

The Zebrafish Information Network: the zebrafish model organism database

Judy Sprague*, Leyla Bayraktaroglu, Dave Clements, Tom Conlin, David Fashena, Ken Frazer, Melissa Haendel, Douglas G. Howe, Prita Mani, Sridhar Ramachandran, Kevin Schaper, Erik Segerdell, Peiran Song, Brock Sprunger, Sierra Taylor, Ceri E. Van Slyke and Monte Westerfield

The Zebrafish Information Network, 5291 University of Oregon, Eugene, OR 97403-5291, USA

Received September 8, 2005; Revised and Accepted October 12, 2005

ABSTRACT

The Zebrafish Information Network (ZFIN; <http://zfin.org>) is a web based community resource that implements the curation of zebrafish genetic, genomic and developmental data. ZFIN provides an integrated representation of mutants, genes, genetic markers, mapping panels, publications and community resources such as meeting announcements and contact information. Recent enhancements to ZFIN include (i) comprehensive curation of gene expression data from the literature and from directly submitted data, (ii) increased support and annotation of the genome sequence, (iii) expanded use of ontologies to support curation and query forms, (iv) curation of morpholino data from the literature, and (v) increased versatility of gene pages, with new data types, links and analysis tools.

INTRODUCTION

The zebrafish has become a well-established model organism, making important contributions to the identification and characterization of genes and pathways involved in development, organ function, behavior and disease. With this success has come the challenge of managing the flood of data and integrating these data with the high volume of information generated by research in other model organisms and humans. ZFIN fills this role by providing a centralized repository and web-based query interface for zebrafish research data, including mutants, genes, genetic markers, mapping panels, links to other genomic resources, publications and community contact information (1). Data integration within ZFIN as well as links to resources outside of ZFIN fosters an understanding of gene function by integrating genotypes, phenotypes and gene

expression with gene sequences and gene models. We continually update and expand the content of ZFIN through ongoing literature curation, bulk data loads and addition of new features. This article describes recent enhancements to ZFIN that increase the utility of this community resource.

DETAILED CURATION OF GENE EXPRESSION

ZFIN provides an integrated representation of gene expression data. Gene expression patterns are annotated with expressed genes, fish genotype, assay, experimental conditions, developmental stage and anatomical structures (Figure 1). We annotate anatomical structures using terms from the zebrafish anatomical ontology. A variety of experimental conditions may be recorded, including temperature, chemicals and use of antisense knockdown reagents such as morpholinos targeted to specific genes.

Gene expression data enter ZFIN by literature curation or through direct data submission. Directly submitted expression data come primarily from a small number of laboratories engaged in large-scale projects. There is no minimum size for submissions, and all researchers are encouraged to submit their published or unpublished, high quality expression data. We provide a standardized template for submissions upon request. ZFIN currently holds >33 000 directly submitted images illustrating the expression of nearly 6000 genes. Each month, ZFIN receives annotated expression patterns for >400 new gene probes from the Thisse *in situ* screening project (2,3).

A significant new feature is detailed curation of gene expression patterns from the scientific literature. We associate each published figure with terms describing the genes, genetic backgrounds, stages, environments and anatomical structures. Images and figure captions are displayed when consistent with journal copyright restrictions. To date, we have curated >1400 figures from ~800 publications, most of which were

*To whom correspondence should be addressed. Tel: +1 541 346 2355; Fax: +1 541 346 0322; Email: judys@cs.uoregon.edu

Hans et al., 2004 - Pax8 and Pax2a function synergistically in otic specification, downstream of the Foxi1 and Dlx3b transcription factors. Development 131(20):5091-5102

