ZFIN, the knowledgebase for *Danio rerio* research

Zebrafish Information Network, the knowledgebase for *Danio rerio* research

Yvonne M. Bradford¹, Ceri E. Van Slyke¹, Leyla Ruzicka¹, Amy Singer¹, Anne Eagle¹, David Fashena¹, Douglas G. Howe¹, Ken Frazer¹, Ryan Martin¹, Holly Paddock¹, Christian Pich¹, Sridhar Ramachandran¹, Monte Westerfield¹

¹The Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403-1254, USA

*To whom correspondence should be addressed:

Yvonne M. Bradford
The Zebrafish Information Network (ZFIN)
5291 University of Oregon
Eugene, OR 97403-5291
ybradford@zfin.org

Keywords: Danio rerio, zebrafish, knowledgebase, database

© The Author(s) (2022). Published by Oxford University Press on behalf of the Genetics Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contactjournals.permissions@oup.com
Abstract

The Zebrafish Information Network (ZFIN, zfin.org) is the central repository for Danio rerio genetic and genomic data. ZFIN has served the zebrafish research community since 1994, expertly curating, integrating, and displaying zebrafish data. Key data types available at ZFIN include, but are not limited to, genes, alleles, human disease models, gene expression, phenotype, and gene function. ZFIN makes zebrafish research data Findable, Accessible, Interoperable, and Reusable through nomenclature, curatorial and annotation activities, web interfaces, and data downloads. Recently, ZFIN and six other model organism knowledgebases have collaborated to form the Alliance of Genome Resources, aiming to develop sustainable genome information resources that enable the use of model organisms to understand the genetic and genomic basis of human biology and disease. Here we provide an overview of the data available at ZFIN including recent updates to the gene page to provide access to single-cell RNA sequencing (scRNA-seq) data, links to Alliance web pages, ribbon diagrams to summarize the biological systems and Gene Ontology terms that have annotations, and data integration with the Alliance of Genome Resources.

Introduction

The Zebrafish Information Network (ZFIN, zfin.org) is the knowledgebase for the model organism Danio rerio (zebrafish). Since 1994, ZFIN has served the zebrafish research community by collecting, integrating, and making available zebrafish data. ZFIN’s biocurators, who have expertise in genetics, developmental, cellular, molecular, and evolutionary biology, have annotated over 16,000 zebrafish research publications for data that include genes, gene function, sequences, alleles, mutant and transgenic lines, human disease models, gene expression, phenotype, orthology, sequence targeting reagents, and antibodies. Additionally, ZFIN supports the zebrafish research community by providing wiki resources to view antibody and protocol information, as well as ZFIN pages for researchers, laboratories, and companies.

ZFIN is the nomenclature authority for zebrafish genes and alleles, and provides official nomenclature support for the zebrafish community. This facilitates the discovery and knowledge integration of gene data, making these data accessible and reusable. The ZFIN Nomenclature Coordinator provides a core service of coordinating nomenclature with the HUGO Gene Nomenclature Committee (HGNC, Povey et al. 2001) and the
ZFIN aims to make zebrafish research data Findable, Accessible, Interoperable, and Reusable (FAIR, Wilkinson *et al.* 2016) by contributing to data annotation standards, the use of biomedical ontologies for data annotations, and persistent identifiers for annotations and metadata, as well as making data freely available at ZFIN (Köhler *et al.* 2013, Vasilevsky *et al.* 2013, Van Slyke *et al.* 2018). ZFIN has provided guidance to the zebrafish scientific community on minimum requirements for data submissions to ensure accurate integration with other data as well as encouraging FAIR standards (Howe *et al.* 2011, Howe *et al.* 2016). In an effort to support FAIR data principles further, ZFIN and the Mouse Genome Database (MGD, Bult *et al.* 2019), the Rat Genome Database (RGD, Laulederkind *et al.*), WormBase (Lee *et al.* 2018), FlyBase (Thurmond *et al.* 2019), and the Gene Ontology Consortium (Carbon *et al.* 2019) have collaborated to create the Alliance of Genome Resources (Alliance, Bult *et al.* 2019a, Agapite *et al.* 2020). The Alliance aims to provide a centralized multi-species platform that provides access to integrated, harmonized model organism data that facilitate the use of model organisms to understand the genetic and genomic basis of human biology and disease (Howe *et al.* 2018). Working in conjunction with the Alliance, ZFIN has continued work on data standardization, incorporating new data visualizations on the gene page, as well as automated textual gene descriptions, and providing zebrafish researchers access to multispecies gene expression, phenotype, human disease, and orthology data.
Gene Page

