

The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources

Leyla Ruzicka¹*, Douglas G. Howe, Sridhar Ramachandran, Sabrina Toro, Ceri E. Van Slyke, Yvonne M. Bradford, Anne Eagle, David Fashena, Ken Frazer, Patrick Kalita, Prita Mani, Ryan Martin, Sierra Taylor Moxon, Holly Paddock, Christian Pich, Kevin Schaper, Xiang Shao, Amy Singer and Monte Westerfield

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

Received September 16, 2018; Revised October 18, 2018; Editorial Decision October 19, 2018; Accepted October 25, 2018

ABSTRACT

The Zebrafish Information Network (ZFIN) (<https://zfin.org/>) is the database for the model organism, zebrafish (*Danio rerio*). ZFIN expertly curates, organizes and provides a wide array of zebrafish genetic and genomic data, including genes, alleles, transgenic lines, gene expression, gene function, mutant phenotypes, orthology, human disease models, nomenclature and reagents. New features at ZFIN include increased support for genomic regions and for non-coding genes, and support for more expressive Gene Ontology annotations. ZFIN has recently taken over maintenance of the zebrafish reference genome sequence as part of the Genome Reference Consortium. ZFIN is also a founding member of the Alliance of Genome Resources, a collaboration of six model organism databases (MODs) and the Gene Ontology Consortium (GO). The recently launched Alliance portal (<https://alliancegenome.org>) provides a unified, comparative view of MOD, GO, and human data, and facilitates foundational and translational biomedical research.

INTRODUCTION

The Zebrafish Information Network (ZFIN) (<https://zfin.org/>) is the central repository for zebrafish (*Danio rerio*) genetic and genomic data. ZFIN collects, organizes, and makes available a wide range of data on zebrafish, including genes, gene function, sequences, mutants, transgenic lines, human disease models, expression, phenotype, orthology, sequence targeting reagents and antibodies. Data at ZFIN can be accessed through ZFIN search interfaces, download files, and ZebrafishMine (zebrafishmine.org), a data mining resource (1). Community services include nomenclature

support for genes and alleles, ZFIN pages for researchers, laboratories and companies, and a wiki for researchers to share antibody and protocol information. Here we describe recent updates and additions to the ZFIN resource, including support for non-coding genes and genomic regions, and more expressive Gene Ontology annotations. We review the role of ZFIN in maintaining the zebrafish reference genome sequence and how the community can submit requests for updates to the genome sequence. Finally, we discuss ZFIN as a member of the Alliance of Genome Resources, a collaboration among six model organism databases and the Gene Ontology consortium.

ZFIN STATISTICS

ZFIN staff expertly curate detailed gene expression and phenotype data from publications and from user submissions (2,3). As of 30 August 2018, there are a total of 367 885 curated gene expression patterns and 144 581 distinct phenotypes in ZFIN; a total of 13 500 genes have expression data, and 7400 genes have phenotype data. Curation of perturbed gene expression patterns in mutants, referred to as gene expression phenotypes, started in early 2016 (4). As of 30 August 2018, 13 864 expression phenotypes had been curated at ZFIN (<https://zfin.org/downloads>, files ‘Zebrafish Gene Expression by Stage and Anatomy Term’, ‘Fish with Phenotypes’, ‘Gene Expression Phenotypes’, ‘Antibody Labeling Phenotypes’).

ZFIN has been collecting structured information on mutation details at the DNA, RNA and protein levels since June 2016 (4). Examples of mutation details include the nucleotide change and the location of the mutation (‘G>T’, ‘C>A in Donor Splice Site of Intron 2’ or ‘8 bp deleted from position 725 to 732 in RefSeq:NM_131183.1 in Exon 1’), as well as the transcript and protein consequences of the mutation (‘Splice Site’, ‘Exon Loss of Exon 2’, ‘Premature stop at position 128’, ‘Polypeptide truncation: Arg>Stop at

*To whom correspondence should be addressed. Tel: +1 541 346 4979; Email: leyla@zfin.org

position 499'). The mutation details displayed in ZFIN reflect the amount of detail provided in publications and submissions since 2016. Mutation details may be less complete for older mutants. Of the 48 816 genomic features in ZFIN, 36 992 had mutation details as of 1 September 2018 (<https://zfin.org/downloads>, file 'Genomic Features and their affected genes').