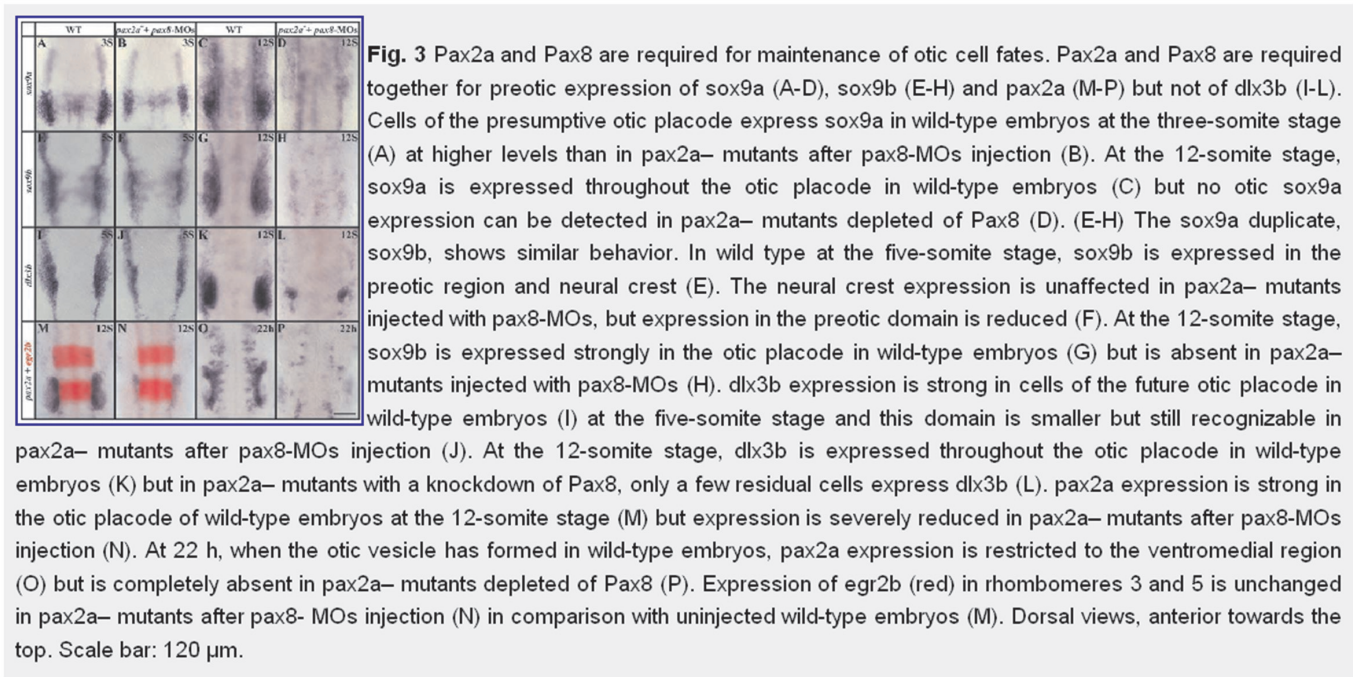
ADDITIONAL FIGURES

Genes : [dlx3b](#) ▼, [egr2b](#) ▼, [pax2a](#) ▼, [sox9a](#) ▼, [sox9b](#) ▼

Genetic Background : [no^{tu29a}](#), [WT](#)

Anatomical Terms: [ectoderm](#), [neural crest](#), [otic placode](#), [otic vesicle](#), [presumptive hindbrain](#)

Stage Range : [1-4 somites](#) to [26+ somites](#)



Gene expression details

Gene	Fish	Stage	Qualifier	Anatomy	Assay
dlx3b ▲	WT	5-9 somites		ectoderm	ISH
		10-13 somites		otic placode	ISH
	no^{tu29a} , MO:pax8	5-9 somites		ectoderm	ISH
		10-13 somites		otic placode	ISH
egr2b ▲	WT	10-13 somites		presumptive hindbrain	ISH
	no^{tu29a} , MO:pax8	10-13 somites		presumptive hindbrain	ISH
pax2a ▲	WT	10-13 somites		otic placode	ISH
		26+ somites		otic vesicle	ISH

Figure 1. A typical published figure with curated gene expression annotation. For brevity, only a subset of the annotation is shown. Genes, mutants, morpholinos and anatomy terms are all linked to their respective pages in ZFIN. Figure reproduced from Hans et al., 2004 (4).

published in the last two years. First priority for figure curation is given to current publications, and figures from older papers are curated on an *ad hoc* basis. An enhanced gene expression query interface allows complex searches of published expression data, amplifying the utility of the scientific literature. Gene expression queries can include gene, genetic background, stage, anatomical structure or morpholino target gene. Searches may be performed of all expression data, or constrained to include only published, directly submitted or recently entered data.

