

# The Zebrafish Model Organism Database: new support for human disease models, mutation details, gene expression phenotypes and searching

Douglas G. Howe\*, Yvonne M. Bradford, Anne Eagle, David Fashena, Ken Frazer, Patrick Kalita, Prita Mani, Ryan Martin, Sierra Taylor Moxon, Holly Paddock, Christian Pich, Sridhar Ramachandran, Leyla Ruzicka, Kevin Schaper, Xiang Shao, Amy Singer, Sabrina Toro, Ceri Van Slyke and Monte Westerfield

The Institute of Neuroscience, University of Oregon, Eugene, OR 97403-1254, USA

Received September 09, 2016; Revised October 25, 2016; Editorial Decision October 26, 2016; Accepted October 27, 2016

## ABSTRACT

The Zebrafish Model Organism Database (ZFIN; <http://zfin.org>) is the central resource for zebrafish (*Danio rerio*) genetic, genomic, phenotypic and developmental data. ZFIN curators provide expert manual curation and integration of comprehensive data involving zebrafish genes, mutants, transgenic constructs and lines, phenotypes, genotypes, gene expressions, morpholinos, TALENs, CRISPRs, antibodies, anatomical structures, models of human disease and publications. We integrate curated, directly submitted, and collaboratively generated data, making these available to zebrafish research community. Among the vertebrate model organisms, zebrafish are superbly suited for rapid generation of sequence-targeted mutant lines, characterization of phenotypes including gene expression patterns, and generation of human disease models. The recent rapid adoption of zebrafish as human disease models is making management of these data particularly important to both the research and clinical communities. Here, we describe recent enhancements to ZFIN including use of the zebrafish experimental conditions ontology, 'Fish' records in the ZFIN database, support for gene expression phenotypes, models of human disease, mutation details at the DNA, RNA and protein levels, and updates to the ZFIN single box search.

## INTRODUCTION

The Zebrafish Model Organism Database (ZFIN; <http://zfin.org>) is the community resource for genetic and genomic data involving zebrafish (*Danio rerio*). Data in ZFIN in-

cludes mutants, gene expression, phenotypes, knockdown reagents, antibodies, transgenic constructs, and reporter lines. These data are integrated from direct data submissions from the research community and expert curation of published literature by biocurators. ZFIN aims to provide the best support possible for current needs of the scientific research community. In this paper, we describe new features added to ZFIN in the past two years. These include use of a new experimental conditions ontology, curation of gene expression phenotypes, zebrafish models of human disease and molecular details of mutants, as well as the addition of a Reporter Lines category and a figure gallery to the single box search interface. The release schedule for these new projects, descriptive metrics for the new data types, and descriptive metrics for usage and data content of the ZFIN resource in general can be found in the Supplementary Tables S1–S4.

## THE 'FISH' CONCEPT AND EXPERIMENTAL CONDITIONS

Gene expression patterns, phenotypes, and models of human disease are all functions of the combined effects of the genetic constitution of an organism (genotype and genetic background), gene knockdown reagents, and the experimental conditions under which observations are made. At ZFIN, we have long been curating gene expression and phenotype from the zebrafish literature to a combination of a genotype plus the experimental conditions (1–3). Historically, ZFIN included the sequence targeting reagents (STRs) morpholinos, TALENs, and CRISPRs in experimental conditions when they were used as transient genetic modifiers (i.e. not used as germline mutagens). In order to obtain a better distinction between the genetic modifiers of a fish and the non-genetic experimental conditions, the STRs were moved out of the experimental conditions and included along with the genotype in a new database ob-

\*To whom correspondence should be addressed. Tel: +1 541 346 2355; Email: [dhowe@zfin.org](mailto:dhowe@zfin.org)

ject we call a ‘Fish’ (Figure 1). The Fish data model allows grouping of the transient and non-transient genetic modifiers, allowing researchers to more easily understand the number and variety of affected genes for which gene expression and phenotype observations are made. Fish records get unique and persistent identifiers in ZFIN in the form ZDB-FISH-#####-##### (see example in Figure 1). All gene expression and phenotype annotations, and models of human disease are currently annotated to a Fish record associated with specific experimental conditions. Information about Fish records (genotype, background, and STRs when applicable) can be found on ZFIN Fish record pages and in downloads files. Data from different model organism databases is currently annotated to records having varying degrees of specificity with regard to the genotype involved. In some cases annotations are made directly to allele records and in other cases annotations are made to more complex records like genotypes or ‘Fish’. This presents a challenge when trying to work with or combine these data. The ‘Fish’ database object aligns with ‘effective genotype’ (GENO:0000525) term from the GENO ontology, which was developed by Monarch Initiative to model genotype representations (4,5). ‘Fish’ records include the classical genotype representing the non-transient genetic modifiers, equivalent to the GENO ontology ‘intrinsic genotype’ (GENO:0000000), and the transient genetic modifiers (STRs), equivalent to the GENO ontology ‘extrinsic genotype’ (GENO:00005234). Use of an ontology such as GENO will be important for efforts aimed at combining data that may be annotated to different levels of granularity in the genotype/allele hierarchy.

Experimental conditions are the conditions under which experiments are performed and data are gathered. They can range from ‘Standard’, as defined in the Zebrafish Book (6) or ‘Generic Control’ (non-standard control treatments such as a vehicle DMSO injection) to any sort of manipulation of the environment of the fish (as in exposure to a chemical or a non-standard light/dark cycle) or physical manipulation of the fish itself (as in amputation). Historically, experimental conditions in ZFIN were limited to a small list of frequently used conditions. Details regarding the conditions, such as the structure amputated or the chemical used, were recorded in a text note which lacked the necessary structure for effective searching. To address those limitations, ZFIN experimental conditions are now annotated using a new ontology called the Zebrafish Experimental Conditions Ontology (ZECO). ZECO provides a more expressive representation of experimental conditions used in zebrafish research while providing an ontological framework to support searching, data reporting, and interoperability of ZFIN annotations using experimental conditions. The full scope and use of the ZECO ontology will be described in another publication.