Summary section update

The Gene page integrates and displays data pertinent to the gene, providing a comprehensive and current understanding of gene expression, mutant phenotypes, associated diseases, gene functions, alleles, and orthology. The layout and functionality of the gene page was recently updated to provide better navigation and usability (Howe et al. 2021a). The ZFIN gene page has a layout similar to the Alliance gene page, providing users a comparable view and functionality that is helpful when navigating between sites. At the top of the gene page, the gene summary section provides basic information about the gene including symbol, links to nomenclature history, name, previously used aliases, textual gene description, and links to gene pages at other databases (Figure 1). The textual gene description is new to the summary section of the gene page and is computationally generated and provided by the Alliance (Kishore et al. 2020). In addition, comparative information was recently added to the summary section, displaying a hexagonal icon that represents the model organism databases of the Alliance. Clicking the icon directs users to the orthology section of the corresponding Alliance gene page, providing quick access to homologous gene information. The summary section is followed by sections that present data pertaining to the gene. A navigation pane on the left side of the page allows quick access to the data sections of interest. Currently, ZFIN contains data on 37,466 genes that are categorized as protein coding genes, ncRNA genes, or pseudogenes.

Gene Expression Section Update

The ‘Expression’ section on the Gene page focuses primarily on the wild-type (WT) expression of the gene. This section begins with links to: 1) ‘All Expression Data’ which links to a summary of gene expression in both WT and mutant fish; 2) external sites with cross-species comparison data; 3) links to high throughput data sets; and 4) Thisse large scale WT screen (Thisse et al. 2001) (Figure 2). Recently links to UCSC scRNAseq data were added to the high throughput data section, providing users access to the new single cell RNA-seq atlas of zebrafish gene expression during organogenesis (Farnsworth et al. 2020). Single cell RNA-seq is a technique that has been increasingly used in zebrafish to understand transcriptional changes at the cellular level during development (Tang et al. 2017, Farrell et al. 2018, Wagner et al. 2018). These data, along with traditional in situ data that ZFIN annotates, help provide an understanding of changes in gene expression during development. The Wild Type Expression Summary
provides a high-level visual summary ribbon that denotes the anatomical systems, stages, and GO cellular components that have annotations and is discussed in more detail in the following section.

ZFIN curates gene expression using gene symbols, genotypes and strains, the Zebrafish Experimental Conditions Ontology (ZECO, Bradford et al. 2016), the Zebrafish Anatomical Ontology (ZFA, van Slyke et al. 2014), the Zebrafish Stage ontology (ZFS, van Slyke et al. 2014), GO cellular compartment ontology (GO-CC, Ashburner et al. 2000, Carbon et al. 2021), and Spatial ontology (BSPO, Dahdul et al. 2014). Through the use of metadata identifiers and annotation standardization, ZFIN gene expression annotations comply with FAIR standards. ZFIN currently has 14,350 genes with expression data and has curated 216,696 zebrafish gene expression assays; of those 134,233 are in wild-type backgrounds with standard or control conditions. In addition, ZFIN has curated 31,964 transgenic reporter assays.

Expression, Phenotype, Gene Ontology Ribbons

The Expression, Phenotype, and Gene Ontology (GO) sections of the gene page utilize a ribbon diagram to visually summarize which anatomical systems, developmental stages, or GO biological process, molecular function, or cellular compartment terms have annotations (Figures 2, 3, and 4). The ribbon diagrams were recently added to ZFIN gene pages and are adapted from the ribbon diagrams displayed on Alliance gene pages (Howe et al. 2021). The terms used in the expression and phenotype summary ribbon were chosen to support grouping of annotations and are either hierarchically high-level ontology terms that subsume many biological concepts or terms that have been used in many annotations in ZFIN. For example, the anatomy portion of the ribbon summary uses high level anatomical system terms from the ZFA ontology that have child terms that describe the anatomical structures that are part of the indicated anatomical system. The ribbon summary also lists terms that may not be child terms of an anatomical system but are of particular interest to zebrafish researchers and have a high number of expression or phenotype annotations. These include ‘fin’, ‘integument’, ‘neural tube’, ‘primary germ layer’, and ‘somite’. The ‘Other Structures’ term groups all other annotations that use terms that are not child terms of an anatomical system.