Zebrafish is an important model organism for studying human diseases. ZFIN established support for curating and providing data on human disease models in June 2015 (4,5). By September 2016, 411 disease models, corresponding to 175 human diseases had been annotated (5). As of 1 September 2018, ZFIN included 1066 curated zebrafish models for 325 human diseases (<https://zfin.org/downloads>, file 'Human Disease/Models').

NEW PROJECTS

Genomic regions and Non-Coding RNA genes

The first phase of the ZFIN Regions Project was completed in 2017 to increase support for genomic regions. A subset of genomic regions was selected from the Sequence Ontology to be supported in ZFIN (<http://www.sequenceontology.org/>) (6) (Table 1). Experimentally validated genomic regions can now be added to the ZFIN database from publications or through data loads. Genomic regions such as conserved non-coding elements (CNE), enhancers, or promoters can be curated and linked to genes with which they interact, or to constructs that contain them. Regions can be browsed in the 'Markers/Clones' category of the ZFIN single box search.

Support for non-coding RNA genes and their transcripts in ZFIN has also been increased: non-coding RNA genes are now classified into Sequence Ontology (SO) based sub-categories, such as miRNA Gene, lincRNA Gene, and tRNA Gene. Each non-coding RNA gene type has a unique identifier prefix in ZFIN (Table 2). Types of non-coding RNA genes and their transcripts can be visualized in the Gene/Transcript category in the ZFIN single box search (<https://zfin.org/search?q=&fq=category%3A%22Gene±%2F±Transcript%22>) (Figure 1).

Gene Ontology

Gene Ontology (GO) annotations describe the molecular function enabled by a gene product, the biological process in which it is involved, and the cellular components in which it is performing its activity. These annotations are created by associating a gene product with a single GO term, supporting evidence on which this annotation was made (e.g. experimental evidence), and literature reference (7).

These simple GO annotations are limited in expressivity, and the context of the activity/function/cellular localization is often lost. For example, *her8a* (ZDB-GENE-030131-2376) was annotated to the term 'negative regulation of transcription by RNA polymerase II' (GO:000122) (8); however, curators could not specify in the GO annotation which genes were regulated by *her8a*. *axin1* (ZDB-GENE-000403-1) was annotated as a regulator of the Wnt signaling pathway (negative regulation of canonical Wnt signaling pathway, GO:0090090) (9), but the fact that this

regulation was part of the anterior/posterior regionalization of the neural plate was lost. The cell type in which an activity/function occurred or the cellular component in which it was located was also lost in these basic GO annotations; for example, curators could not specify that *clndnd* (ZDB-GENE-010328-4) exercises its function in the 'bicellular tight junction' (GO:0005923) of the bile canaliculus of the liver (10).

To create more descriptive GO annotations, 'annotation extensions' were introduced by the Gene Ontology Consortium, allowing the recording of the biological context (e.g. molecular targets, locational context, biological process context) in which the gene product performs its function (11). Annotation extensions are created by associating a basic GO annotation with a relation and an entity, which is an identifier for a database object or an ontology term (11).

These annotation extensions support recording that *her8a* regulates the transcription of *dtx1* (ZDB-GENE-130503-1) (8), *axin1* regulates the canonical Wnt signaling pathway involved in the neural plate anterior/posterior regionalization (9), and *clndnd* exercises its function in the 'bicellular tight junction' (GO:0005923) of the bile canaliculus of the liver (10). Examples of GO annotations including their annotation extensions are shown in Table 3.

The ZFIN database was updated to record and display GO annotations with their annotation extensions, making zebrafish GO annotations more expressive. All zebrafish GO annotations, including those using annotation extensions, are available through the ZFIN Gene Association Files (GAF) at the Gene Ontology Consortium (http://geneontology.org/gene-associations/gene_association.zfin.gz).