## GENE EXPRESSION AS A PHENOTYPE

Gene expression patterns in mutants frequently differ greatly from patterns in wild-type animals in spatial extent, level (increased or decreased), or location. These differences illuminate genetic interactions and pathways. In some cases, particularly during early development, a change in an ex-

pression pattern may be the only notable phenotype. Until recently, phenotype data and gene expression data were entirely separate in ZFIN. The ZFIN database could store expression annotations such as ‘gene *xyz* is expressed in the kidney in *abc* mutant fish’ and phenotype annotations such as ‘the kidney is enlarged in *abc* mutant fish’. However, there was no way to record that the gene expression pattern of gene *xyz* in the kidney of *abc* mutant fish differed from the wild-type pattern. Database support for expression patterns as phenotypes (EaP) now allows these important phenotypes to be curated and searched. Between January and September 2016 ZFIN curators added 3087 EaP annotations from 253 publications. During that period, 59% of the 429 publications curated for gene expression and 20% of all 1281 expertly curated publications contained at least one EaP annotation (Supplementary Table S2). This underscores the importance of these expression phenotype annotations in the more general phenotype annotation landscape.

## Expression phenotype format

Gene expression phenotypes follow the same ‘Entity:Quality:Tag’ (EQT) format as other phenotypes in ZFIN (1,7). EQT annotations are typically composed of an ‘Entity’ term commonly from the Zebrafish Anatomy Ontology (AO) (8) or the Gene Ontology (GO) (9,10) combined with a Quality term from the Phenotypic Quality ontology (PATO) (11) and a Tag (‘normal’, ‘abnormal’, or ‘ameliorated’). The Entity in an EaP annotation combines an anatomical structure with either a gene or an antibody. An antibody is used in cases of antibody labeling data that is not connected with expression of a single gene, e.g. labeling by antibodies that recognize carbohydrate epitopes or multiple members of a gene family. Table 1 describes two phenotypes in the heart, one standard anatomical phenotype and one EaP. The standard anatomical phenotype could be stated in words as ‘the heart has abnormally increased volume’. The EaP could be stated as ‘*lfi2* expression in the heart is abnormally mislocalized.’

There are two possible EaP annotation tags—abnormal and ameliorated. The ameliorated tag is used to describe an expression pattern observed in a mutant that is significantly or completely rescued by an additional mutation or reagent. The EaP Quality is chosen from a small subset of terms from the PATO ontology (Table 2). After extensive discussion within ZFIN and with members of the zebrafish research community, it was concluded that the vast majority of user needs could be met by using only seven PATO quality terms for EaPs having the ‘abnormal’ tag, and three PATO quality terms for EaPs having the ‘ameliorated’ tag. The use of a small collection of PATO quality terms made the addition of EaP curation to the ZFIN curation interface simpler, keeping the day-to-day process of EaP curation quick and efficient.

## Logical representation of EaP data

EaP annotations come in two forms: gene and antibody. Which form is used depends upon whether the expression pattern is for a gene or for an antibody staining pattern that

## Fish = Genotype and Genetic Background + Sequence Targeting Reagents

	Genotype	G.B.	STRs
Fish name:	hand2 <sup>c99/c99</sup>	(AB)	+ MO1-dlx3b + MO1-dlx5a + MO3-dlx4b
Genotype:	hand2 <sup>c99/c99</sup>	(AB)	<input type="checkbox"/>
Targeting Reagent:	MO3-dlx4b	<input type="checkbox"/>	MO1-dlx5a <input type="checkbox"/> MO1-dlx3b <input type="checkbox"/>

**Figure 1.** The Fish concept in ZFIN. Information found at the top of a Fish page in ZFIN. Fish records include the genotype (Green; Genotype), the genetic background (Yellow, G.B.), and the sequence targeting reagents (Red, STRs) when applicable. Fish IDs have the form ZDB-FISH-#####-#####. Fish ZDB-FISH-150901-11301 (<https://zfin.org/ZDB-FISH-150901-11301>) includes a homozygous *hand2<sup>c99/c99</sup>* gene on the AB genetic background plus three morpholinos targeting the *dlx3b*, *dlx5a* and *dlx4b* genes.

**Table 1.** Howe *et al.* NAR database issue 2016

Annotation Type	Entity	Quality
Anatomical phenotype	Heart	Increased volume
Expression phenotype	Heart <i>lfi2</i> expression	Mislocalized

**Table 2.** Howe *et al.* NAR database issue 2016

EaP Qualities with 'Abnormal'	Tag corresponding PATO ID
Increased amount	PATO:0000470
Decreased amount	PATO:0001970
Absent	PATO:0000462
Mislocalized	PATO:0000628
Spatial pattern	PATO:0000060
Increased distribution	PATO:0001671
Decreased distribution	PATO:0001672
<b>EaP Qualities with 'Ameliorated' Tag</b>	
Amount	PATO:0000070
Position	PATO:0000140
Spatial pattern	PATO:0000060

is not clearly attributable to a single gene. In each case the phenotype is manifest in the expression of a gene or epitope in an anatomical location.

- **Gene Form:** <PATO> <RO:inheres.in> some (gene and <RO:expressed in> some <ZFA>)
- **Antibody Form:** <PATO> <RO:inheres.in> some (<SO:epitope> and <RO:expressed in> some <ZFA>)

<PATO> denotes a term from our short list of EaP PATO qualities. <ZFA> denotes a term from the zebrafish anatomical ontology which can also optionally be post-composed with a term from the GO Cellular Component ontology. <RO:expressed in> is the term 'expressed in' (RO:0002206) from the Relations Ontology (RO) (12). <SO:epitope> is the term 'epitope' (SO:0001018) from the Sequence Ontology (SO) (13). In both cases, the phenotype inheres in something being expressed in an anatomical structure.

### Curation of expression phenotypes

Curators add EaP data to ZFIN quickly using a simple extension of the existing 'building block' style gene expression curation interface. In the gene expression section of the ZFIN curation interface, 'Experiments' are created that combine Gene, Antibody (optional), Assay, Fish, and Experimental Conditions. Combining Experiments with a developmental stage and Figure number makes 'Expressions'.

The 'Anatomy pile' is a list of anatomy terms that is built by curators as they work on a particular publication. Associating one or more 'expressions' with one or more anatomical structures from the 'Anatomy pile' produces gene expression annotations (Figure 2). This approach maximizes re-use, so that experiments and anatomy terms are only entered once, and then they are combined as needed for different figures.

EaP terms are represented on the Anatomy Pile along with non-EaP anatomy terms, so the pile might contain 'posterior lateral plate mesoderm' and also 'posterior lateral plate mesoderm: decreased amount'. Adding 'posterior lateral plate mesoderm' to an Expression creates a gene expression entry in the database. Adding 'posterior lateral plate mesoderm: decreased amount' creates both gene expression and gene expression phenotype annotations.