The terms used for the GO ribbon are high-level terms that were chosen through a collaboration with the GO Consortium and the Alliance based on annotation coverage for model species represented at the Alliance. These high-level terms cover broad aspects of the biological process, molecular function, and cellular component branches of GO and have a subset tag, ‘goslim_agr’, to specify that the term is a member of the GO Alliance subset. A subset of an ontology, or slim, is a trimmed down version of the
ontology that contains a subset of terms and is useful for providing a high-level view of the content of an ontology. The Alliance subset is named ‘GO slim AGR subset’ and can be downloaded at http://geneontology.org/docs/download-ontology/#subsets. The expression, phenotype, and GO ribbon diagrams have similar functionality, with shaded boxes denoting the presence of annotations for the indicated system, stage, biological process, molecular function, or cellular compartment. Clicking on the shaded box opens a table view that provides more detailed annotations.

Phenotype Section Update

The ‘Phenotype’ section of the Gene page reports the phenotype of mutant and gene knockdown fish (Figure 3). This section begins with a link to all phenotype data, which directs users to a Phenotype Figure Summary page that lists the phenotype annotations with associated figures and publications for a gene. A new Cross-Species Comparison section provides a link to the phenotype section of the Alliance gene page. The Phenotype Summary provides a high-level ribbon overview of the anatomical systems, stages, molecular functions, and biological processes that have phenotype annotations.

Phenotype annotations provide core insights into gene function and are utilized to facilitate the understanding of disease processes and outcomes. ZFIN annotates phenotype data using the Phenotype and Trait Ontology (PATO, Gkoutos et al. 2005), ZFA, GO, and Chemical Entities of Biological Interest Ontology (ChEBI, Hastings et al. 2016). A phenotype annotation is composed of the mutant or knockdown fish, the experimental conditions, the stage at which the phenotype is observed, the phenotype statement, and the associated publication. Phenotype statements are composed of entities and qualities assembled in the E+Q syntax (Washington et al. 2009) using the aforementioned ontologies. To facilitate cross-species phenotype integration, ZFIN participates as a core member of the UPheno initiative, which aims to reconcile logical definitions across several model organism phenotype ontologies (Matentzoglu et al. 2018). This reconciliation work will benefit the Alliance as it moves toward providing ribbon diagrams for cross-species phenotype comparisons. ZFIN has curated 52,305 phenotype statements for single gene mutant or knockdown fish under standard or control conditions. There are 4,895 genes represented by 7,055 alleles with mutant phenotypes. Additionally, there are 2241 genes that have no mutant alleles but have MO induced phenotypes and 245 genes that have phenotypes reported in animals injected with CRISPRs (CRISPants).
Mutations Section Update

The ‘Mutations’ section of the gene page is intended to provide a high-level summary of the alleles and knockdown reagents that have been used to investigate gene function. As part of the gene page update, the tables in this section were updated from lists to data tables that provide more data and allow users to gain a better understanding of the allele or knockdown reagent without having to go to another web page for more information (Howe et al. 2021b). The Mutations section has two tables, the ‘Mutants’ table, which summarizes information for the alleles of a gene, and ‘Sequence Targeting Reagents’ (STR) table, which lists curated gene knockdown reagents such as morpholinos, CRISPRs, and TALENs. ZFIN facilitates FAIR standards for STR reagents by obtaining STR sequence information from publications, providing distinct nomenclature and identifiers, and reporting this information via user interfaces and download files. In addition, STR sequence targets are verified, and if errors are found authors are contacted to make corrections. Currently ZFIN has records for 10,763 Morpholinos, 6,872 CRISPRs, and 813 TALENs. As the nomenclature authority for allele designations, ZFIN has made allele associations for 19,319 genes and has records for 53,522 alleles.

Human Disease Section Update

The ‘Human Disease’ section on the Gene page provides two types of disease associations, diseases associated to zebrafish genes via orthology to human disease-causing genes and experimentally verified disease models (Figure 4). The ‘Associated with the Human Ortholog’ table lists Disease Ontology terms (DO, Schriml et al. 2019), which links to the ZFIN DO term page where more information can be found about the disease, as well as the OMIM disease name and links to the corresponding OMIM page (Amberger et al. 2015). Recently, links to the Alliance Disease page were added to this table, directing users to the Alliance disease page that provides cross-species information about genes and alleles associated to disease as well as experimental models of disease. Information in the ‘Associated with Human Ortholog’ table is produced by computational mappings of ZFIN curated orthologs to human genes and their disease associations from the genemap and mim2gene files from OMIM (https://omim.org/downloads/). This zebrafish gene to human gene/disease mapping is used to make associations between the ZFIN gene and the DO terms via DO term to OMIM disease mappings in the DO file (Bradford et al. 2017). The ‘Associated Experimental Models’ table lists curated experimental models of human disease. These models have been experimentally verified to model some or all aspects of a human disease. ZFIN currently has 1,900 curated disease models and 3,606 diseases.
ZFIN, the knowledgebase for *Danio rerio* research

associated with 5,154 zebrafish genes via orthology with 4,078 human genes. Individual DO term pages can be found via the search at the top of any ZFIN page.

