Published in final edited form as:

Methods Cell Biol. 2016 ; 135: 451–481. doi:10.1016/bs.mcb.2016.04.010.

A Scientist's Guide for Submitting Data to ZFIN

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, Oregon, USA 97403-1254

Abstract

The Zebrafish Model Organism Database (ZFIN; zfin.org) serves as the central repository for genetic and genomic data produced using zebrafish (*Danio rerio*). Data in ZFIN are either manually curated from peer-reviewed publications or submitted directly to ZFIN from various data repositories. Data types currently supported include mutants, transgenic lines, DNA constructs, gene expression, phenotypes, antibodies, morpholinos, TALENs, CRISPRs, disease models, movies, and images. The rapidly changing methods of genomic science have increased the production of data that cannot readily be represented in standard journal publications. These large data sets require web-based presentation. As the central repository for zebrafish research data, it has become increasingly important for ZFIN to provide the zebrafish research community with support for their data sets and guidance on what is required to submit these data to ZFIN. Regardless of their volume, all data that are submitted for inclusion in ZFIN must include a minimum set of information that describes the data. The aim of this chapter is to identify data types that fit into the current ZFIN database and explain how to provide those data in the optimal format for integration. We identify the required and optional data elements, define jargon, and present tools and templates that can help with the acquisition and organization of data as they are being prepared for submission to ZFIN. This information will also appear in the ZFIN wiki, where it will be updated as our services evolve over time.

Introduction

ZFIN is the central repository of genetic and genomic information for the zebrafish research community. Granting agencies increasingly require that large data sets be submitted to an appropriate database repository such as ZFIN. In light of this, we aim to integrate as much of the zebrafish mutant, transgenic, expression and phenotype data from the research community as possible and to provide links back to source databases when possible. To accomplish that, there must be clear guidance and documentation of the process and requirements for adding data to the ZFIN database. The process of preparing, submitting, and loading data into the ZFIN database is called a “data submission”. Planning ahead for data submission to ZFIN will result in more efficient and timely addition of data to the database. This publication, the associated ZFIN wiki pages, and data submission templates

^{*}Corresponding Author.

referenced herein provide an up-to-date reference resource to support submission of large and small data sets to ZFIN.

Why Load Data Into ZFIN?

In this era of big data it is increasingly important to integrate data from multiple sources to provide a unified view of data at a single location. This integration maximizes the value of each piece of data by allowing queries to return accurate and more complete results, accelerating research and reducing redundant effort and research cost. ZFIN supports a diverse collection of data types including mutants, transgenic lines, expression, phenotypes, constructs, morpholinos, TALENs, CRISPRs, antibodies, and disease models. Data curated from publications and from prior data loads are integrated to provide as complete a picture as possible of the role and function of each gene based on all the information available in the ZFIN database. Often, data stored in lab-specific databases lack the long-term stability, accessibility, and data integration that they will have in ZFIN. The goal at ZFIN is to capture the essential core of the data and any additional details the ZFIN database is able to support. In many cases it is possible to link from ZFIN back to the laboratory web pages, providing easy access to any further details that are not currently included at ZFIN. In the long term, the most important services ZFIN can provide related to data loads are to integrate data from disparate resources, provide a central location presenting a complete picture of what is known about a topic of interest, and provide critical long term data stability and accessibility.

The Structure of the ZFIN Database

Data at ZFIN are stored in a complex relational database consisting of over 300 database tables (<http://zfin.org/schemaSpy>). This database structure allows disparate pieces of data to relate to each other and to be presented in an integrated format. Aligning incoming data with existing ZFIN data and associated database constraints is one of the major challenges for data loads, particularly if alignment is considered only after data are collected. In contrast, if the structure of a potential data submission is understood early in the data gathering process, data collection can be optimized to facilitate a smooth data submission process. Below we describe the data submission process and each major data type we currently support, as well as their components, and we identify which components are required and which are optional.

The Data Submission Process

Data submission requests are typically initiated by an inquiry from a researcher. Once submission of data has been agreed upon, there are several steps to a typical data submission process (Figure 1). A curator assigned to the data load will provide guidance on data gathering, establish necessary records in ZFIN, and assist in getting the data into a format that can be loaded into the ZFIN database.

Once submitted, data are subject to a number of quality control and data validation steps. Inconsistencies in the data are resolved through discussion with the data submitter. Once the

data are free of errors, they are loaded into a test database for final review by the submitter. Once approved by the submitter, the data are loaded into the ZFIN database.

ZFIN does not hold private data. Once submitted, the data load will proceed as part of the normal software release cycle. When there are data that cannot be released to the public until they are published, an initial submission of high quality data from the same experiments that is not part of the publication can be considered. This may include data such as mutations with no obvious or early lethal phenotypes, gene expression where there is either no expression at a particular developmental stage or expression is ubiquitous, or enhancer traps that trap already characterized enhancers. This allows the researcher to become familiar with the data submission process and to validate the data submission file format while protecting the integrity of the unpublished data. Once published, those data can be submitted with confidence, knowing that similar data have already been integrated into ZFIN.

Data Submissions

The data required for a submission often involves multiple files that together provide the information needed to represent the data fully at ZFIN. In this section each of the data types that can be loaded into the ZFIN database are described along with their required and optional components.