### Displaying and searching EaPs

EaP phenotypes are shown along with non-EaP phenotypes in the Phenotype section of Fish and Figure pages. A variety of different phenotypes, including EaP as well as non-EaP phenotypes can be seen in the 'Phenotype Details' section of a typical Figure page (Figure 3).

Searching for EaP annotations is done in the same manner as searching for expression or phenotype annotations. For example, an EaP annotation representing 'fgf8a expression is mislocalized to the retina' can be found in the *fgf8a* gene expression data as well as the retina phenotype data.

**A**

Figure	Gene	Fish	Env	Assay	Ab	Stage Range	Expressed in	Del	Push
Fig. 2	klf1	AB/TL	Standard	ISH		20-25 somites (19.00h-22.00h)	posterior lateral plate mesoderm	<input type="checkbox"/>	<a href="#">to pato</a>
Fig. 2	klf1	it627[2,1,1]AB/TL	Standard	ISH		20-25 somites (19.00h-22.00h)	posterior lateral plate mesoderm(decreased amount)	<input type="checkbox"/>	<a href="#">to pato</a>

Structures: [Hide](#)

Update Structures for Expressions

**B**

Structure	Quality	Stage Range	Remove
<input type="radio"/> lateral plate mesoderm		1-4 somites (10.33h-11.66h) - Prim-5 (24.00h-30.00h)	<input type="checkbox"/>
<input type="radio"/> lateral plate mesoderm	decreased amount	1-4 somites (10.33h-11.66h) - Prim-5 (24.00h-30.00h)	<input type="checkbox"/>

Update Structures for Expressions

**Anatomy** : **Substructure**

AO:

within the

AO:

not  
 absent phenotypic  
 mislocalized  
 position ok  
 decreased distribution  
 increased distribution  
 spatial pattern abnormal  
 spatial pattern ok  
 decreased amount  
 increased amount  
 amount ok

**TERM:** anatomical structure  
**OBO ID [ID]:** ZFA:0000037 [ZDB-TERM-100331-35]  
**Synonyms:** anatomical structures  
**Definition:** Zebrafish anatomical entity that has inherent 3D shape and is generated by coordinated expression of the organism's own genome.  
**has subtype** [acellular anatomical structure](#) • [anatomical group](#) • [cell](#) • [compound organ](#) • [embryonic structure](#) • [extraembryonic structure](#) • [multi-tissue structure](#) • [neuraxis flexure](#) • [organism subdivision](#) • [portion of tissue](#) • [whole organism](#)  
**is a type of** [zebrafish anatomical entity](#)  
**start stage** [Zygote:1-cell](#)  
**end stage** [Adult](#)  
**Comments:**

**Figure 2.** A portion of the ZFIN gene expression curation interface. (A) portion of the ZFIN gene expression curation interface showing EaP and non-EaP expression annotations. (A) Two Expression rows, the first holding an ordinary gene expression annotation in a wild-type fish, the second row holding an EaP annotation in a mutant fish. (B) The Anatomy Pile section of the ZFIN curation interface showing the anatomical structure lateral plate mesoderm in standard format as well as expression phenotype format for lateral plate mesoderm - decreased amount. Image captured on 18 August 2016.

**Phenotype Details**

Fish	Conditions	Stage	Phenotype
AB/TL + MO1-gata1a <input type="checkbox"/>	standard	Prim-5	blood island <i>spi1b</i> expression mislocalised, abnormal <input type="checkbox"/>
kdm1a <sup>it627/it627</sup> (AB/TL) <input type="checkbox"/>	standard	20-25 somites	posterior lateral plate mesoderm <i>alas2</i> expression decreased amount, abnormal <input type="checkbox"/>
		20-25 somites	posterior lateral plate mesoderm <i>hbbe1.1</i> expression decreased amount, abnormal <input type="checkbox"/>
		20-25 somites	posterior lateral plate mesoderm <i>klf1</i> expression decreased amount, abnormal <input type="checkbox"/>
		Prim-25	nucleate erythrocyte decreased amount, abnormal <input type="checkbox"/>
		Day 4	erythroblast immature, abnormal <input type="checkbox"/>
		Day 6	erythroblast immature, abnormal <input type="checkbox"/>
kdm1a <sup>it627/it627</sup> (AB/TL) + MO1-gata1a <input type="checkbox"/>	standard	Prim-5	blood island <i>spi1b</i> expression position, ameliorated <input type="checkbox"/>

**Figure 3.** The Phenotype Details section of the ZFIN figure page. <https://zfin.org/ZDB-FIG-151214-52#expDetail> shows examples of EaP and non-EaP phenotypes displayed in the Phenotype column of the Phenotype Details section for the figure. Image captured on 18 August 2016.

EaP has been integrated into ZFIN's new single box search page. (<https://zfin.org/search>)

A new facet—'Misexpressed Gene', aids EaP discovery in five different data categories: Gene/Transcript, Phenotype, Fish, Mutation/Tg and Figure.

Clicking on a gene symbol in the 'Misexpressed Gene' facet will filter the result set to only include results that have EaP data for that gene. Another way to access EaP data is to use 'misexpression' in the search string. Table 3 includes examples of four common kinds of searches using the Misexpressed Gene facet or the 'misexpression' key word.

### EaP data downloads

Data in ZFIN are available from the downloads page (<http://zfin.org/downloads>), where the 'Phenotype Data' section has two new download files to support EaP annotations:

- Gene Expression Phenotypes [https://zfin.org/downloads/gene\\_expression\\_phenotype.txt](https://zfin.org/downloads/gene_expression_phenotype.txt)
- Antibody Labeling Phenotypes [https://zfin.org/downloads/antibody\\_labeling\\_phenotype.txt](https://zfin.org/downloads/antibody_labeling_phenotype.txt)

### HUMAN DISEASE

The zebrafish has become a well-established vertebrate model organism. The optical clarity, accessibility and rapid development of the zebrafish embryo allow organ systems and tissues to be closely observed as they form. There are close parallels between the genomes of zebrafish and humans. Sequencing of the zebrafish genome has revealed that 71% of human genes have at least one zebrafish ortholog while 69% of zebrafish genes have at least one human ortholog. In addition, 82% of human disease genes in OMIM have a zebrafish ortholog (14). A variety of powerful tools for genome manipulation are now available, facilitating the creation of zebrafish mutants that model specific human diseases. This section will highlight recent enhancements at ZFIN that allow users to access and navigate human disease information and zebrafish disease models for use in translational research.