**Gene Ontology Section**

The ‘Gene Ontology’ section of the gene page reports the roles of a gene or gene product as described by the Gene Ontology (GO) term annotations for the gene. ZFIN manually produces GO annotations while curating publications (Sprague *et al.* 2008). These annotations are created by associating a gene or gene product with GO terms that describe molecular functions (MF) that the gene product enables, the biological process (BP) the gene product is involved in, and the cellular component (CC) where the gene product performs its function. ZFIN also downloads and displays GO annotations that are produced based on electronic annotation from InterPro2GO and UniProt (Finn *et al.* 2017, Dimmer *et al.* 2012). The GO section was recently updated to include a high-level summary ribbon view of the molecular function, biological process, and cellular component annotations for a gene. Clicking on shaded boxes opens a table with more detailed GO annotations that include the annotated GO term, evidence code, with/from field, and citation (Figure 5). ZFIN has 23,047 genes with GO annotations.

**Orthology and Other Data sections**

The ‘Orthology’ section of the gene page provides an overview of the orthology relationships between zebrafish, human, and mouse genes. An expert curator using conserved synteny, gene family tree analysis, and amino acid alignments has vetted orthology displayed in this section. The primary resources used in the process of determining orthology and gene nomenclature are Ensembl (Howe *et al.* 2021a), NCBI (Agarwala *et al.* 2018), and Panther (Thomas *et al.* 2003). This level of orthology verification is necessary due to the whole genome duplication in the teleost lineage (Postlethwait *et al.* 2004) that causes many computational algorithms used for orthology to misidentify co-orthologs of mammalian genes. This section has new links to the orthology section of the zebrafish gene pages at the Alliance where the displays highlight computationally derived orthology and paralogy.

The other sections on the gene page provide additional information about the gene and links to related data both at ZFIN and other sites. The Protein Domains section contains two tables which provide data from InterPro (Blum *et al.* 2021). The Domain, Family, and Site Summary table lists protein binding sites, domains, and families with links to InterPro and the Domain Details Per Protein table denotes which domain/site/family is
ZFIN, the knowledgebase for *Danio rerio* research

associated with each specific Uniprot entry (Bateman *et al.* 2021). The transcript section
links to several genome browsers and displays transcript diagrams, links to ZFIN
Transcript Pages, and withdrawn transcripts. ZFIN links to SignaFish for interactions
when known (Csályi *et al.* 2016). Antibodies targeting the gene, plasmids with the
coding region at Addgene, and constructs containing the gene are linked in their
respective sections. There is also a section that links BACs, ESTs, and cDNAs
associated with the gene as well as a section that links to sequences at GenBank and
UniProt.

**ZFIN data at the Alliance**

ZFIN is a founding member of the Alliance of Genome Resources (alliancegenome.org,
*Bult et al.* 2019a) which has the primary mission of developing and maintaining
comparative genome information resources that facilitate the use of multiple model
organisms to understand the genetic and genomic basis of human biology and disease.

As of the Alliance 4.1 release (August 2021, see Agapite *et al.* this issue), ZFIN
contributes the following data to the Alliance: genome data, wild-type expression,
phenotype data, mutant and transgenic alleles, variants, disease models, and orthology.

Through the integration of model organism gene data, the Alliance provides
comprehensive gene orthology data. The Alliance utilizes ZFIN data, and other model
organism database data, to produce species-specific genome browsers that display the
genome, genes, and variants. In addition, the Alliance uses ZFIN data to produce
zebrafish-specific gene and allele pages, and provides links to data pages at ZFIN
including the gene, allele, fish, and disease pages. ZFIN’s InterMine-based data mining
database, ZebrafishMine, has been replaced by the InterMine instance developed at the
Alliance, AllianceMine (<https://www.alliancegenome.org/alliancemine>) (Smith *et al.*
2012, Kalderimis *et al.* 2014, Van Slyke *et al.* 2018). AllianceMine has gene, orthology,
allele, gene ontology, and disease ontology data and search templates for zebrafish and
other model organisms. AllianceMine will continue to be updated with additional ZFIN
data and data types with each Alliance release. As discussed previously, reciprocal
links are provided from ZFIN pages to appropriate Alliance data pages. The Alliance
makes ZFIN data available along with data from all the participating model organism
databases via either swagger APIs from the Alliance
(<https://www.alliancegenome.org/api/swagger-ui/>) or as download files
(<https://www.alliancegenome.org/downloads>).
ZFIN technical implementation