INTEGRATION OF SEQUENCES

The zebrafish genome is being sequenced by the Wellcome Trust Sanger Institute. The Sanger Institute provides access to the annotated genome sequence through the Ensembl and Vega (Vertebrate Genome Annotation) databases. Ensembl presents a view of the automated genome analyses based on pre-computed genome alignments to other sequences and an initial set of gene models, while the manual annotation provides a refined set of curated gene models that are displayed in Vega (<http://vega.sanger.ac.uk/>). The Sanger Institute and

ZFIN collaborate extensively to present the manual annotation that can be viewed in Vega. We integrate gene and clone annotations with the existing genomic, genetic and phenotype information in ZFIN. We compare manually annotated genes from Vega with genes in ZFIN to ensure correct associations and to identify the manually annotated genes that are new to ZFIN. Database records are created in ZFIN for novel genes and are assigned temporary nomenclature. ZFIN clone records are created for all the sequenced and annotated genomic clones. Reciprocal links between the ZFIN and Vega gene and clone records are established to complete the data integration process and to facilitate user access to relevant information at either database. A system is currently under development to expedite the renaming of novel zebrafish genes with more informative nomenclature. We continuously use information from orthologous and paralogous Human and mouse genes to revise and update nomenclature for novel zebrafish genes. ZFIN, the Human Gene Nomenclature Committee, and Mouse Genome Informatics strive to achieve uniform nomenclature for orthologous genes among these species whenever possible. Official nomenclature of zebrafish genes ultimately follows the established and approved nomenclature for orthologous Human and mouse genes.

ZFIN is also integrating cDNA clones from the Zebrafish Gene Collection (ZGC) (<http://zgc.nci.nih.gov>), an NIH sponsored program supporting the production of a complete set of full-length cDNA clones and sequences of expressed zebrafish genes. These clones play an important role in improving gene identification, morpholino construction and array development for expression analyses. We run ZGC clones through a similar process as established for the genome sequence annotations to ensure correct associations and to identify sequences representing genes new to ZFIN. ZFIN curators assign temporary names for novel genes. Detailed analyses of orthology with Human and mouse genes then facilitates assignment of more informative nomenclature. ZFIN clone records are created for all ZGC clones and are fully integrated with ZFIN gene records.

Clones from both the Sanger Institute genome sequencing effort and the ZGC initiative are available without restriction to the scientific community and can be obtained via a direct link from ZFIN gene and clone records.

EXPANDED USE OF ONTOLOGIES

The zebrafish anatomical ontology

ZFIN serves as the central repository for development and dissemination of the zebrafish anatomical ontology. The zebrafish anatomical ontology is a hierarchical vocabulary of zebrafish anatomical terms, including many definitions and synonyms. The anatomical ontology provides standard terminology for annotating gene expression and phenotypes, thus providing a link between these two types of data commonly used to study gene function. We are developing the zebrafish anatomical ontology in collaboration with other model organism communities, including fly and mouse, to promote cross-species comparisons.

The zebrafish anatomical ontology includes structures from each of the 44 defined stages of zebrafish development, arranged by functional system. This arrangement makes it

simple to search for anatomical structures. Each term exists within a defined range of developmental stages and can have multiple relationships to other anatomical terms in the ontology, making the zebrafish anatomical ontology more robust than a simple dictionary of structures. The anatomical ontology includes the relationship types *is_a*, *part_of* and *develops_from*. Some examples are: The optic cup *develops_from* the optic vesicle, which is *part_of* the eye. The eye is *part_of* the visual system, which *is_a* sensory system. These relationship types aim to capture not only the form and function of anatomical structures, but also the dynamic nature of their development.