A link to human disease information in ZFIN is provided through two different methods, one automated, one curated. An automated pipeline links OMIM diseases associated with human genes with the orthologous zebrafish genes using curated orthology. Orthology curation is done

Table 3. Howe *et al.* NAR database issue 2016

Search	Category	Filter settings
Which Fish have perturbed gene expression in the heart?	Fish	Affected Anatomy: heart Misexpressed Gene: Any
Which phenotypes include <i>nrg1</i> misexpression?	Phenotype	Expression Anatomy: heart
Which figures show <i>lefl</i> misexpression?	Figure	Misexpressed Gene: <i>nrg1</i>
Which phenotypes include misexpression of members of the <i>bmp</i> gene family?	Phenotype	Misexpressed Gene: <i>lefl</i> No Filters set Query string = 'bmp misexpression'

**A**

**DISEASE ASSOCIATED WITH *atxn2* HUMAN ORTHOLOG**

Disease Ontology Term	OMIM Term	OMIM Phenotype ID
<a href="#">amyotrophic lateral sclerosis type 13</a> <input type="checkbox"/>	Spinocerebellar ataxia 2	183090
	{Amyotrophic lateral sclerosis, susceptibility to, 13}	183090
<a href="#">Parkinson's disease</a> <input type="checkbox"/>	{Parkinson disease, late-onset, susceptibility to}	168600
<a href="#">spinocerebellar ataxia type 2</a> <input type="checkbox"/>	Spinocerebellar ataxia 2	183090
	{Amyotrophic lateral sclerosis, susceptibility to, 13}	183090

**B**

**Term Name:** Parkinson's disease

**Synonyms:** paralysis agitans, Parkinson disease

**Definition:** A synucleinopathy that has\_material\_basis\_in degeneration of the central nervous system that often impairs motor skills, speech, and other functions.[http://en.wikipedia.org/wiki/Parkinson%27s\\_disease](http://en.wikipedia.org/wiki/Parkinson%27s_disease)

**References:** EFO:0002508, ICD10CM:G20, ICD9CM:332 (all 33) ▶

**Ontology:** Human Disease

Search Ontology:

**Relationships**

**is a type of:** [synucleinopathy](#)

**has subtype:** [autosomal dominant Parkinson disease 1](#)   
[autosomal dominant Parkinson disease 8](#)   
[autosomal recessive early-onset Parkinson disease 7](#)   
[autosomal recessive early-onset Parkinson disease 15](#)   
[autosomal recessive juvenile Parkinson disease 2](#)   
... Show All 7 Terms

**GENES INVOLVED**

Human Gene	Zebrafish Ortholog	OMIM Term	OMIM Phenotype ID
ATXN2	<i>atxn2</i>	{Parkinson disease, late-onset, susceptibility to}	168600
DNAJC6	<i>dnajc6</i>	Parkinson disease 19, juvenile-onset	615528
EIF4G1	<i>EIF4g1a</i> , <i>EIF4g1b</i>	{Parkinson disease 18}	614251
FBXO7	<i>fbxo7</i>	Parkinson disease 15, autosomal recessive	260300
GBA	<i>gba</i>	{Parkinson disease, late-onset, susceptibility to}	168600

▼ Show all 14

**ZEBRAFISH MODELS**

Fish	Conditions	Citations
<a href="#">AB</a> <input type="checkbox"/>	chemical treatment: oxidopamine <input type="checkbox"/>	(1)
<a href="#">WT</a> <input type="checkbox"/>	chemical treatment: pharmaceutical <input type="checkbox"/>	(1)
	chemical treatment: N-methyl-4-phenylpyridinium <input type="checkbox"/>	(1)
	chemical treatment: pharmaceutical <input type="checkbox"/>	(1)
<a href="#">slc39a14<sup>u801/u801</sup> (AB)</a> <input type="checkbox"/>	chemical treatment: pharmaceutical <input type="checkbox"/>	(1)
<a href="#">AB + MO1-atp13a2</a> <input type="checkbox"/>	standard	(1)
<a href="#">AB + MO1-pink1 + MO2-pink1</a> <input type="checkbox"/>	standard	(1)
<a href="#">WT + MO1-park2 + MO2-park2</a> <input type="checkbox"/>	standard	(1)
<a href="#">ot80Tg</a> <input type="checkbox"/>	chemical treatment: pharmaceutical <input type="checkbox"/>	(1)

**Figure 4.** Human disease information in ZFIN. (A) The human disease section from the *atxn2* gene page showing OMIM phenotypes associated with the human ortholog of zebrafish *atxn2*. (B) The Disease page for Parkinson's disease showing details from the DO including disease name, synonyms, definition, cross references to other resources and term relationships within DO. The zebrafish models section shows models of this disease generated in zebrafish with chemical treatments or genetic alterations. Image captured on 24 August 2016.

The screenshot shows the Zebrafish Model Organism Database search interface. The search bar contains 'Parkinson's disease' and shows 194 results. On the left, a category list includes Gene / Transcript (23), Human Disease (11), Fish (5), Reporter Line (1), Mutation / Tg (1), Construct (1), Marker / Clone (2), Anatomy / GO (3), Community (7), and Publication (141). The search results are sorted by Relevance and show four entries:

- Parkinson's disease** (DOI:14330): Synonyms: paralysis agitans, Parkinson disease. Definition: A synucleinopathy that has\_material\_basis\_in degeneration of the central nervous system that often impairs motor skills, speech, and other functions. Related zebrafish gene (16) | Disease Models (9) | Publication (11). Matching Text: Gene Full Name: **parkinson** protein 7. Category: Human **Disease**. Name: **Parkinson's disease**. Alias: **Parkinson disease**. An arrow labeled 'Related Data Links' points to the 'Related zebrafish gene' link.
- postencephalitic Parkinson disease** (DOI:14332): Synonyms: Postencephalitic parkinsonism, Postencephalitic Parkinsonism, Postencephalitic parkinsonism (disorder). Matching Text: Category: Human **Disease**. Name: postencephalitic **Parkinson disease**.
- secondary Parkinson disease** (DOI:13548): Synonyms: disorder presenting primarily with parkinsonism, secondary parkinsonism, unspecified, Symptomatic parkinsonism (disorder), secondary parkinsonism, unspecified (disorder), secondary parkinsonism (disorder) [Ambiguous], secondary Parkinsonism, secondary parkinsonism (disorder). Matching Text: Category: Human **Disease**. Name: secondary **Parkinson disease**.
- Kufor-Rakeb syndrome** (DOI:0060556): Synonyms: PARK9, autosomal recessive Parkinson disease 9. Definition: A form of juvenile-onset atypical Parkinson disease associated with supranuclear gaze palsy, spasticity, and dementia. Some patients have neuroradiologic evidence of iron deposition in the basal. It is caused by homozygous or compound heterozygous mutation in the ATP13A2 gene, which encodes a lysosomal type 5 ATPase, on chromosome 1p36. Related zebrafish gene (1). Matching Text: Category: Human **Disease**. Definition: A form of juvenile-onset atypical **Parkinson disease** associated with. Alias: autosomal recessive **Parkinson disease 9**.