Mutant and Transgenic Line Submission

Mutant features are genomic alterations often generated by an applied mutagen, whereas transgenic features are genomic alterations generated by insertion of one or more copies of a transgenic construct. These may or may not result in alleles of genes. Mutant and transgenic lines are strains of fish that contain one or more heritable transgenic or mutant features. Each transgenic feature may contain one or more transgenic constructs. The exact insertion site or sites may or may not be known. Lines that contain multiple distinct known transgenic insertion loci will have a distinct genomic feature designated for each insertion site. The transgenic line is then composed of a combination of those distinct insertions. Data submissions containing genomic features not yet in ZFIN include information to create new feature records. Below we describe the various elements of data that can be included in a mutant or transgenic line data submission.

Genotypes—The genotype represents the primary genomic sequence alterations present in a fish. The genotype conveys zygosity information about specific loci having known sequence variants or transgenic insertions as well as the genetic background. To define a specific genotype in ZFIN completely and uniquely, each genomic feature and the zygosity of the locus where it resides are required. To define the genotype further, information about the parental zygosity of each locus and the genetic background can also optionally be provided (Table 1). The required and optional data elements for transgenic line submission (Table 2) and for mutant submission (Table 3) are similar, but distinct.

Genomic Feature—Each genomic feature has a unique name representing a specific genomic alteration or transgenic insertion event. Genomic feature names are composed of the line designation and a unique number identifier. Transgenic features have the suffix Tg,

Et, Pt, or Gt depending on whether they are standard transgenic, enhancer trap, promoter trap, or gene trap features, respectively. Line designations are institution specific and can be obtained from the ZFIN nomenclature coordinator (nomenclature@zfin.org), and the researcher supplies the number identifier.

Feature Zygosity—Feature zygosity is the zygosity of the locus with respect to the specific genomic feature. Valid choices include: homozygous, heterozygous, and unknown. Each genomic feature in a genotype will have zygosity information associated with it. Transgenes that are present on only a single copy of the locus are considered heterozygous in ZFIN.

Feature Maternal and Paternal Zygosity—The male and female parental zygosity of the locus at which there is a known genomic feature can be included to provide information about how the genomic feature was inherited. Parental zygosity of each mutant or Tg locus is reported as either homozygous, heterozygous, wild type or unknown. Transgenes that are only present on a single copy of the male or female parental locus are considered heterozygous for the parent in ZFIN.

Genetic Background—The Genetic background specifies on which standard or wild-type line (Table 4) the genotype is being carried. If there are no mutations involved, the genotype may be one of these standard lines. More detailed information about these standard lines can be found on the ZFIN Wild-Type Lines web page at <http://zfin.org/action/feature/wildtype-list>.

Affected Gene Symbol—If the genomic alteration is a point mutation, small insertion, small deletion, indel, transgene insertion, or results in a change in the coding region of one gene, that gene should be listed as an affected gene to indicate that the genomic alteration is an allele of that gene. For other mutation types (translocations, inversions, deficiencies), which may affect multiple genes, a separate file is used to specify the affected genes and how they relate to the genomic feature (Table 5). Genomic features with multiple affected genes are provided with one row per affected gene in this data file.

Affected Gene Accession—When submitting a mutant or transgenic line with an affected gene it is essential to supply a unique identifier for the gene to remove any ambiguity in gene identification. ZFIN gene record IDs (ZDB-GENE IDs) are the best identifiers for unambiguous identification of genes in ZFIN, so they are preferred whenever possible. If ZDB-GENE IDs are difficult to obtain for your data, Ensembl IDs (OTTDART or ENSDARG) are the next best choice because they fit most easily into existing gene identification pipelines at ZFIN. Minimally, a sequence accession number, such as the GenBank ID for the longest transcript of the gene, must be provided to ensure accurate identification of genes for the incoming data. Without such gene-specific identifiers, it may not be possible to load all data in a submitted data set.

Transgene type—In cases where trapping has occurred, transgene type abbreviations are added to the end of the line designation to denote the type of transgene (Et for enhancer trap,

Gt for gene trap and Pt for promoter trap). It is acceptable to use Tg to name trap lines, but the more specific name using Et, Pt or Gt is helpful in denoting their function.

Mutation Type—ZFIN supports many types of genomic sequence alterations derived from the sequence ontology (SO) (2) (Table 6). One of these types must be assigned to each mutant submitted to ZFIN.

Mutagen—The mutagen used to generate the genomic feature or transgenic line must be provided. Valid mutagens come from a constrained set of terms (Table 7). If a TALEN or CRISPR was the mutagen, then the name of the specific TALEN or CRISPR should be provided as the mutagen. For transgenic lines using only a construct, the mutagen is “DNA”. If a TALEN or CRISPR was also used during creation of a transgenic insertion, the TALEN/CRISPR name should be provided as the mutagen. The construct name is provided in a separate column (see below). Details for TALENs, CRISPRs, and constructs not already in ZFIN must be specified in a separate file dedicated to these new record types. See the TALEN, CRISPR, and MO section and the Transgenic Constructs section for details on supplying these data.

Subject—This is the type of original recipient animal of the mutagen treatment. Subject is used in combination with “mutagen” data to capture how the genomic alteration was generated. Valid choices for the subject come from the following constrained set: Adult Males, Adult Females, Embryos, Sperm.

Construct—The name of the transgenic construct used to make a transgenic line should be specified. If the construct is already in ZFIN, the ZFIN construct name should be provided. When multiple constructs are inserted simultaneously there should be multiple rows of data for the line designation specifying only one construct per row. Transgenic constructs that are new to ZFIN are described in a separate section dedicated to new construct records. See the Transgenic Constructs section of this chapter for details on submitting new construct data.

Laboratory of Origin—All genomic features require a laboratory of origin, the laboratory from which the feature originated. This laboratory must have a laboratory record in ZFIN. New laboratory records can be created upon request at zfinadm@zfin.org.

