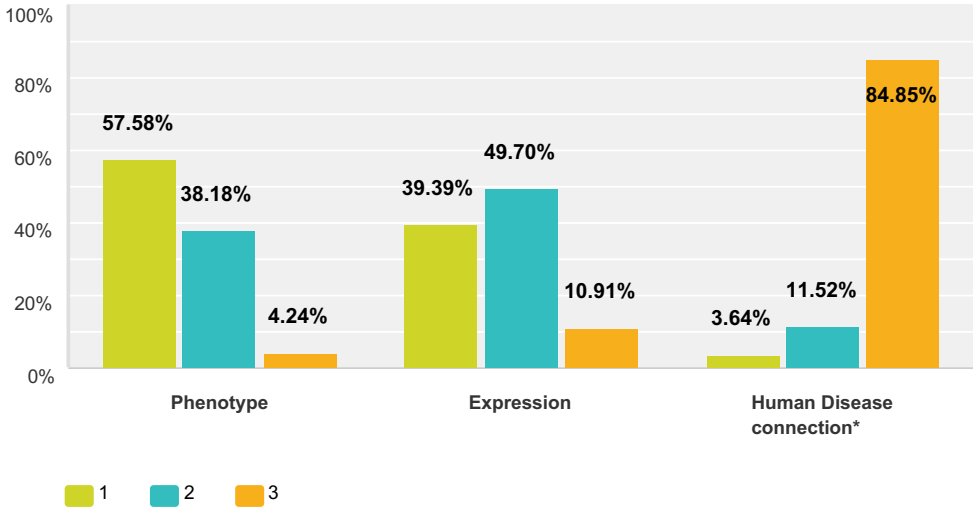
Answered: 138 Skipped: 34



Answer Choices	Responses	
Continue to separate based on evidence, as this is useful to me	62.32%	86
Do not separate based on evidence, but highlight where the evidence is experimental (e.g. by an icon)	23.91%	33
Do not separate based on evidence, I find the evidence statement sufficient to differentiate	10.87%	15
Don't know	2.90%	4
Total		138

Q4 In order of preference, rank the following data types you would like to see summarized in ribbon displays, where 1 is most valuable and 3 is the least.

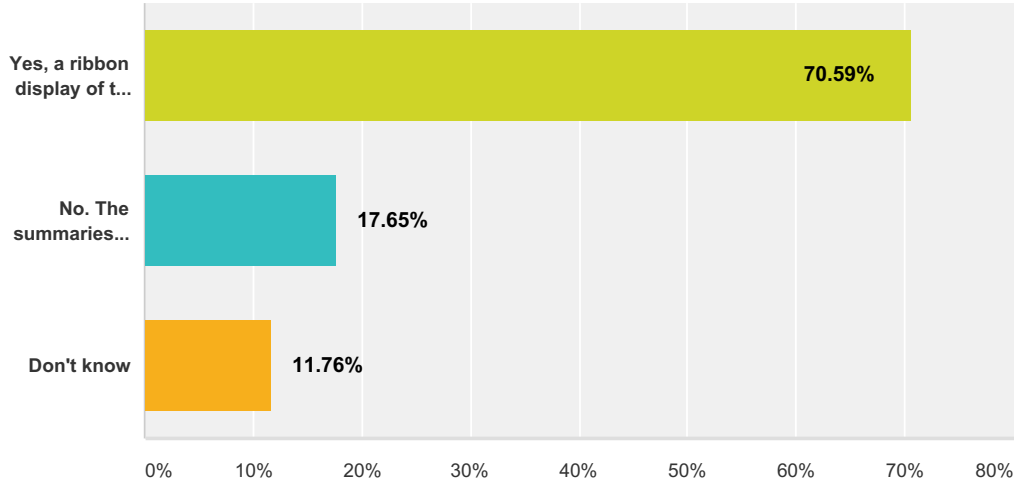
Answered: 166 Skipped: 6



	1	2	3	Total	Score
Phenotype	**57.58%** 95	**38.18%** 63	**4.24%** 7	165	2.53
Expression	**39.39%** 65	**49.70%** 82	**10.91%** 18	165	2.28
Human Disease connection*	**3.64%** 6	**11.52%** 19	**84.85%** 140	165	1.19

Q5 The allele reports contain associated phenotype data. Along with gene-level ribbon summaries in gene reports, should we also build ribbons specific for alleles to be included in the individual allele reports? An example of how phenotype data could be displayed for an allele of Cdk1 is shown below.Choose one option.

Answered: 170 Skipped: 2



Answer Choices	Responses	
Yes, a ribbon display of this data in allele reports would be useful to me	**70.59%**	120
No. The summaries should be only on the gene reports	**17.65%**	30
Don't know	**11.76%**	20
Total		**170**