**Figure 5.** Single box search results. Search results for Parkinson's disease using the single box search. Categories of data matching the text search 'Parkinson's Disease' are in the left column along with the count of records found in that category. Image captured on 24 August 2016.

using the most recent human and zebrafish genome assemblies available at Ensembl at the time of each analysis (15). Expert curation of publications link zebrafish models of human disease, human disease, and the relevant publications utilizing the Disease Ontology (DO) (16). Each of these methods is described in more detail below. Phenotype based comparison approaches for human disease is under consideration (17). A tool that has been developed by the Monarch Initiative presents a grid detailing similarities between human and zebrafish phenotypes (18). ZFIN is evaluating integration of this widget on the disease term pages to provide users with a different option for linking data to human disease.

### Gene-disease association via OMIM

Gene-to-disease associations are brought into ZFIN through a data exchange with NCBI. This association is established through an automated pipeline that relates OMIM disease IDs with orthologous zebrafish genes at ZFIN via the NCBI human Gene ID. Zebrafish genes that are orthologous to human genes implicated in disease display OMIM disease data on the ZFIN gene pages, providing users with a snapshot of diseases for which the zebrafish gene could provide a model (Figure 4A). Human genes and orthologous zebrafish genes are also made available on the ZFIN human disease page (Figure 4B). It is important to note that this gene-to-disease relationship is derived solely through orthology determinations

between zebrafish and human genes and are not reviewed by curators at ZFIN.

### Disease annotation through literature curation

Expert manual curation of zebrafish models of human disease is another way in which ZFIN supports human disease model discovery. Zebrafish disease models are annotated in ZFIN through a combination of Fish, experimental conditions and terms from the Disease Ontology. The Fish could include a wild-type strain, or a mutant or transgenic line. The experimental conditions refer to non-standard treatment conditions such as temperature, diet or chemical treatment applied to the fish to produce a phenotype. If an experimental treatment was not used on the fish, 'standard' or generic control' is used as the experimental condition. The DO is used to make an annotation to the human disease and represents diseases in an ontological hierarchy, with cross references to disease concepts from OMIM, SNOMed, NCI Thesaurus and MeSH (16). The data structure that ZFIN uses to record disease models can be summarized as: 'Fish' in specific 'experimental conditions' is\_a\_model\_of 'disease X' as described in 'Citation N'. A summary of disease models in this format can be seen on the Parkinson's disease page shown in Figure 4B. In addition, terms from the DO are used to establish a relationship between a human disease and a journal article about the disease, regardless of whether a zebrafish model is created.

Every annotation linking a zebrafish model of human disease to a disease term is supported by an evidence code

**A** ZFIN ID: ZDB-ALT-080318-3 Your Input Welcome

**Genomic Feature:** **b1127**

**Synonyms:**

**Affected Genes:** [fam20b \(1\)](#)

**Type:** Point Mutation (2)

**Protocol:** adult males treated with ENU

**Lab Of Origin:** [Postlethwait Zebrafish Genetics Laboratory](#)

**Location:** Chr: 20 [Details](#)

**Sequence:**

**Current Sources:**

**MUTATION DETAILS**

**DNA/cDNA Change:** T>C at position 991 in Exon 6 (1)

**Transcript Consequence:** [Missense](#)  (1)

**Protein Consequence:** [Amino Acid Substitution](#) : Cys>Arg at position 331 (1)

---

**B** ZFIN ID: ZDB-GENE-040724-125 Your Input Welcome

**Gene Name:** *family with sequence similarity 20, member B (H. sapiens)*

**Gene Symbol:** *fam20b*

**Previous Names:** [si:ch211-195o19.3](#), [unm b1127 \(1\)](#), [unm\\_b1127 \(1\)](#), [zgc:136944](#)

**Location:** Chr: 20 [Mapping Details/Browsers](#)

[Nomenclature History](#)

**GENE EXPRESSION**

**All Expression Data:** [Fig. 4 from Eames et al., 2011](#)

**Wild-type Stages, Structures:** [Blastula:1k-cell \(3.0h-3.33h\)](#) to [Larval:Protruding-mouth \(72.0h-96.0h, 3.5mm\)](#)  
[brain](#) , [ceratohyal cartilage](#) , [ceratohyal cartilage chondrocyte](#) , [ethmoid cartilage](#)  (all 7) [▶](#)

**Curated Microarray Expression:** [GEO \(1\)](#)

**MUTATIONS AND SEQUENCE TARGETING REAGENTS**

Allele	Type	Localization	Consequence	Mutagen	Suppliers
<a href="#">b1125</a>	Point Mutation	Exon 7	Premature Stop	ENU	
<a href="#">b1127</a>	Point Mutation	Exon 6	Missense	ENU	

**Figure 6.** Mutation details as seen on the genomic feature page and the gene page. Mutation details are displayed on the genomic feature and gene pages in ZFIN. (A) Snap shot of the *b1127* point mutation feature page. The mutation details section (bracket) shows the DNA/cDNA change with the corresponding location within the gene, and the consequences at the transcript and protein level. (B) Snap shot of the *fam20b* gene page. The mutations available for this gene are displayed with details regarding their localization within the gene, and the transcript consequences of the mutation (bracket). Image captured on 18 August 2016.