Sequence Accession—When genomic features are located by sequencing, a GenBank accession number may be provided to describe the altered genomic sequence or transgene insertion site. The GenBank number is included on the genomic feature page in ZFIN and may also be used to locate features on a genome browser.

Link to Alternate Resource—Some laboratories provide access to their data via a dedicated laboratory web site. A URL pointing to such an alternate web-based resource about the mutant or transgenic line can be provided. This URL is used to link from ZFIN back to the original resource web page where additional information about the feature may be available.

Citations—All data added to ZFIN must be associated with a publication record specifically created at ZFIN to represent the data load. The publication record created at ZFIN must minimally have a list of authors, a title, and an abstract. Data that have been published in a peer reviewed journal article may also be attributed to that publication by providing the PubMed ID (Table 8).

Note—A free text note with further details about the mutant or transgenic line can be included. For a mutant line these details may include transition/transversion, location info, null/hypomorph/gain of function, amino acid changes, etc. These details clarify the nature and location of the mutation, particularly for details that can not be stored in a dedicated field in the ZFIN database. For a transgenic line the note may include information such as the insertion site construct copy number, the consequence of the insertion (null, hypomorph, etc.), or specific details about the function or composition of the line.

Sperm Samples—ZFIN accepts feature data only for sperm samples that have been accepted by a resource center such as the Zebrafish International Resource Center (ZIRC). Data loads involving sperm samples will be arranged after the resource center contacts ZFIN regarding the sperm samples they expect to receive.

Transgenic Constructs

Transgenic constructs are engineered nucleotide sequences injected into fish resulting in heritable transgenic insertions into the genome.. ZFIN creates construct records only for constructs used to make stable and heritable transgenic insertions. Information about the functional portions of constructs is collected, but information about plasmid propagation is not. The required and optional data elements for creating new transgenic construct records in ZFIN are describe below (Table 9). Transgenic construct design varies widely and, as a result, a multiple construct submission formats are required. The information requested here will support only basic constructs with a single functional cassette. When construct data do not fit into this simple set of fields, we are happy to work with data submitters to accept construct data as completely as possible. This can typically be accomplished by submitting an extended version of the simple data we typically request.

Construct Name—The name ZFIN uses to label constructs is composed of the functional parts of the construct, mainly the promoter elements and expressed genes. A single colon is used to separate the promoter element from the expressed gene. Basal promoters are not recorded, but any gene specific promoter elements should be listed in the promoter section of the name (Figure. 2). Additionally, the type of transgenic construct is encoded in the name by prepending the name with Tg if the construct is a general transgenic construct, Gt if the construct is a gene trap, Et if the construct is an enhancer trap, and Pt if the construct is a promoter trap. Additional names for constructs can be provided as aliases if desired. Assistance with naming transgenic constructs is provided by the ZFIN nomenclature coordinator (nomenclature@zfin.org).\

Promoter Gene Symbol—Many transgenic constructs drive gene expression using the promoter of a gene from zebrafish or another species. The symbol for the promoter gene

should be provided, including an abbreviation for the species from which the sequence was derived if it is not a zebrafish gene (Hsa for human, Mmu for mouse, etc.). In some cases, like promoter trap constructs, there may not be a regulatory region to include in the construct name.

Promoter Gene Accession—When submitting data specifying a gene it is essential to supply a stable and unique identifier for the gene to remove any ambiguity in gene identification. ZDB-GENE IDs are the best identifiers for this purpose. Please see the Affected Gene Accession paragraph in the Mutants and Transgenic Lines section of this chapter for more details on providing unique gene identifiers.

Coding Sequence Gene Symbol—Most transgenic constructs drive expression of a gene or a reporter in the derived transgenic fish. The symbol for the product that will be expressed from the integrated construct should be provided here, including an abbreviation for the species from which it was derived if it is not a zebrafish gene.

Coding Sequence Gene Accession—When submitting data specifying a gene it is essential to supply a unique and immutable identifier for the gene to remove any ambiguity in gene identification. ZDB-GENE IDs are the best identifiers for this purpose. Please see the Affected Gene Accession paragraph in the Mutants and Transgenic Lines section of this chapter for more details on providing unique gene identifiers.

Engineered Region Name—Constructs often include additional elements in them such as binding or cleavage sites, IRES sequences, etc. ZFIN refers to these as Engineered Regions (ERs). ERs are typically not included in the name of the construct. Instead, they are linked to the construct using a “construct contains” relationship, making the ER details available on the construct page. The complete list of Engineered Regions currently supported at ZFIN is available via a marker search at ZFIN. New Engineered Region records can be requested through the nomenclature coordinator (nomenclature@zfin.org).

Construct Sequence Accession—The GenBank number for a construct sequence can optionally be provided. This GenBank number will be visible on the construct record in ZFIN. If a BAC or PAC clone was used to generate the construct, the accession number of the BAC or PAC should be provided.

Construct Map Image Name—The name of an image file (jpg or png) can be provided showing a graphical representation of the construct map. The image file itself can be provided via email, ftp, or a file sharing service. The image of the construct map will be included on the ZFIN construct page.

Link to Alternate Construct Resource—Laboratories may have a website where more details about the construct are provided. A URL for such a web-based resource can be optionally provided. This is used to support a link to the alternate construct page from the ZFIN construct page.

Citation—All data loaded into ZFIN require a data load citation to which the data are attributed. This will be a ZDB-PUB ID for the data load publication and an optional PubMed ID for any related journal article. Please see “Citations” in the Mutant and Transgenic line section of this chapter for further details.

Construct Note—Unfortunately, the ZFIN database cannot support every detail for every construct. Sometimes there are additional pieces of information that further illuminate the nature of a construct. A note can be included here to provide salient details.