ZFIN adds and updates anatomical terms, definitions and stage ranges regularly. For relatively simple cases, term definitions are derived from the literature. For more complex cases, a consortium of researchers who serve as experts for particular sets of anatomical structures, or researchers who specialize in a particular field are consulted. Members of the zebrafish anatomy consortium can be found at http://zfin.org/zf_info/anatomy/dict/mem.html. User suggestions for new terms or changes to existing terms are welcome, and can be made through the 'Your Input Welcome' button found on most ZFIN web pages.

Gene expression can now be queried using terms from the anatomical ontology at <http://zfin.org/cgi-bin/webdriver?Mlval=aa-xpatsselect.apg>. In the future, queries for phenotypes of mutants, transgenics and genetic knockdown experiments will also make use of the anatomical ontology.

The zebrafish anatomical ontology is available at the Open Biological Ontologies (OBO) website (<http://obo.sourceforge.net>) or from the ZFIN downloads page (http://zfin.org/zf_info/downloads.html#ad).

Gene ontology

The gene ontology (GO) is a set of three orthogonal controlled vocabularies designed to facilitate annotation of the molecular functions, biological processes and cellular components of gene products (for details about GO see the Gene Ontology article in this issue). ZFIN curators have been adding GO annotations to gene records in ZFIN as a routine part of literature curation since the end of 2003. As of August 2005, 2366 manual annotations have been made on 738 unique genes. An electronic GO annotation pipeline based on translation of InterPro domains, enzyme commission numbers and SwissProt keywords to GO terms is also in place. This electronic annotation process has produced 25 780 annotations on 5850 unique genes. GO term or gene based queries of zebrafish (and many other species) GO data can be made using AmiGO, the web based GO query interface provided by the GO consortium (<http://godatabase.org>).

MORPHOLINO CURATION

Morpholinos are synthetic oligonucleotides that bind to complementary sequences of RNA, disrupting translation initiation or pre-mRNA splicing. The proven effectiveness of morpholinos to knock down gene function has resulted in their widespread use for evaluating gene function in zebrafish. ZFIN now curates morpholinos from the literature, making it easy to locate morpholinos that have been used effectively by others

to target a specific gene, and to locate papers that report using a specific morpholino. We assign each morpholino a unique name in the format MO#-targeted-gene-symbol, which has been approved by the Zebrafish Nomenclature Committee (http://zfin.org/zf_info/nomen_comm.html). Published morpholino names are retained as aliases linked to the appropriate publications. We verify morpholino sequences by sequence analysis before entering them into ZFIN. If there are apparent discrepancies, we contact the authors and describe any resulting changes to the published sequence in the notes field in the morpholino record. ZFIN currently contains records for 399 morpholinos targeting 236 genes.

Morpholino records can be found in ZFIN in several ways. From the Genes/Markers/Clones search page you can search specifically for morpholinos by selecting 'Morpholino' from the 'Types' menu, and entering all or part of the gene symbol for the targeted gene in the search box. Morpholino information can also be found at the top of gene pages, in the 'Mutants and Targeted Knockdowns' section. Morpholinos associated with a specific publication are also listed in the 'Additional Information' section located at the bottom of ZFIN publication records.

GENE PAGE ENHANCEMENTS

The gene page is a central hub from which a variety of gene-specific information is accessible. This wealth of summarized information has made the gene page the most frequently visited page in ZFIN. Gene pages are continuously updated owing to ongoing curation and the addition of new data types. Keeping informed of these changes can present a challenge to even the most seasoned users. Recent changes to the ZFIN gene pages include the following.

Mutants and targeted knockdowns

In addition to displaying the mutant locus that is known to correspond to a specific gene, this section now also displays knockdown reagents designed to target the gene. This is currently limited to morpholinos, but may include other types of knockdown reagents in the future.

Gene products

This section contains a list of GO terms that we curate from the literature as well as electronically. A detailed view of the GO annotations is available, where publications supporting the annotations can be found. In addition to GO, links to external databases storing information on protein families, domains and sites found in each gene product are also located in this section. A link to a 'Gene Product Description' has been added here as well. This link displays the detailed description of the gene product as shown in the UniProt record associated with that gene.