Reporter Line  Go New Feedback

Remove All Reporter Line × Expression Anatomy: Meckel's cartilage × Stage: Larval (72 hpf - 29.99 dpf) × Reporter Gene: mCherry ×

**▼ Reporter Gene**

mCherry

**▼ Expression Anatomy**

Meckel's cartilage

cartilage element (2)

chondrocranium (2)

chondrocranium cartilage (2)

cranial cartilage (2)

[Show All](#)

**▼ Regulatory Region**

col2a1a (1)

cryaa (1)

**▼ Stage**

Larval (72 hpf - 29.99 dpf)

Gastrula (5.25 - 10.32 hpf) (1)

Segmentation (10.33 - 23.99 hpf) (1)

Pharyngula (24.00 - 47.99 hpf) (1)

Hatching (48.00 - 71.99 hpf) (1)

[Show All](#)

Download 2 results Browse Images Sorted by Relevance

**hu5910Tg** ZDB-FISH-150901-28629 Reporter Line

Affected Gene	Line / Reagent	Mutation Type	Construct	Parental Zygosity
<a href="#">hu5910Tg</a>	<a href="#">hu5910Tg</a>	Transgenic Insertion	<a href="#">TgBAC(col2a1a:mCherry)</a>	Unknown

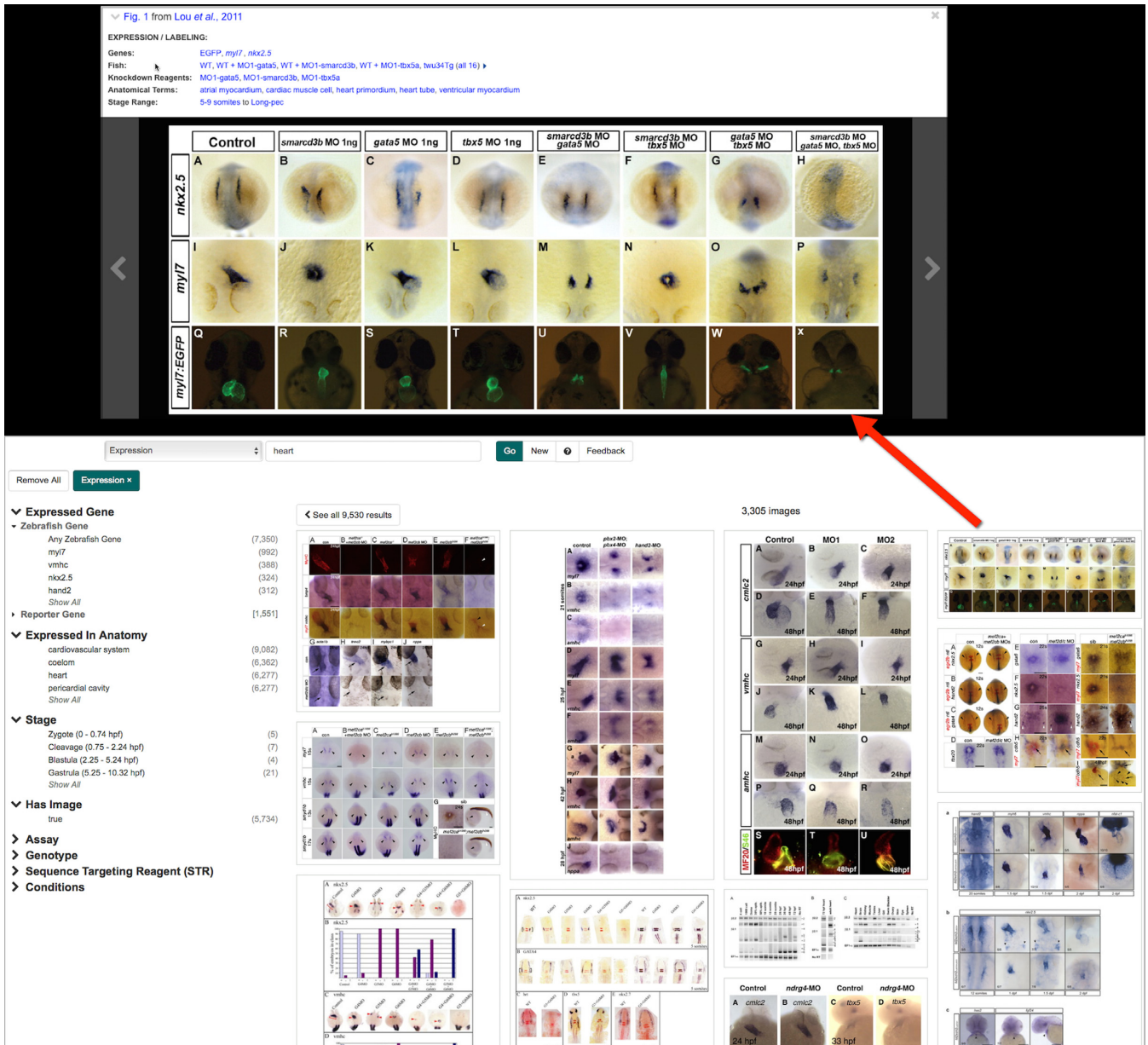
[Expression \(5\)](#) | [Publication \(10\)](#)

**nu15Tg** ZDB-FISH-150901-20383 Reporter Line

Affected Gene	Line / Reagent	Mutation Type	Construct	Parental Zygosity
<a href="#">nu15Tg</a>	<a href="#">nu15Tg</a>	Transgenic Insertion	<a href="#">Tg(Mmu.Foxa2-cryaa:mCherry)</a>	Unknown

[Expression \(16\)](#) | [Publication \(3\)](#)

**Figure 7.** An example search in the Reporter Lines Category in ZFIN single box search interface. In this example, the Reporter Line Category was selected and values were set in the Expression Anatomy, Stage, and Reporter Gene filters to answer the question 'Are there any reporter lines that express mCherry in Meckel's cartilage during the larval stage?'. Image captured on 24 August 2016.



**Figure 8.** The single box search figure gallery. The figure gallery provides a tiled view of images associated with results of a single box search. Here, the Expression category has been searched for 'heart' and the 'show images' button was clicked to get to the figure gallery view (lower). Clicking on any figure produces a magnified view of that image (red arrow). The magnified view includes the figure caption and curated data associated with that figure. Left and Right arrows provide continued navigation through the image set in the magnified view. The particular image clicked here is Figure 1 from Lou *et al.* (23). Images captured on 24 August 2016.

from the Evidence and Conclusion Ontology (ECO) (19). The evidence code 'Traceable Author Statement (TAS)' is used when authors make an explicit assertion about a Fish plus experimental condition as a model for a specific disease. The 'Inferred by Curator (IC)' evidence code is used if authors do not make an explicit assertion about the disease model but instead report phenotypes similar to the human patient pathologies, and the intent of their publication is to elucidate portions of the disease mechanism based on results obtained in zebrafish.

As of 10 October 2016 there are 372 curated zebrafish disease models and annotations using 357 DO terms. 16% of

curated publications between 17 June and 10 October 2016 included a disease model (Supplementary Table S2).