Genome maintenance

The Genome Reference Consortium (GRC) uses a single, preferred tiling path to produce a consensus representation of the genome. Issues with the quality of the zebrafish genome assembly were catalogued and corrected by the Genome Reference Informatics Team (GRIT) at the Wellcome Trust Sanger Institute. The GRIT group, which is a member of the GRC, developed the latest release of the reference genome assembly, GRCz11. This was released in May 2017 and includes nearly 1000 finished clone sequences and the resolution of more than 400 previous issues with the assembly.

GRCz11 was built using clone sequences ordered and oriented according to genetic markers and BioNano data (12). Gaps and assembly issues were resolved using selected contigs from whole genome sequencing assemblies WGS31 and WGS32. These are novel whole genome shotgun assemblies generated from Illumina sequencing of a single double-haploid Tübingen strain fish. Variant sequence representations for particular regions are displayed using alternate loci (ALT_REF_LOCI) scaffolds. These sequences were derived from clones exhibiting at least 5 kb of unique sequence that was not present in the primary chromosomal sequence. Further details regarding the GRCz11 assembly can be found at: <https://www.ncbi.nlm.nih.gov/grc/zebrafish>.

Table 1. Genomic Regions supported in ZFIN

Sequence Ontology (SO) term name (ID)	Sequence Ontology term definition ^a	New ZFIN ID Prefix
binding_site (SO:0000409)	A biological_region of sequence that, in the molecule, interacts selectively and non-covalently with other molecules. A region on the surface of a molecule that may interact with another molecule. When applied to polypeptides: Amino acids involved in binding or interactions. It can also apply to an amino acid bond which is represented by the positions of the two flanking amino acids.	ZDB-BINDSITE-
biological_region (SO:0001411)	A region defined by its disposition to be involved in a biological process.	ZDB-BR-
DNA_motif (SO:0000713)	A motif that is active in the DNA form of the sequence.	ZDB-DNAMO-
enhancer (SO:0000165)	A cis-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter.	ZDB-ENHANCER-
enhancer_binding_site (SO:0001461)	A binding site that, in the enhancer region of a nucleotide molecule, interacts selectively and non-covalently with polypeptide residues.	ZDB-EBS-
epigenetically_modified_region (SO:0001720)	A biological region implicated in inherited changes caused by mechanisms other than changes in the underlying DNA sequence.	ZDB-EMR-
histone_modification (SO:0001700)	Histone modification is a post translationally modified region whereby residues of the histone protein are modified by methylation, acetylation, phosphorylation, ubiquitination, sumoylation, citrullination, or ADP-ribosylation.	ZDB-HMR-
ligand_binding_site (SO:0001657)	A binding site that, in the molecule, interacts selectively and non-covalently with a small molecule such as a drug, or hormone.	ZDB-LIGANDBS-
locus_control_region (SO:0000037)	A DNA region that includes DNase hypersensitive sites located 5' to a gene that confers the high-level, position-independent, and copy number-dependent expression to that gene.	ZDB-LCR-
modified_DNA_base (SO:0000305)	A modified nucleotide, i.e. a nucleotide other than A, T, C, G.	ZDB-MDNAB-
nc_conserved_region (SO:0000334)	Non-coding region of sequence similarity by descent from a common ancestor.	ZDB-NCCR-
nucleotide_binding_site (SO:0001655)	A binding site that, in the molecule, interacts selectively and non-covalently with nucleotide residues.	ZDB-NCBS-
nucleotide_motif (SO:0000714)	A region of nucleotide sequence corresponding to a known motif.	ZDB-NUCMO-
promoter (SO:0000167)	A regulatory_region composed of the TSS(s) and binding sites for TF_complexes of the basal transcription machinery.	ZDB-PROMOTER-
protein_binding_site (SO:0000410)	A binding site that, in the molecule, interacts selectively and non-covalently with polypeptide molecules.	ZDB-PROTBS-
regulatory_region (SO:0005836)	A region of sequence that is involved in the control of a biological process.	ZDB-RR-
RNA_motif (SO:0000715)	A motif that is active in RNA sequence.	ZDB-RNAMO-
TF_binding_site (SO:0000235)	A region of a nucleotide molecule that binds a Transcription Factor or Transcription Factor complex [GO:0005667].	ZDB-TFBS-
transcription_regulatory_region (SO:0001679)	A regulatory region that is involved in the control of the process of transcription.	ZDB-TRR-
translation_regulatory_region (SO:0001680)	A regulatory region that is involved in the control of the process of translation.	ZDB-TLNR-

^aAll SO terms, IDs, and definitions are from the Sequence Ontology (<http://www.sequenceontology.org/>) (6).