The ZFIN technical stack is substantially unchanged since our last publication (Howe et al 2021). In brief the ZFIN web architecture is primarily written in Java using the Spring Framework and served by JSP in Apache/Tomcat. The ZFIN user interface is implemented using React, GWT, Angular, jQuery, and plain JavaScript. Groovy, SQL, and Perl are used to process and load bulk data. Hibernate serves as the object-relational mapping library from Java to the relational PostgreSQL database. ZFIN uses both Solr and Java/Spring to support our search interfaces. Data from papers are entered via a web-based curation interface primarily written in GWT with a few implementations of AngularJS interfaces. The community wiki is powered by Atlassian Confluence software (http://www.atlassian.com/software/confluence/). A detailed and browsable view of the current ZFIN data model can be found at http://zfin.org/schemaSpy/index.html. Our current plan is to move from our servers to cloud servers to provide dynamic response to load and smooth our server costs.

Data Accessibility

As the knowledgebase for the zebrafish model organism research community, one of the main goals for ZFIN is to make the data that ZFIN annotates and integrates as accessible as possible. As discussed earlier, ZFIN makes integrated data available on the Gene Page. ZFIN also produces data-specific web pages for alleles, fish, antibodies, constructs, STR, clones, publications, and ontology terms where information that is more detailed can be found for each data type. In addition, data at ZFIN can be searched using the single box search as well as dedicated search forms for specific data types. For computational exploration of data, ZFIN maintains a downloads page that provides access to data-specific download files (https://zfin.org/downloads) and archives https://zfin.org/downloads/archive. ZFIN also can provide custom download files upon request. Data provided in download files are FAIR compliant denoting the date the file was generated, as well as all identifiers, symbols, names, and ontology terms. As previously noted, ZFIN data are also available via AllianceMine, API, and download files at the Alliance.

Funding

National Human Genome Research Institute at the US National Institutes of Health [U41 HG002659 (ZFIN) and U24 HG010859 (Alliance of Genome Resources)].
References

1. Agapite J., L. P. L.-P. Albou, S. Aleksander, J. Argasinska, V. Arnaboldi, et al., 2020
   Alliance of Genome Resources Portal: Unified model organism research platform. Nucleic Acids Res. 48: D650–D658. https://doi.org/10.1093/nar/gkz813

2. Agarwala R., T. Barrett, J. Beck, D. A. Benson, C. Bollin, et al., 2018 Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 46: D8–D13. https://doi.org/10.1093/nar/gkx1095

3. Amberger J. S., C. A. Bocchini, F. Schiettecatte, A. F. Scott, and A. Hamosh, 2015 OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. Nucleic Acids Res. 43: D789–D798. https://doi.org/10.1093/nar/gku1205

4. Ashburner M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, et al., 2000 Gene ontology: Tool for the unification of biology. Nat. Genet. 25: 25–29.

5. Bateman A., M. J. Martin, S. Orchard, M. Magrane, R. Agivetova, et al., 2021 UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Res. 49: D480–D489. https://doi.org/10.1093/nar/gkaa1100

6. Blum M., H.-Y. Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, et al., 2021 The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49: D344–D354. https://doi.org/10.1093/nar/gkaa977

7. Bradford Y. M., C. E. Van Slyke, S. Toro, and S. Ramachandran, 2016 The zebrafish experimental conditions ontology: Systemizing experimental descriptions in ZFIN, in CEUR Workshop Proceedings.

8. Bradford Y. M., S. Toro, S. Ramachandran, L. Ruzicka, D. G. Howe, et al., 2017 Zebrafish models of human disease: Gaining insight into human disease at ZFIN. ILAR J. 58: 4–16. https://doi.org/10.1093/ilar/ilw040

9. Bult C. J., J. A. Blake, B. R. Calvi, J. M. Cherry, V. DiFrancesco, et al., 2019a The alliance of genome resources: Building a modern data ecosystem for model organism databases. Genetics 213: 1189–1196.