Morpholinos, TALENs, and CRISPRs

Morpholinos, TALENs, and CRISPRs are sequence-specific reagents used to disrupt gene translation or splicing, or to introduce sequence alterations at sequence-targeted locations. Once these tools are designed and tested, data describing them can be submitted to ZFIN. The required and optional data elements for submitting these records to ZFIN are described here (Table 10).

MO/TALEN/CRISPR Name—Morpholinos, TALENs, and CRISPRs are given unique names at ZFIN using a type-specific prefix with a unique numerical index and the gene symbol of the targeted gene appended. For example TALEN1-pax2a would be the first TALEN added to ZFIN that targets the pax2a gene. These names are assigned in sequence as each new record is created in ZFIN. Any name can be included on the record as a synonym.

Target Sequence 1—The target sequence reported is the 5'–3' genomic sequence being targeted by the TALEN, CRISPR, or Morpholino.

Target Sequence 2—The target sequence reported is the 5'–3' genomic sequence for the second genomic target of a TALEN. These data are not provided for CRISPRs or morpholinos, which only have a single target sequence.

Target Gene Symbol—The target gene symbol is for the gene being affected by the morpholino, TALEN, or CRISPR. If there are multiple target genes for a morpholino, TALEN, or CRISPR, there should be one row per target gene in the data submission file.

Target Gene Accession—When submitting data specifying a gene, it is essential to supply a unique and immutable identifier for the gene to remove any ambiguity in gene identification. ZDB-GENE IDs are the best identifiers for this purpose. Please see the Affected Gene Accession paragraph in the Mutants and Transgenic Lines section of this chapter for more details on providing unique gene identifiers.

Link to Alternate Resource—Some laboratories maintain databases specifically to house their own TALEN, CRISPR, or morpholino data. ZFIN can link back to those resources if a URL with long-term stability is provided.

Citations—All data loaded into ZFIN require a data load citation to which the data are attributed. This will be a ZDB-PUB ID for the data load publication and an optional PubMed

ID for any related journal article. Please see “Citations” in the Mutant and Transgenic line section of this chapter for further details.

Expression Data

Zebrafish are used extensively to assay gene expression and antibody labeling patterns. The information for an expression data submission is described below (Table 11).

Expressed Gene Symbol—Expression data are often recorded for a specific gene. This could be a zebrafish gene or a reporter transgene. The symbol of the gene whose expression is measured should be included. In some cases, the labeled gene may not be known or the antibody may be made using immunogens other than gene products, such as lipids, or chemicals. In those cases, the expressed gene data are not included.

Expressed Gene Accession—When submitting data specifying a gene, it is essential to supply a unique and immutable identifier for the gene to remove any ambiguity in gene identification. ZDB-GENE IDs are the best identifiers for this purpose. Please see the Affected Gene Accession paragraph in the Mutants and Transgenic Lines section of this chapter for more details on providing unique gene identifiers.

Genotype—The genotype of the fish in which the expression observation was made, including the genetic background, must be specified. See the Mutant Transgenic line section in this chapter for details on reporting genotypes to ZFIN.

Morpholinos, TALENs and CRISPRs—Morpholinos, TALENs and CRISPRs can be injected into zygotes to induce anatomically localized modification of gene expression or gene structure. When used in this way during an expression analysis, the name of the morpholino, TALEN, or CRISPR should be reported. Morpholinos, TALENs, and CRISPRs that already have records in ZFIN can be listed by their ZDB ID or ZFIN name. When a morpholino, TALEN, or CRISPR is not already in ZFIN, a separate file must be submitted for creation of these new records. Please see the Morpholinos, TALENs, and CRISPRs section of this paper for details on submitting these new records to ZFIN.

Anatomical Structure—The anatomical structure where the expression observation was made must be specified. To use a common vocabulary, the anatomical term should be selected from the Zebrafish Anatomy Ontology (ZFA), which describes zebrafish anatomy from larvae to adult (3). Any anatomical terms that are not drawn from the ZFA will need to be mapped to equivalent ZFA terms. This can be time consuming work before your submission can be accepted, so it is advised that the ZFA terms be used from the outset of data gathering. The most recent update of the ZFA is available at GitHub: <https://github.com/cerivs/zebrafish-anatomical-ontology>. More complex compound anatomical terms, including organ parts or sub-cellular localization (such as “mitochondrion part_of retinal ganglion cell”), can also be supported. Contact ZFIN if you would like to use such a term so a data format can be established for your particular data load.

Developmental Stage—The developmental stage is the developmental time point at which the experimental observation was made. The developmental stage is reported in hours

post fertilization and is selected from the Zebrafish Developmental Stages ontology (ZFS), which is based on developmental stages as described in Kimmel et al. 1995 (4). The current ZFS can be found here http://zfin.org/zf_info/zfbook/stages/index.html. Developmental stage terms that do not come from the ZFS will need to be mapped to equivalent ZFS terms. This can be time consuming work before a data submission can be accepted, so it is advised that the ZFS terms be used from the outset of data gathering.

Experimental Conditions—Experimental conditions provide details of the pertinent experimental treatments present at the time data was collected. Experimental conditions are represented in ZFIN by a constrained set of choices (Table 12). New conditions can be requested if necessary.

Experimental Conditions Note—A free text note can be used to capture further details about an experimental condition. If “chemical” is selected as the experimental condition the chemical name, supplier number, and a chemical identifier from ChEBI (5) should be supplied in the note.