The screenshot shows the ZFIN search results page. At the top, there's a search bar with 'Gene / Transcript' entered and a 'Go' button. Below the search bar, there are filters for 'Gene / Transcript' and 'Type: Gene'. A 'Type' dropdown menu is open, showing a list of gene types with their counts: Gene (37,489), Protein Coding Gene (36,442), ncRNA Gene (516), miRNA Gene (426), lincRNA Gene (27), tRNA Gene (24), snoRNA Gene (22), lncRNA Gene (3), Pseudogene (420), and Engineered Foreign Gene (111). The 'ncRNA Gene' category is highlighted with a red box. Below the filter, the search results are displayed, showing 37,489 results sorted by Relevance. The first result is 'pax2a' (ZDB-GENE-990415-8) with details like Gene Name, Previous Names, and Location. The second result is 'EGFP' (ZDB-EFG-070117-1).

Figure 1. Non-coding genes in ZFIN's Single Box Search. ZFIN data can be searched or browsed using the single box search (22). To search for a specific gene, enter the gene symbol in the Search box on any ZFIN page. To browse all genes in ZFIN, as shown in this figure, click on the 'Go' next to the search box on the ZFIN home page to get to the Search page (<https://zfin.org/search?q=>), then select the 'Gene/Transcript' category, and then select 'Gene' from the 'Type' filter. To further limit the Gene results list to non-coding genes, select 'ncRNA Gene' or one of the more specific ncRNA gene subtypes in the 'Type' filter.

Table 2. Non-coding gene types supported in ZFIN

Sequence Ontology (SO) term name (ID)	Sequence Ontology term definition ^a	New ZFIN ID Prefix	Examples in ZFIN (gene name, gene symbol, ZFIN gene ID)
lincRNA_gene (SO:0001641)	A gene that encodes a long, intervening non-coding RNA.	ZDB-LINCRNAG-	<i>linc RNA nlrp12</i> <i>linc-nlrp12</i> ZDB-LINCRNAG-170802-1
lncRNA_gene (SO:0002127)	A gene that encodes a long non-coding RNA.	ZDB-LNCRNAG-	<i>lnc RNA kalrna</i> <i>lnc-kalrna</i> ZDB-LNCRNAG-180126-1
miRNA_gene (SO:0001265)	-	ZDB-MIRNAG-	<i>microRNA 1-1</i> <i>mir1-1</i> ZDB-MIRNAG-041228-15
ncRNA_gene (SO:0001263)	A gene that encodes a non-coding RNA.	ZDB-NCRNAG-	<i>RNA, U6 small nuclear 33</i> <i>rnu6-33</i> ZDB-NCRNAG-170419-1
piRNA_gene (SO:0001638)	A gene that encodes for an piwi associated RNA.	ZDB-PIRNAG	<i>none</i>
rRNA_gene (SO:0001637)	A gene that encodes for ribosomal RNA.	ZDB-RRNAG-	<i>RNA ribosomal 5.8S</i> <i>rna5-8s</i> ZDB-RRNAG-180607-1
scRNA_gene (SO:0001266)	-	ZDB-SCRNAG	<i>RNA, 7SL, cytoplasmic 1</i> <i>rn7sl1</i> ZDB-SCRNAG-171018-1
snoRNA_gene	-	ZDB-SNORNAG-	<i>small nucleolar RNA, C/D box 44</i> <i>snord44</i> ZDB-SNORNAG-120309-3
SRP_RNA_gene (SO:0001269)	-	ZDB-SRPRNAG-	<i>none</i>
tRNA_gene (SO:0001272)	-	ZDB-TRNAG-	<i>tRNA alanine, mitochondrial</i> <i>mt-ta</i> ZDB-TRNAG-011205-25

^aAll SO terms, IDs, and definitions are from the Sequence Ontology (<http://www.sequenceontology.org/>) (6).