### Searching for disease models

Human disease information in ZFIN can be queried by entering a disease name or gene of interest into the search box on the home page. The results can be refined further by selecting a data category on the left of the search results or by selecting a 'related data' link in the body of one of the search results (Figure 5).



Disease term search results link to the Disease term page, which provides more details on the Human disease including definitions, synonyms and cross-references in other databases (Figure 4B). The relationship of the disease in the DO to other terms is also provided. The zebrafish models section of the disease page contains details of the genetic background of the fish, experimental conditions and the citation referring to the use of this model. Following either the Fish or Citations link in the models section provides access to phenotype information for each individual model. A list of established disease models and the ability to search for them will be very useful to researchers and/or clinicians studying a particular disease.

## MUTATION DETAILS

Information about nucleotide or peptide changes for alleles has previously been provided in a free text note on genomic feature (allele) pages. Annotation and display of these data have now been updated to provide better reporting and searching of mutation details. The genomic feature page now displays detailed information about the DNA/cDNA change, as well as associated transcript and peptide changes for alleles (Figure 6A). These data originate from publications, direct submissions to the nomenclature coordinator, or via data loads. As of October 2016, ZFIN includes DNA, transcript, or protein mutation details for 83%, 82% and 3% of applicable feature records respectively (Supplementary Table S2).

Various types of DNA/cDNA changes are described based on the type of mutation that has occurred. Nucleotide substitutions, like Adenine to Thymine, are reported for point mutations while the numbers of base pairs inserted and/or deleted are reported for mutations that are described as insertion, small deletion, or indel. The position of the nucleotide alteration within the DNA/cDNA and the associated reference sequence as well as the corresponding location within the gene (such as exon, intron, splice acceptor site, promoter) are reported when these details are provided. Complex mutations, where there are multiple substitutions, are reported in notes.

Nucleotide changes can have multiple effects on the resulting mutant transcript producing missense or splicing variants that can encode a frameshift or premature stop for example. These changes can be described by multiple transcript consequences. Terms from the SO (20) are used to annotate these sequence variants. A single mutation can have different effects on multiple transcripts originating from a single gene. ZFIN does not currently report the effect of mutations on each specific transcript.

Ultimately, nucleotide changes and transcript alterations can result in modifications at the protein level. The 'Protein consequence' section of the mutation details includes these changes such as protein truncation, amino acid insertion, or amino acid deletion. Amino acid sequence modification, such as amino acid change (e.g. Cys > Arg, Tyr > Stop) and the number of amino acids inserted or deleted are also reported, as well as the position of this modification and its sequence of reference when available. Information that cannot be effectively annotated into a particular section (for

example protein domains affected by the mutation) remains in the notes.

Currently mutation details are only recorded for point mutations, small deletions, insertions, indels, and transgenic insertions where the insertion site in a gene is known. It is important to note that mutation details as described above are rarely reported in their entirety. In addition to being displayed on the genomic feature page, mutation consequences for a particular allele are also displayed on the associated gene page, where a detailed view of known mutations for a single gene is provided (Figure 6B). Transcript consequences are also displayed in single box search results in the Mutants and Transgenics category, and can be used to filter the search results.

## UPDATES TO SINGLE BOX SEARCH

The ZFIN single box search was updated to a Solr-backed faceted interface in October 2014 (15). Since that initial release, there have been numerous small updates and several major additions. Here we describe two major new features that were recently added: The Reporter Genes category and the Figure Gallery.

### Reporter lines category

Reporter lines have become indispensable tools for a wide array of zebrafish research projects. However, locating a reporter line that meets specific criteria for the reporter gene, the promoter driving the reporter expression, or the anatomical location and developmental time at which the reporter is expressed can be challenging. To help address this need, the ZFIN single box search now has a 'Reporter Lines' category. This Category is limited to lines that express a reporter gene, such as EGFP or mCherry, and excludes lines those lines with any mutant alleles, sequence targeting reagents, or experimental treatments. Four parameters are provided to further refine reporter line searches; reporter gene, anatomical structure, transgene regulatory region, and reporter expression developmental stage. Together, these filters readily support inquiries such as 'Are there any reporter lines that express mCherry in Meckel's cartilage during the larval stage?' ZFIN currently includes two such lines: hu5910Tg and nu15Tg (Figure 7). Queries involving a specific transgene regulatory region are also supported. For example, 'Are there any reporter lines that use the regulatory region of the *agr2* gene to drive expression of any reporter in the anterior crista?' ZFIN currently includes one reporter line meeting those criteria: as11Tg. The search results provide links to all the expression data for each reporter line as well as direct access to associated figures. The new figure gallery (discussed below) provides another way to rapidly review images of reporter gene expression.

### Figure gallery

Quick and easy access to images is a favorite feature for ZFIN users. In response to user input, a new figure gallery has been added to the single box search interface. All single box search results that return images now include a button to 'Browse Images'. The image browsing view is a familiar

web paradigm for surveying sets of images in a tiled format (Figure 8). Clicking on a figure in the tiled format displays a magnified view that includes curated figure summary data. Left and right arrows allow navigation through the image set in the magnified view format.

## DISCUSSION

ZFIN aims to provide the best support possible for the zebrafish research community and for the broader biomedical community having an interest in the data, tools, and resources that zebrafish research generates. Expanded support for curated zebrafish models of human disease will be especially useful to biomedical researchers.

A great strength of zebrafish as a research tool is their transparency during early development, allowing the detailed observation of gene expression patterns in wild-type and mutant fish embryos and larvae. New features have been added to ZFIN to support curation, search and display of phenotypic expression patterns.

As mutagenesis techniques have evolved, zebrafish researchers have asked ZFIN to provide more details about mutations and their consequences. In response to this, more details about the nature of mutations and their molecular consequences is now provided in a structured format. These new data are curated from literature and added via data loads. Results of mutation searches can be filtered by the transcript consequences. These data will provide an additional way for researchers to evaluate the utility of a specific mutant for their research.

Searching for data is a critical component of any on-line data resource. A new ‘Reporter Line’ Category has been added to the ZFIN single box search to streamline and simplify searches for reporter lines with particular attributes of interest. Images are an important part of many searches. The new Figure Gallery feature brings simple image browsing to the single box search. These new additions to ZFIN will help members of the research community find data of interest and generate new hypotheses.