Gene expression

The gene expression section now includes links to all the expression data in ZFIN for the specific gene. ZFIN provides separate links for 'All expression data' and 'Directly submitted data'. A 'current status' link alerts users to the kinds of expression data supported by ZFIN and to the curation status of older literature.

Segment (clone and probe) relationships

This section provides links to ZFIN cDNA and genomic DNA segment records. The relationship between the gene and the nucleic acid segment is also indicated. We continue to add cDNA clones from the ZGC, as well as BACs and PACs used in the Genome Sequencing Project. You can find links to DNA segment pages supporting ZFIN mapping and expression data in this section. DNA segments that can be ordered from various sources have an 'order this' hyperlink beside them.

Sequence information

This section contains sequences associated with a specific gene, categorized as cDNA, genomic, polypeptide or sequence cluster (UniGene). A complete list of cDNA, genomic, polypeptide and cluster sequences associated with the gene are found on a separate page accessed by selecting the 'All Sequence Information' link. We have added additional information such as length and sequence type to each sequence. A pull-down menu of sequence analysis options beside each sequence provides increased functionality. Analysis options may include NCBI BLAST, Ensembl BLAST, UCSC BLAT or SIB BLAST depending on the sequence type. Selecting one of these options prepares the selected query form to analyze the associated sequence.

Other gene/marker pages

This section now includes direct links to zebrafish gene pages in Entrez Gene and in the Sanger Institute's Vertebrate Genome Annotation database (Vega). Marker pages that are part of The Sanger Institute's fingerprinting map of the zebrafish genome (Fingerprint Contig or FPC) are also now available in this section of the ZFIN gene page.

Orthologs

The redesigned orthology display includes evidence codes to indicate the type of data that supports each assertion of orthology. Work is currently under way to include chromosome location for mouse and human orthologs in the near future.

IMPLEMENTATION

ZFIN is currently implemented with the IBM/Informix relational database management system (server version 9.4). A web interface of HTML-based forms combined with JavaScript, Java, Perl and CGI scripts provides access to the database. The current ZFIN data model may be viewed at <http://zfin.org/DataModel>.

FUTURE DIRECTIONS

Expanded use of ontologies will provide better support for curation and querying for data, and it will facilitate cross-species comparative genomics. This broader implementation of ontologies will also support phenotype annotation, with the goal of providing comprehensive information about mutant and morpholino phenotypes. Annotation tools are currently available for laboratories that are generating phenotype or expression data. We encourage investigators interested in submitting data directly to ZFIN to use these tools. For more

information on this process email us at zfinadm@zfin.org. We are developing new query forms and page displays that will fully integrate phenotypes with other data in ZFIN.

ACKNOWLEDGEMENTS

Funds for the development of the Zebrafish Information Network are provided by the NIH HG002659. Funding to pay the Open Access publication charges for this article was provided by NIH HG002659.

Conflict of interest statement. None declared.

REFERENCES

1. Sprague,J., Clements,D., Conlin,T., Edwards,P., Frazer,K., Schaper,K., Segerdell,E., Song,P., Sprunger,B. and Westerfield,M. (2003) The Zebrafish Information Network (ZFIN): the zebrafish model organism database. *Nucleic Acids Res.*, **31**, 241–243.
2. Thisse,B., Pfumio,S., Fürthauer,M., Loppin,B., Heyer,V., Degraeve,A., Woehl,R., Lux,A., Steffan,T., Charbonnier,X.Q. *et al.* (2001) Expression of the zebrafish genome during embryogenesis. ZFIN on-line publication.
3. Thisse,B. and Thisse,C. (2004) Fast release clones: a high throughput expression analysis. ZFIN on-line publication.
4. Hans,S., Liu,D. and Westerfield,M. (2004) Pax8 and Pax2a function synergistically in otic specification, downstream of the Foxi1 and Dlx3b transcription factors. *Development*, **131**, 5091–5102.