Table 3. Examples of GO annotations and their annotation extensions

Gene	Primary GO term	Evidence code ^a	Annotation extension	Ref
<i>mcm1bp</i> (ZDB-GENE-030131-9676)	chromatin binding (MF; GO:0003682)	ISS	<i>happens_during</i> mitotic S phase (GO:0000084)	(10)
<i>her8a</i> (ZDB-GENE-030131-2376)	DNA-binding transcription factor activity, RNA polymerase II-specific (MF; GO:0000981)	IDA	<i>part_of</i> negative regulation of transcription by RNA polymerase II (GO:0000122)	(8)
<i>mcm1bp</i> (ZDB-GENE-030131-9676)	DNA-dependent DNA replication (BP; GO:0006261)	ISS	<i>happens_during</i> mitotic S phase (GO:0000084)	(10)
<i>gna11b</i> (ZDB-GENE-041121-9)	G-protein coupled acetylcholine receptor signaling pathway (BP; GO:0007213)	ISS	<i>occurs_in</i> pinealocyte (CL:0000652)	(10)
<i>her8a</i> (ZDB-GENE-030131-2376)	negative regulation of transcription by RNA polymerase II (BP; GO:0000122)	IMP	<i>has_input</i> dtx1 (ZDB-GENE-130503-1)	(8)
<i>fscn1a</i> (ZDB-GENE-031009-1)	receptor internalization (BP; GO:0031623)	IMP	<i>part_of</i> positive regulation of gene expression (GO:0010628)	(23)
<i>axin1</i> (ZDB-GENE-000403-1)	negative regulation of canonical Wnt signaling pathway (BP; GO:0090090)	IMP	<i>part_of</i> neural plate anterior/posterior regionalization (GO:0021999)	(9)
<i>gna11b</i> (ZDB-GENE-041121-9)	photoreceptor outer segment (CC; GO:0001750)	ISS	<i>part_of</i> retina (UBERON:0000966)	(10)
<i>cldnd</i> (ZDB-GENE-010328-4)	bicellular tight junction (CC; GO:0005923)	ISS	<i>part_of</i> liver (UBERON:0002107) <i>part_of</i> bile canaliculus (UBERON:0001283)	(10)

^aInferred from Sequence or structural Similarity (ISS), Inferred from Direct Assay (IDA), Inferred from Mutant Phenotype (IMP).

After the release of GRCz11, maintenance of the zebrafish reference sequence was transferred to ZFIN. ZFIN will undertake future curation of the zebrafish assembly in response to user-provided reports of specific assembly issues. User reports for specific issues can be submitted to the GRC using the ‘Report a Genome Problem’ form at the GRC, available at <https://www.ncbi.nlm.nih.gov/grc/report-an-issue>. In addition, genome assembly problems can be submitted directly to ZFIN via email to zfinadm@zfin.org. Future updates to the zebrafish genome assembly are likely to be provided as patches. Patches are alternate sequence scaffolds that are aligned to the primary assembly and leave the primary assembly chromosomal coordinates unchanged (<https://www.ncbi.nlm.nih.gov/grc/help/patches/>). It is important to note that ZFIN’s reference sequence maintenance will not extend to the curation or reannotation of gene models.

Genome versions in ZFIN

ZFIN is now using GRCz11 as the primary zebrafish genome assembly. In addition to GRCz11, two previous versions of the genome assembly, GRCz10 and Zv9, are available as alternate data sources in ZFIN’s implementation of the genome browser GBrowse (13). The version of the genome assembly can be selected in GBrowse using a pull-down menu at the top left of the ZFIN GBrowse web page. Some data tracks in GBrowse are available only on older genome assemblies; for example, the locations of the viral insertion mutants from the Burgess and Lin labs (14) are currently only available as a Zv9 GBrowse track, reflecting the location data provided to ZFIN by the submitters. The location of mutations in the Sanger Institute large-scale Zebrafish Mutation Resource (ZMP) mutagenesis screen (15) had originally been mapped to Zv9, and subsequently updated to GRCz10 and GRCz11. Note that due

to changes in the genome assemblies, not all ZMP alleles can be mapped to newer assemblies. Although most ZMP mutants are mapped to all three genome assemblies, fewer than 0.3% are mapped only to Zv9.