Citations—All data loaded into ZFIN require a data load citation to which the data are attributed. This will be a ZFIN publication ID (ZDB-PUB ID) for the data load publication and an optional PubMed ID for any related journal article. Please see “Citations” in the Mutant and Transgenic line section of this chapter for further details.

Assay Type—The assay used to detect expression should be listed. Valid options are: immunohistochemistry, western blot, mRNA in situ hybridization, DNA in situ hybridization, intrinsic fluorescence, northern blot and reverse transcription PCR.

Antibody Name—When expression data are gathered by immunohistochemistry, the specific antibody can be included. If the antibody is already found in ZFIN, then the ZFIN ID for that antibody should be provided. However, if the antibody is not in ZFIN, then an antibody record can be created. A number of pieces of information are required to create the antibody record at ZFIN (Table 13). Information not listed as optional is required.

Probe GenBank Accession #—Expression data are often gathered by in situ hybridization using a labeled sequence specific nucleotide probe. If there is a GenBank number for the probe sequence, that information can be included here.

Images and movies—Images and movies are often the best medium to convey gene expression information. Submission of images or movies in conjunction with the expression data text files is encouraged. Image files should be in JPEG or PNG format. Movies should be in MP4 format with a 10MB per file size limit. Media file names need to be unique within the load. Encoding information about the content of the image/movie in the file name can be helpful. For example, expressed gene, stage, and image number could all be in the file name. If expression was in a non-standard background or if multiple alleles were used, this information could be included in the file name. For mutant phenotypes the allele and stage of observation could be included in the file name. If an antibody is used then the antibody could be encoded in the file name.

Phenotype Data

Phenotypes at ZFIN are recorded in the bipartite E:Q syntax (6)(7). The “E” is an entity (such as “eye” (ZFA:0000107)) in which the phenotype manifests. The “Q” is a phenotypic quality describing the nature of the phenotype (“decreased size” (PATO:0000587), for example). Together, this EQ combination can be used to describe a phenotype of “small eyes”. Phenotype data must be in this format before it can be loaded into ZFIN. The required and optional data elements for a phenotype data submission are listed in table 14.

Genotype—The genotype of the fish in which the phenotype was observed, including the feature name, zygosity, and genetic background, must be specified. See the Genotypes portion of the Mutant and Transgenic section elsewhere in this paper for details on how to provide genotype information.

Morpholinos, TALENS, and CRISPRs—Morpholinos, TALENs, and CRISPRs can be injected into zygotes to induce anatomically localized modification of gene expression and/or gene structure. When used in this way during a phenotypic analysis, the morpholino, TALEN, or CRISPR name should be reported. If the reagent already has a record in ZFIN, it can be listed by its ZFIN name or ZDB ID. When the reagent is not already in ZFIN, a separate file needs to be submitted to support creation of these new records. Please see the Morpholinos, TALENs, and CRISPRs section of this paper for details on submitting these new records to ZFIN.

Developmental Stage—The developmental stage is a term selected from the Zebrafish Developmental Stages ontology (ZFS). This should report the developmental stage at which the phenotype observation was made. Developmental stage terms that do not come from the ZFS will need to be mapped to ZFS terms before the data submission can be loaded. This data mapping can be time consuming, so it is recommended that the ZFS terms be used from the outset of data gathering.

Experimental Condition—All phenotype annotations need to have the experimental conditions specified. Valid values must come from the constrained set of experimental conditions supported at ZFIN (Table 12).

Phenotype Entity—The phenotype entity can either be an anatomical term selected from the Zebrafish Anatomy Ontology (ZFA) (e.g. “heart”), or a process term selected from the Gene Ontology (GO) (e.g. “cell migration”) (8). Any phenotype entity terms that are not drawn from these two sources will need to be mapped to equivalent terms from these ontologies. This can be time consuming work, so it is advised that the GO and ZFA terms be used from the outset of data gathering. Example terms include “eye” from the ZFA or “cell migration” from the GO. More complex phenotype entities can be supported by combining terms such as “cell migration occurs_in heart”. Contact ZFIN if you would like to use more complex entities so we can review the details together for your particular data load.

Phenotype Quality—The phenotype quality is a term selected from the Phenotypic Quality Ontology (PATO) (9). The current PATO ontology can be found at BioPortal (<http://>

purl.bioontology.org/ontology/PATO). The quality term describes the nature of the phenotype as it pertains to the phenotype entity. Any phenotype quality terms that are not drawn from the PATO ontology will need to be mapped to equivalent terms from PATO. This can be time consuming work, so it is advised that the PATO terms be used from the outset of data gathering. Note that some PATO terms can be used only with anatomical structures (Example: “decreased size” (PATO:0000587)), and others are exclusively used with biological process terms (e.g. “increased rate” (PATO:0000912)).

Tag—Phenotype annotations can capture observations of “abnormal” or “normal” processes or structures relative to control conditions. The “abnormal” tag is used when a phenotype involves an altered morphology or process compared to a control. For example, a mutant with small eyes relative to a control would be captured as Entity=“eye”, PATO=“decreased size”, Tag=“abnormal”. If it is notable that no abnormal morphological phenotype is observed, that can be annotated using the “normal” tag to capture the phenotype as Entity=“whole organism”, PATO=“morphology”, Tag=“normal”.

Media File Name—Phenotype data can be submitted with accompanying images or movies to illustrate the phenotype. The image or movie file name is specified here. Please see “Images and Movies” in the Expression section of this article for details about the media that can be provided.

Citations—All data loaded into ZFIN require a data load citation to which the data are attributed. This will be a ZDB-PUB ID for the data load publication and an optional PubMed ID for any related journal article. Please see “Citations” in the Mutant and Transgenic line section of this chapter for further details.