The ZFIN team strives to have as complete a set of data as possible. Community input is always welcomed to help supplement the data in ZFIN. New information regarding disease models, gene expression phenotypes, and mutation details are particularly relevant now that support has been added for those new features. To make a contribution, contact us at [zfinadm@zfin.org](mailto:zfinadm@zfin.org) or use the ‘Your Input Welcome’ button found on the top right of every page in ZFIN.

## ZFIN TECHNICAL

The ZFIN web architecture is written primarily in Java using the Spring Framework and served via JSP in Apache/Tomcat. Perl, SQL and Groovy are used primarily to process and load data in bulk. Hibernate serves as the object-relational mapping software from Java to an IBM Informix relational database. ZFIN uses both Solr and Java/Spring to facilitate search interfaces. Data from papers is entered via a web-based curation interface primarily written in GWT with some interfaces implemented in AngularJS. The community wiki is powered by Atlassian Confluence software (<http://www.altassian.com/software/confluence/>).

ZFIN supports an InterMine (21) instance, ZebrafishMine, here: <http://ZebrafishMine.org> (<https://github.com/ZebrafishMine/intermine>), which also serves ZFIN data via web services (15). InterMine software is produced by the Generic Model Organism Database (GMOD; <http://gmod.org>) project (22). A detailed and browsable view of the current ZFIN data model can be found at <http://zfin.org/schemaSpy/index.html>. Basic metrics describing web site usage are included in Supplementary Table S4.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Human Genome Research Institute of the National Institutes of Health [HG002659]. Funding for open access charge: The NHGRI grant that funds ZFIN.

*Conflict of interest statement.* None declared.

## REFERENCES

- Howe, D.G., Bradford, Y.M., Conlin, T., Eagle, A.E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S.A.T. *et al.* (2013) ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Res.*, **41**, D854–D860.
- Howe, D.G., Bradford, Y.M., Eagle, A., Fashena, D., Frazer, K., Kalita, P., Mani, P., Martin, R., Moxon, S.T., Paddock, H. *et al.* (2016) Chapter 23—a scientist’s guide for submitting data to ZFIN. *Methods Cell Biol.*, **135**, 451–481.
- Sprague, J., Bayraktaroglu, L., Clements, D., Conlin, T., Fashena, D., Frazer, K., Haendel, M., Howe, D.G., Mani, P., Ramachandran, S. *et al.* (2006) The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res.*, **34**, D581–D585.
- Brush, Matthew H, Mungall, Chris J, Washington, Nicole and Haendel, M.A. (2013) What’s in a genotype?: an ontological characterization for integration of genetic variation data. In: *ICBO 2013 - Proceedings of the 4th International Conference on Biomedical Ontology*. CEUR-WS, Montreal, pp. 105–108.
- McMurry, J.A., Kohler, S., Washington, N.L., Balhoff, J.P., Borromeo, C., Brush, M., Carbon, S., Conlin, T., Dunn, N., Engelstad, M. *et al.* (2016) Navigating the phenotype frontier: the monarch initiative. *Genetics*, **203**, 1491–1495.
- Westerfield, M. (2007) *The Zebrafish Book*, 5th Edition, A guide for the laboratory use of zebrafish (*Danio rerio*) 5th Editio. University of Oregon Press.
- Mungall, C.J., Gkoutos, G. V, Smith, C.L., Haendel, M.A., Lewis, S.E. and Ashburner, M. (2010) Integrating phenotype ontologies across multiple species. *Genome Biol.*, **11**, R2.
- Van Slyke, C.E., Bradford, Y.M., Westerfield, M. and Haendel, M.A. (2014) The zebrafish anatomy and stage ontologies: representing the anatomy and development of *Danio rerio*. *J. Biomed. Semantics* **5**, 12.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
- Gene Ontology Consortium (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
- Gkoutos, G. V, Green, E.C.J., Mallon, A.-M., Hancock, J.M. and Davidson, D. (2005) Using ontologies to describe mouse phenotypes. *Genome Biol.*, **6**, R8.
- Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L. and Rosse, C. (2005) Relations in biomedical ontologies. *Genome Biol.*, **6**, R46.
- Cunningham, F., Moore, B., Ruiz-Schultz, N., Ritchie, G.R. and Eilbeck, K. (2015) Improving the Sequence Ontology terminology for genomic variant annotation. *J. Biomed. Semantics*, **6**, 32.

14. Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L. *et al.* (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, **496**, 498–503.
15. Ruzicka, L., Bradford, Y.M., Frazer, K., Howe, D.G., Paddock, H., Ramachandran, S., Singer, A., Toro, S., Van Slyke, C.E., Eagle, A.E. *et al.* (2015) ZFIN, the zebrafish model organism database: Updates and new directions. *Genesis*, **53**, 498–509.
16. Schriml, L.M. and Mitra, E. (2015) The Disease Ontology: fostering interoperability between biological and clinical human disease-related data. *Mamm. Genome*, **26**, 584–589.
17. Mungall, C.J., Washington, N.L., Nguyen-Xuan, J., Condit, C., Smedley, D., Kohler, S., Groza, T., Shefchek, K., Hochheiser, H., Robinson, P.N. *et al.* (2015) Use of model organism and disease databases to support matchmaking for human disease gene discovery. *Hum. Mutat.*, **36**, 979–984.
18. Haendel, M.A., Vasilevsky, N., Brush, M., Hochheiser, H.S., Jacobsen, J., Oellrich, A., Mungall, C.J., Washington, N., Kohler, S., Lewis, S.E. *et al.* (2015) Disease insights through cross-species phenotype comparisons. *Mamm. Genome*, **26**, 548–555.
19. Chibucos, M.C., Mungall, C.J., Balakrishnan, R., Christie, K.R., Huntley, R.P., White, O., Blake, J.A., Lewis, S.E. and Giglio, M. (2014) Standardized description of scientific evidence using the Evidence Ontology (ECO). *Database (Oxford)*, **2014**, bau075.
20. Eilbeck, K., Lewis, S.E., Mungall, C.J., Yandell, M., Stein, L., Durbin, R. and Ashburner, M. (2005) The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, **6**, R44.
21. Lyne, R., Sullivan, J., Butano, D., Contrino, S., Heimbach, J., Hu, F., Kalderimis, A., Lyne, M., Smith, R.N., Štěpán, R. *et al.* (2015) Cross-organism analysis using InterMine. *Genesis*, **53**, 547–560.
22. Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
23. Lou, X., Deshwar, A.R., Crump, J.G. and Scott, I.C. (2011) Smarcd3b and Gata5 promote a cardiac progenitor fate in the zebrafish embryo. *Development*, **138**, 3113–3123.