The location of knockdown reagents such as morpholinos, TALENs, and CRISPRs are mapped to the genome by ZFIN using Bowtie (16,17). They can be visualized in the ‘Knockdown Reagents’ track in GBrowse for all three genome assemblies. Tracks present in ZFIN GBrowse for Zv9, GRCz10 and GRCz11 are summarized in Table 4.

The Alliance of Genome Resources

The Alliance of Genome Resources (<https://alliancegenome.org>) was founded by six model organism databases (MODs: ZFIN, Mouse Genome Informatics Database (<http://informatics.jax.org/>), Rat Genome Database (<https://rgd.mcg.edu/>), FlyBase (<http://flybase.org/>), WormBase (<https://wormbase.org/>), Saccharomyces Genome Database (<https://yeastgenome.org/>) and the Gene Ontology Consortium (<http://geneontology.org/>). The Alliance, a centralized and integrated platform for human and model organism data, serves the biological and medical research communities and facilitates translational research (18). The initial public release of the Alliance portal in October 2017 included data on genes, functional annotations, orthology, and human disease associations (19). Since then, additional releases have added data types such as phenotypes, alleles, molecular interactions, computed gene descriptions and links to expression data. Upcoming releases will include wild-type gene expression annotations, ortholog-based model organism gene to human disease associations, and tables of molecular interactions on gene pages.

Table 4. ZFIN GBrowse tracks on zebrafish genome assemblies

ZFIN GBrowse tracks	Zebrafish genome build		
	Zv9	GRCz10	GRCz11
Genes	x	x	x
Genes with expression	x	x	x
Genes with phenotype	x	x	x
Genes with antibodies	x	x	x
Ensembl transcripts	x	x	x
Vega transcripts	x	x	x
Knockdown reagents	x	x	x
Transgenic insertions	x		
Assembly	x	x	x
Complete assembly clones	x	x	x
dbSNP	x	x	x
Zebrafish Mutation Project alleles	x	x	
Zebrafish Mutation Project – available to purchase	x		x
Restriction sites	x	x	x
Conserved nongenic elements	x		
NHGRI-1 invariant	x		
NHGRI-1 stopgain	x		

Alliance and MOD gene pages are reciprocally linked. Researchers can travel from ZFIN gene pages to the corresponding Alliance gene page by clicking on the ‘Alliance’ link in the ‘OTHER GENE PAGES’ section. The ZFIN identifier in the top left corner of each Alliance zebrafish gene page links to the ZFIN gene page. Alliance gene pages provide novel functionalities not found in ZFIN, such as a JBrowse (20,21) sequence feature viewer (the GBrowse viewer is supported at ZFIN), high-level graphical summaries of Gene Ontology annotations, and a comparative table of computed and curated orthologs.

ZFIN technical implementation

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 a PostgreSQL relational database (migrated from Informix in 2017). 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.atlassian.com/software/confluence/>). ZFIN supports an InterMine instance, ZebrafishMine, (<http://ZebrafishMine.org>; <https://github.com/ZebrafishMine/intermine>), which also serves ZFIN data via web services. A detailed and browsable view of the current ZFIN data model can be found at <http://zfin.org/schemaSpy/index.html>.

FUNDING

National Human Genome Research Institute at the US National Institutes of Health [U41 HG002659]. Funding for open access charge: The NHGRI grant that funds ZFIN.