10. Bult C. J., J. A. Blake, C. L. Smith, J. A. Kadin, J. E. Richardson, et al., 2019b Mouse Genome Database (MGD) 2019. Nucleic Acids Res. 47: D801–D806. https://doi.org/10.1093/nar/gky1056

11. Carbon S., E. Douglass, N. Dunn, B. Good, N. L. Harris, et al., 2019 The Gene
Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47: D330–D338. https://doi.org/10.1093/nar/gky1055

12. Carbon S., E. Douglass, B. M. Good, D. R. Unni, N. L. Harris, et al., 2021 The Gene Ontology resource: Enriching a GOld mine. Nucleic Acids Res. 49: D325–D334. https://doi.org/10.1093/nar/gkaa1113

13. Csályi K., D. Fazekas, T. Kadlecsik, D. Türei, L. Gul, et al., 2016 SignaFish: A Zebrafish-specific signaling pathway resource. Zebrafish 13: 541–544. https://doi.org/10.1089/zeb.2016.1277

14. Dahdul W. M., H. Cui, P. M. Mabee, C. J. Mungall, D. Osumi-Sutherland, et al., 2014 Nose to tail, roots to shoots: Spatial descriptors for phenotypic diversity in the Biological Spatial Ontology. J. Biomed. Semantics 5: 1–13. https://doi.org/10.1186/2041-1480-5-13

15. Dimmer E. C., R. P. Huntley, Y. Alam-Faruque, T. Sawford, C. O’Donovan, et al., 2012 The UniProt-GO Annotation database in 2011. Nucleic Acids Res. 40: D565. https://doi.org/10.1093/NAR/GKR1048

16. Farnsworth D. R., L. M. Saunders, and A. C. Miller, 2020 A single-cell transcriptome atlas for zebrafish development. Dev. Biol. 459: 100–108. https://doi.org/10.1016/j.ydbio.2019.11.008

17. Farrell J. A., Y. Wang, S. J. Riesenfeld, K. Shekhar, A. Regev, et al., 2018 Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science (80-. ). 360. https://doi.org/10.1126/science.aar3131

18. Finn R. D., T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, et al., 2017 InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 45: D190–D199. https://doi.org/10.1093/nar/gkw1107

19. Gkoutos G. V, E. C. J. Green, A. M. Mallon, J. M. Hancock, and D. Davidson, 2005 Using ontologies to describe mouse phenotypes. Genome Biol. 6. https://doi.org/10.1186/gb-2004-6-1-r8

20. Hastings J., G. Owen, A. Dekker, M. Ennis, N. Kale, et al., 2016 ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res. 44: D1214–D1219. https://doi.org/10.1093/nar/gkv1031

21. Howe D. G., K. Frazer, D. Fashena, L. Ruzicka, Y. Bradford, et al., 2011 Data extraction, transformation, and dissemination through ZFIN. Methods Cell Biol. 104: 311–325. https://doi.org/10.1016/B978-0-12-374814-0.00017-3

22. Howe D. G. G., Y. M. M. Bradford, A. Eagle, D. Fashena, K. Frazer, et al., 2016 A scientist’s guide for submitting data to ZFIN. Methods Cell Biol. 135: 451–481. https://doi.org/10.1016/bs.mcb.2016.04.010
23. Howe D. G., J. A. Blake, Y. M. Bradford, C. J. Bult, B. R. Calvi, et al., 2018 Model organism data evolving in support of translational medicine. Lab Anim. (NY). 47: 277–289. https://doi.org/10.1038/s41684-018-0150-4

24. Howe K. L., P. Achuthan, J. J. Allen, J. J. Allen, J. Alvarez-Jarreta, et al., 2021a Ensembl 2021. Nucleic Acids Res. 49: D884–D891. https://doi.org/10.1093/nar/gkaa942

25. Howe D. G., S. Ramachandran, Y. M. Bradford, D. Fashena, S. Toro, et al., 2021b The zebrafish information network: Major gene page and home page updates. Nucleic Acids Res. 49: D1058–D1064. https://doi.org/10.1093/nar/gkaa1010

26. Kalderimis A., R. Lyne, D. Butano, S. Contrino, M. Lyne, et al., 2014 InterMine: extensive web services for modern biology. Nucleic Acids Res. 42. https://doi.org/10.1093/nar/gku301

27. Kishore R., V. Arnaboldi, C. E. Van Slyke, J. Chan, R. S. Nash, et al., 2020 Automated generation of gene summaries at the Alliance of Genome Resources. Database J. Biol. Databases Curation 2020. https://doi.org/10.1093/database/baaa037

28. Köhler S., S. C. Doelken, B. J. Ruef, S. Bauer, N. Washington, et al., 2013 Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Research 2: 30. https://doi.org/10.12688/f1000research.2-30.v1

29. Laulederkind S. J. F., G. T. Hayman, S. J. Wang, J. R. Smith, V. Petri, et al., 2018 A primer for the rat genome database (RGD), pp. 163–209 in Methods in Molecular Biology.