Genome Browser Tracks

It is now common for data to be placed into the context of the zebrafish genome by viewing them as a track in a genome browser. This can include gene-centric information, such as expression data, as well as epigenetic data such as DNA methylation status, specific binding sites, etc. Track files can be submitted to ZFIN and added to the ZFIN Track Hub (10). Tracks in the ZFIN track hub can be viewed in the UCSC Genome Browser (11) by adding the ZFIN Track Hub URL (<http://trackhub.zfin.org/zfintracks/hub.txt>) to the list of available hubs in the “My Data” tab at UCSC. The data necessary to submit a track for inclusion in the ZFIN track hub are described below.

Track Files—Track hub files and configuration are well documented on the UCSC web site: <https://genome.ucsc.edu/goldenPath/help/hgTrackHubHelp.html>. Please consult that resource for detailed information about track hubs and file formats. Track hubs submitted to ZFIN must be in one of the compressed binary formats supported by the Genome Browser: bigBed, bigGenePred, bigWig, BAM, HAL, or VCF.

Track Configuration and Description—All tracks are generated using the coordinate system from a specific zebrafish reference genome such as Zv9 or GRCz10 (1). Tracks from the ZFIN track hub will be visible in the UCSC genome browser only while viewing the

reference genome for which the track was generated. Tracks generated using the Zv9 coordinate system will not be visible while viewing the GRCz10 build at UCSC. The track submitter provides track description and configuration details. The UCSC track hub documentation provides a comprehensive review of the options available for track configuration. The ZFIN track hub has the following minimum set of configuration information that must be provided by the track submitter (Table 15).

Citation—All data loaded into ZFIN require a data load citation to which the data are attributed. This will be a ZDB-PUB ID for the data load publication and an optional PubMed ID for any related journal article. Please see “Citations” in the Mutant and Transgenic line section of this chapter for further details.

Track Maintenance—When new versions of genome builds are released, such as the transition from Zv9 to GRCz10, ZFIN will not port submitted tracks to the newer genome coordinate system. New track files must be provided by the original track submitter. Previously submitted tracks will remain available in the ZFIN track hub when viewing the data source for which the track was made.

Disease Models

ZFIN recently added the ability to annotate disease models created using zebrafish. A disease model can be created using a mutant, a morpholino, TALEN, CRISPR, and/or manipulation of the experimental conditions such as application of a chemical or modification of diet. Mutants and experimental conditions can be used in any combination to generate a disease model. All of the following pieces of data are required to submit a new disease model to ZFIN.

Genotype—The genotype of the fish used in the disease model must be specified. See the section dedicated to reporting genotypes to ZFIN the Mutants and Transgenics section of this paper.

Experimental condition—A disease model may exist as a wild-type or mutant fish with a chemical treatment. Therefore, the experimental conditions for all disease model observations must be provided. ZFIN currently uses a constrained list of experimental conditions (Table 12).

Morpholino, TALEN, CRISPR—Some disease models are generated using morpholinos, TALENs, or CRISPRs. The name of the morpholino, TALEN, or CRISPR must be specified when one is used as an experimental treatment rather than as a germline mutagen.

Disease Term ID—The Disease Ontology (DO) (12) is used to curate disease data in ZFIN. Any disease model data must therefore include the DO identifier for the DO term representing the intended disease. The DO can be searched at <http://disease-ontology.org/>. If your disease terms are not found in DO, let us know and we can get them added to the ontology.

Citation—All data loaded into ZFIN require a data load citation to which the data are attributed. This will be a ZDB-PUB ID for the data load publication and an optional PubMed ID for any related journal article. Please see “Citations” in the Mutant and Transgenic line section of this chapter for further details.

What happens to your data after submission to ZFIN?

Data that are submitted to ZFIN have many individual pieces of information that identify specific things such as genes, anatomical structures, antibodies, probes, constructs, genotypes, etc. Before submitted data can be loaded into ZFIN, it must be established whether records in the incoming data already exist in the ZFIN database. Quality control standards are also applied that may go beyond those used during data collection. Below we describe some of the primary data validation and quality control processes we use to treat data before entering them into the ZFIN database.

Gene Identification

When genes are specifically identified in a data submission, it is essential that the genes be correctly identified in the ZFIN database. Because gene symbols can change, we request a gene symbol and an additional stable accession number to identify the gene. This is very straightforward if ZDB-GENE IDs are provided. If another sequence accession number is provided, genes are identified first by matches to the provided accession number. Accession numbers for genes that cannot be identified by that method will be run through a set of scripts designed to help curators identify the correct gene in ZFIN using a BLAST analysis process known as the “Redundancy Pipeline”. When the redundancy pipeline fails to locate an existing gene in ZFIN based on the provided sequence accession, a new gene record may be created depending on the nature of the data being submitted. A second BLAST analysis method is then used to help curators determine the best name possible for this new gene based on sequence similarity to mouse and human genes. This is known as the “Nomenclature Pipeline”. These BLAST-based analyses can generate significant work and hence can significantly delay completion of a data load.

Morpholino, TALEN, CRISPR Identification and Target Validation

Incoming sequences for morpholinos, TALENs, and CRISPRs will be used to identify any that may already have a record in ZFIN. When target genes are not provided in the data submission, morpholino, TALEN, and CRISPR sequences will be used to identify the target genes unambiguously in the ZFIN database. If an exact match is not found for a target gene or a morpholino, TALEN, or CRISPR, a new record for it may be created to support the incoming data depending on the nature of the data being submitted.