Conflict of interest statement. None declared

REFERENCES

1. Van Slyke, C.E., Bradford, Y.M., Howe, D.G., Fashena, D.S., Ramachandran, S., Ruzicka, L. and ZFIN Staff. (2018) Using ZFIN: data types, organization, and retrieval. *Methods Mol. Biol.*, **1757**, 307–347.
2. 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.
3. Sprague, J., Bayraktaroglu, L., Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Haendel, M., Howe, D.G., Knight, J. *et al.* (2008) The Zebrafish Information Network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Res.*, **36**, D768–D772.
4. 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.* (2017) The Zebrafish Model Organism Database: new support for human disease models, mutation details, gene expression phenotypes and searching. *Nucleic Acids Res.*, **45**, D758–D768.
5. Bradford, Y.M., Toro, S., Ramachandran, S., Ruzicka, L., Howe, D.G., Eagle, A., Kalita, P., Martin, R., Taylor Moxon, S.A., Schaper, K. *et al.* (2017) Zebrafish models of human disease: Gaining insight into human disease at ZFIN. *ILAR J.*, **58**, 4–16.
6. 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.
7. The Gene Ontology Consortium. (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.*, **45**, D331–D338.
8. Cheng, Y.-C., Huang, Y.-C., Yeh, T.-H., Shih, H.-Y., Lin, C.-Y., Lin, S.-J., Chiu, C.-C., Huang, C.-W. and Jiang, Y.-J. (2015) Deltex1 is inhibited by the Notch–Hairy/E(Spl) signaling pathway and induces neuronal and glial differentiation. *Neural Dev.*, **10**, 28.
9. van de Water, S., van de Wetering, M., Joore, J., Esseling, J., Bink, R., Clevers, H. and Zivkovic, D. (2001) Ectopic Wnt signal determines the eyesh phenotype of zebrafish masterblind mutant. *Development*, **128**, 3877–3888.
10. GOA, HGNC, AgBase, UniProtKB (2007) Manual transfer of experimentally-verified manual GO annotation data to orthologs by curator judgment of sequence similarity. *ZFIN Direct Data Submission*, <https://zfin.org>.
11. Huntley, R.P., Harris, M.A., Alam-Faruque, Y., Blake, J.A., Carbon, S., Dietze, H., Dimmer, E.C., Foulger, R.E., Hill, D.P., Khodiyar, V.K. *et al.* (2014) A method for increasing expressivity of Gene Ontology annotations using a compositional approach. *BMC Bioinformatics*, **15**, 155.
12. 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.
13. Stein, L.D. (2013) Using GBrowse 2.0 to visualize and share next-generation sequence data. *Brief. Bioinform.*, **14**, 162–171.
 14. Varshney, G.K., Lu, J., Gildea, D.E., Huang, H., Pei, W., Yang, Z., Huang, S.C., Schoenfeld, D., Pho, N.H., Casero, D. *et al.* (2013) A large-scale zebrafish gene knockout resource for the genome-wide study of gene function. *Genome Res.*, **23**, 727–735.
 15. Kettleborough, R.N.W., Busch-Nentwich, E.M., Harvey, S.A., Dooley, C.M., de Bruijn, E., van Eeden, F., Sealy, I., White, R.J., Herd, C., Nijman, I.J. *et al.* (2013) A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature*, **496**, 494–497.
 16. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods. Nature Publishing Group*, **9**, 357–359.
 17. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
 18. Howe, D.G., Blake, J.A., Bradford, Y.M., Bult, C.J., Calvi, B.R., Engel, S.R., Kadin, J.A., Kaufman, T.C., Kishore, R., Lauderkind, S.J.F. *et al.* (2018) Model organism data evolving in support of translational medicine. *Lab. Anim. (NY)*, **47**, 277–289.
 19. Skrzypek, M.S., Nash, R.S., Wong, E.D., MacPherson, K.A., Hellerstedt, S.T., Engel, S.R., Karra, K., Weng, S., Sheppard, T.K., Binkley, G. *et al.* (2018) Saccharomyces genome database informs human biology. *Nucleic Acids Res.*, **46**, D736–D742.
 20. Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
 21. Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J. and Holmes, I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
 22. 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.
 23. Liu, Z., Ning, G., Xu, R., Cao, Y., Meng, A. and Wang, Q. (2016) Fscn1 is required for the trafficking of TGF- β family type 1 receptors during endoderm formation. *Nat. Commun.*, **7**, 12603.