30. Lee R. Y. N., K. L. Howe, T. W. Harris, V. Arnaboldi, S. Cain, et al., 2018 WormBase 2017: Molting into a new stage. Nucleic Acids Res. 46: D869–D874. https://doi.org/10.1093/nar/gkx998

31. Matentzoglu N., J. P. Balhoff, S. M. Bello, C. F. Boerkoel, Y. M. Bradford, et al., 2018 Phenotype Ontologies Traversing All The Organisms (POTATO) workshop aims to reconcile logical definitions across species. https://doi.org/10.5281/ZENODO.2382757

32. Postlethwait J., A. Amores, W. Cresko, A. Singer, and Y. L. Yan, 2004 Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet. 20: 481–490. https://doi.org/10.1016/j.tig.2004.08.001

33. Povey S., R. Lovering, E. Bruford, M. Wright, M. Lush, et al., 2001 The HUGO Gene Nomenclature Committee (HGNC). Hum. Genet. 109: 678–680.

34. Schriml L. M., E. Mitraka, J. Munro, B. Tauber, M. Schor, et al., 2019 Human Disease Ontology 2018 update: Classification, content and workflow expansion.
Nucleic Acids Res. 47: D955–D962. https://doi.org/10.1093/nar/gky1032

35. Slyke C. E. Van, Y. M. Bradford, M. Westerfield, and M. A. Haendel, 2014 The zebrafish anatomy and stage ontologies: Representing the anatomy and development of Danio rerio. J. Biomed. Semantics 5. https://doi.org/10.1186/2041-1480-5-12

36. Slyke C. E. Van, Y. M. Bradford, D. G. Howe, D. S. Fashena, S. Ramachandran, et al., 2018 Using ZFIN: Data types, organization, and retrieval, pp. 307–347 in Methods in Molecular Biology, NIH Public Access.

37. Smith R. N., J. Aleksic, D. Butano, A. Carr, S. Contrino, et al., 2012 Databases and ontologies InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. 28: 3163–3165. https://doi.org/10.1093/bioinformatics/bts577

38. Sprague J., L. Bayraktaroglu, Y. Bradford, T. Conlin, N. Dunn, et al., 2008 The Zebrafish Information Network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. Nucleic Acids Res. 36: D768-72. https://doi.org/10.1093/nar/gkm956

39. Tang Q., S. Iyer, R. Lobbardi, J. C. Moore, H. Chen, et al., 2017 Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. J. Exp. Med. 214: 2875–2887. https://doi.org/10.1084/jem.20170976

40. Thisse B., S. Pflumio, M. Fürthauer, B. Loppin, V. Heyer, et al., 2001 Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). ZFIN Direct Data Submission.

41. Thomas P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, et al., 2003 PANTHER: A library of protein families and subfamilies indexed by function. Genome Res. 13: 2129–2141. https://doi.org/10.1101/gr.772403

42. Thurmond J., J. L. Goodman, V. B. Strelets, H. Attrill, L. S. Gramates, et al., 2019 FlyBase 2.0: The next generation. Nucleic Acids Res. 47: D759–D765. https://doi.org/10.1093/nar/gky1003

43. Vasilevsky N. A., M. H. Brush, H. Paddock, L. Ponting, S. J. Tripathy, et al., 2013 On the reproducibility of science: Unique identification of research resources in the biomedical literature. PeerJ 2013: e148. https://doi.org/10.7717/peerj.148

44. Wagner D. E., C. Weinreb, Z. M. Collins, J. A. Briggs, S. G. Megason, et al., 2018 Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science (80-. ). 360: 981–987. https://doi.org/10.1126/science.aar4362

45. Washington N. L., M. A. Haendel, C. J. Mungall, M. Ashburner, M. Westerfield, et al., 2009 Linking human diseases to animal models using ontology-based...
Figure 1. The ZFIN gene page. Web page providing information as it relates to a gene. The top portion provides general information about the gene, and the left navigation panel allows users to navigate directly to areas of interest. Arrows indicate new links or information provided by the Alliance.

Figure 2. Expression section of the Gene page. Section of the gene page providing information about where and when the gene is expressed. Links to all expression data, Alliance and Bgee for cross-species expression data, high throughput data, and Thisse (Thisse et al. 2001) expression data are provided. Circles indicate new links. The Wild Type Expression Summary provides an overview of expression data, with color blocks indicating where annotations exist. Clicking on shaded boxes opens a table with more detailed gene expression information as shown.