Anatomy Term and Stage Validation

Anatomy terms in the ZFA have developmental stages from the ZFS assigned to them to indicate when they develop (start stage) and when they disappear (end stage). For expression data, the ZFIN database requires that the developmental stage at which the expression data was collected overlaps with the developmental stages at which the labeled structure is

present according to the ZFA. This relationship is validated during expression loads and updates are requested for instances that violate this ZFS constraint on the ZFA.

EQ Syntax Validation

There are quality control restrictions that govern which PATO qualities are valid for use with which phenotype entities. For example, it is not permitted to submit an EQ observation for “cell migration”:”decreased size” because the PATO term “decreased size” is disallowed for use with GO Biological Process terms and is only allowed for use with physical entities such as “eye”. Although many of these constraints follow common sense, strict validation ensures more consistent results for data integration, computed reasoning, and searches.

Antibody Identification

When information about a new antibody is being submitted, the ZFIN antibody records are queried to determine whether the antibody already has a record in ZFIN and to confirm that the antibody details provided in the data submission agree with details for the antibody that already exists in ZFIN. When an antibody record in ZFIN cannot be unambiguously identified, a new antibody record may be created to support the data load depending on the nature of the data being submitted.

Data Submission Templates

Several pieces of data included in data submissions come from ontologies or constrained lists of valid choices (Table 16). Submission of data that do not come from these term sets creates a challenge during data submissions.

To help promote use of correct ontology terms and the constrained term sets, a Google Spreadsheet has been produced that includes tabs for each of the data files that can be submitted (https://docs.google.com/spreadsheets/d/1p7e6LyxU1wSOBd4q8Fon0f6Kf5w-O_xPF-QTmDZSwJc/edit?usp=sharing). To use the spreadsheet, log in to a Google account and save a copy of the workbook for your data. Columns that accept values from a constrained set offer a pick list of valid choices to choose from and restrict entry to valid terms only. Columns that accept terms from specific ontologies use the OntoMaton (13) plug-in for Google Spreadsheets to restrict valid choices to terms from the current version of the correct ontology at BioPortal (14). The OntoMaton plugin will not be available if the workbook is saved as an Excel file. Once your data are gathered, you can share the files with ZFIN or export to Excel and send them to ZFIN. These basic data templates are offered to help researchers gather the correct data for a submission to ZFIN. Files that may be required for each type of data submission, regardless of how these are produced, are listed in table 17.

Citations

1. LaFave MC, Varshney GK, Vemulapalli M, Mullikin JC, Burgess SM. 2014 A defined zebrafish line for high-throughput genetics and genomics: NHGRI-1. *Genetics* 198:167–70. [PubMed: 25009150]
2. Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, Ashburner M. 2005 The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol* 6:R44. [PubMed: 15892872]

3. Van Slyke CE, Bradford YM, Westerfield M, Haendel MA. 2014 The zebrafish anatomy and stage ontologies: representing the anatomy and development of *Danio rerio*. *J Biomed Semantics* 5:12. [PubMed: 24568621]
4. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995 Stages of embryonic development of the zebrafish. *Dev Dyn* 203:253–310. [PubMed: 8589427]
5. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. 2015 ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res* gkv1031.
6. Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE. 2009 Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS Biol* 7:e1000247. [PubMed: 19956802]
7. Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K, Knight J, Mani P, Martin R, Moxon SAT, Paddock H, Pich C, Ramachandran S, Ruef BJ, Ruzicka L, Schaper K, Shao X, Singer A, Sprunger B, Van Slyke CE, Westerfield M. 2013 ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Res* 41:D854–60. [PubMed: 23074187]
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–9. [PubMed: 10802651]
9. Mungall CJ, Gkoutos G V, Smith CL, Haendel MA, Lewis SE, Ashburner M. 2010 Integrating phenotype ontologies across multiple species. *Genome Biol* 11:R2. [PubMed: 20064205]
10. Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D, Kent WJ. 2014 Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30:1003–5. [PubMed: 24227676]
11. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002 The human genome browser at UCSC. *Genome Res* 12:996–1006. [PubMed: 12045153]
12. Schriml LM, Mitraka E. 2015 The Disease Ontology: fostering interoperability between biological and clinical human disease-related data. *Mamm Genome* 26:584–9. [PubMed: 26093607]
13. Maguire E, González-Beltrán A, Whetzel PL, Sansone S- A, Rocca-Serra P. 2013 OntoMaton: a bioportal powered ontology widget for Google Spreadsheets. *Bioinformatics* 29:525–7. [PubMed: 23267176]
14. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, Musen MA. 2011 BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res* 39:W541–5. [PubMed: 21672956]

Figure 1. Summary of the data submission process.

Black boxes indicate work done by the data submitter, light gray boxes indicate work done by ZFIN, gradient filled boxes indicate iterative steps where work is shared between ZFIN and the data submitter.

Figure 2.
The basic elements of a transgenic construct

Table 1.

Data for Submitting Genotypes to ZFIN

REQUIRED	OPTIONAL
Genomic Feature	Feature Maternal Zygosity
Feature Zygosity	Feature Paternal Zygosity
Genetic Background	

Table 2.

Data for Submitting Transgenic Lines to ZFIN

REQUIRED	OPTIONAL
Genomic Feature	Link to Alternate Resource
Affected Gene Symbol	Insertion Accession
Affected Gene Accession	Note
Transgene Type	
Mutagen	
Subject	
Construct	
Laboratory of Origin	
Citation	

Table 3.

Data for Submitting Mutants to ZFIN

REQUIRED	OPTIONAL
Genomic Feature	Link to Alternate Resource
Affected Gene Symbol	Mutant Sequence Accession
Affected Gene Accession	Note
Mutation Type	
Mutagen	
Subject	
Laboratory of Origin	
Citation	

Table 4.