Figure 3. Phenotype section of the gene page. Section of the gene page providing mutant and gene knockdown phenotype information. Links to Alliance gene page provided in Cross-Species Comparison section. The Phenotype summary provides a high-level overview of systems, stages, molecular functions, and biological processes that have annotations. Clicking shaded boxes opens a table where more detailed annotation information is provided, as shown.

Figure 4. Human Disease section of the Gene page. Provides information about human diseases associated with the gene via orthology to human genes and experimentally validated models of human disease. Links to Alliance disease pages (circled) provide access to multi-species disease information including associated genes, and experimental models.
Figure 5. Gene Ontology Section of the gene page. This section provides an overview of the GO molecular function, biological process, and cellular component annotations for a gene. Shaded boxes indicate presence of data; clicking shaded boxes, as shown, opens a table where more detailed annotation information is provided.
**pax2a**

**ID**  
ZDB-GENE-990415-8

**Name**  
*paired box 2a*

**Symbol**  
*pax2a*

**Previous Names**  
Pax-2, pax-b, *pax2.1*, Pax2.1, *pax2a1*, *paxb*, cb378 (1), *no isthmus*, noi, *pax[zf-b]*

**Type**  
protein_coding_gene

**Location**  
Chr: 13  
Mapping Details/Browsers

**Description**  
Exhibits sequence-specific DNA binding activity. Involved in several processes, including animal organ development; cilium assembly; and peripheral nervous system development. Localizes to nucleus. Used to study coloboma. Human ortholog(s) of this gene implicated in focal segmental glomerulosclerosis 7 and renal coloboma syndrome. Is expressed in several structures, including immature eye; mesoderm; nervous system; neural keel; and pronephros. Orthologous to human *PAX2* (paired box 2).

**Genome Resources**  
Alliance (1), Gene:30425 (1), VEGA:OTTDARG00000027765 (1), Ensembl(GRCz11):ENSARG00000028148 (2)

**Note**  
None

**Comparative Information**

**Expression**

**All Expression Data**  
767 figures from 535 publications

**Cross-Species Comparison**  
Alliance, Bgee
Phenotype Summary

- Exclude altered gene expression phenotypes
- Exclude morphant phenotypes

Phenotype in digestive system

| Phenotype                                      | Stage Observed | Citations          |
|------------------------------------------------|----------------|--------------------|
| gut epithelium edematous, abnormal             | 0 hpf          | Lun et al., 1998   |
|                                                | 2.00 hpf       |                    |
|                                                | 3.33 hpf       |                    |
|                                                | 5.66 hpf       |                    |
|                                                | 10.33 hpf      |                    |
|                                                | 24 hpf         |                    |
|                                                | 60 hpf         |                    |
|                                                | 7 dpf          |                    |
|                                                | 90 dpf         |                    |
## Associated With pax2a Human Ortholog

| Disease Ontology Term                  | Multi-Species Data | OMIM Term                                         | OMIM Phenotype ID |
|---------------------------------------|--------------------|---------------------------------------------------|-------------------|
| focal segmental glomerulosclerosis 7  | Alliance           | Glomerulosclerosis, focal segmental, 7            | 616002            |
| renal coloboma syndrome               | Alliance           | Papillorenal syndrome                             | 120330            |

**1 - 2 of 2**

## Associated With pax2a Via Experimental Models

| Human Disease | Fish     | Conditions     | Citations |
|---------------|----------|----------------|-----------|
| coloboma      | pax2a tu29a/tu29a | standard conditions | (2)       |

**1 - 1 of 1**
## Gene Ontology

### All annotations
- catalytic activity
- carbohydrate metabolic process
- DNA binding
- DNA replication
- DNA repair
- DNA transcription activity
- receptor activity
- receptor binding
- small molecule binding

### All molecular functions
- DNA binding
- histone binding
- transcription factor activity

### All biological processes
- cellular component organization
- DNA metabolic process
- gene expression
- immune system process

### Other molecular functions
- carbohydrate metabolic process
- cell cycle

### Other biological processes
- cell death
- DNA metabolic process

---

| Qualifier       | Term              | Annotation Extension | Evidence     | With/From | Citations        |
|-----------------|-------------------|----------------------|--------------|-----------|------------------|
| cilium assembly | IMP               | MO1-pax2a            |              |           | Shim et al., 2016|
| cilium assembly | IMP               | MO6-pax2a            |              |           | Shim et al., 2016|