Standard Lines in ZFIN

Abbreviation	Full Name	Description
AB	AB	Any of the inbred lines derived from the original Streisinger A and B incrosses. Includes AB* and ABC.
AB/TL	AB/Tupfel long fin	Mixed AB/ Tupfel long fin line either a maintained inbred line or a novel cross.
AB/TU	AB/ Tuebingen	Mixed AB/Tuebingen line either a maintained inbred line or a novel cross.
C32	C32	Either C32 derivatives from the Steve Johnson laboratory or the Kimmel lab
KOLN	Cologne	A Wild-Type line originally from the Campos-Ortega laboratory that has short fins.
DAR	Darjeeling	Wild-type line collected in Darjeeling, India by Heiko Bleher in 1987. Line maintained by inbreeding.
EKW	Ekkwil	Wild-type line from Ekkwill Breeders in Florida
HK	Hong Kong	Stock obtained from Hong Kong fish dealer.
IND	India	Stock obtained from expedition to Darjeeling (wild isolate).
NA	Nadia	Wild type line from the Nadia district. Original stock collected from stagnant ponds and flood plain. Inbred.
NHGR-1	NHGR-1	Fully sequenced inbred line derived from a Tuebingen/AB cross (1)
RW	RIKEN WT	Wild type line distributed by RIKEN.
SAT	Sanger AB Tuebingen	The AB/Tuebingen line derived from double haploid fish used by Sanger for genomic sequencing.
SJA	SJA	AB derived line that is bred to reduce polymorphism.
SJD	SJD	Sibling line to Darjeeling. Inbred to reduce polymorphisms.
TU	Tuebingen	Short fins, original stock from a Tuebingen pet shop.
TL	Tupfel long fin	Homozygous for <i>leo</i> ^{tl} and <i>lof</i> ^{dt2} .
TLN	Tupfel long fin nacre	The TL-derived TLN wild type strain carries a mix of molecularly uncharacterized <i>mitfa</i> (nacre) ^{s170} and <i>mitfa</i> (nacre) ^{s184} in the background. TL is homozygous for <i>cx41.8</i> (<i>leo</i>) ^{tl} and <i>lof</i> ^{dt2} .
WIK	WIK	The WIK line is very polymorphic relative to the TU line.
WT	Wild type	Used to denote any wild type not listed above.

Table 5.

Data Required for Specifying Multiple Affected Genes

Required Data	Description
Genomic Feature	The unique identifier for the mutant
Affected Gene	The symbol for the affected gene
Affected Gene Accession	A ZDB-GENE ID or sequence accession number for the affected gene
Relationship	The relationship between the genomic feature and the affected gene. One of: gene missing, gene present, gene moved, is allele of gene

Table 6.

Mutation Types Supported in ZFIN

Mutation type	Definition	Notes	Data to provide	SO ID
point mutation	A single nucleotide change which has occurred at the same position of a corresponding nucleotide in a reference sequence.	Can be an allele of a single gene.	Affected gene and its ZDB-GENE ID	SO:1000008
small deletion	The point at which one or more contiguous nucleotides were excised.	The excision is within a single gene.	Affected gene and its ZDB-GENE ID	SO:0000159
insertion	The sequence of one or more nucleotides added between two adjacent nucleotides in the sequence.	Usually an allele of a single gene.	Affected gene and its ZDB-GENE ID	SO:0000667
indel	A sequence alteration which includes an insertion and a deletion, affecting 2 or more bases.	Usually an allele of a single gene.	Affected gene and its ZDB-GENE ID	SO:1000032
translocation	A region of nucleotide sequence that has translocated to a new position. The observed adjacency of two previously separated regions.	Has at least one breakpoint, frequently within a gene. Has genes that are in a new genomic context	Genes at breakpoint and their ZDB-GENE IDs. Genes that have been relocated and their associated ZDB-GENE IDs	SO:0000199
inversion	A continuous nucleotide sequence is inverted in the same position.	Has two break points which may occur in one or more genes and may have additional genes within the inverted sequence.	Genes at breakpoint and their ZDB-GENE IDs. Genes that have been relocated and their associated ZDB-GENE IDs	SO:1000036
deficiency	An incomplete chromosome.	The chromosome is missing more than a single gene. Has two break points. Other genes existing between the break points may have also have been lost.	Genes at breakpoint and their ZDB-GENE IDs. Genes that have been lost and their associated ZDB-GENE IDs	SO:1000029
unknown	A mutation where the lesion type is unknown.	May be an allele of a gene or in an unknown location.	Affected gene and its ZDB-GENE ID if known	NA

Table 7.**Mutagen Types Supported in ZFIN**

Mutagen Type	Description
TALEN	Transcription activator-like effector nucleases (TALENs) are nucleases specifically designed to cleave a DNA sequence of interest. Provide the name of the TALEN.
CRISPR	Clustered regularly interspaced short palindromic repeats (CRISPRs) are specifically designed to recruit the Cas9 enzyme to cleave DNA at a targeted DNA locus. Mutations often result from the subsequent DNA repair event. Provide the name of the CRISPR.
ENU	N-ethyl-N-nitrosourea, ENU, is a chemical alkylating agent and mutagen when applied to animals.
TMP	4,5',8-trimethylpsoralen is a DNA cross-linking agent which often produces deletion mutations.
Gamma Rays	Ionizing electromagnetic radiation used to induce mutations. Often produces large deletions and chromosomal aberrations.
Spontaneous	de novo mutations not generated by the application of an external mutagen.
Zinc Finger Nuclease	Zinc finger nucleases are artificial restriction enzymes designed to target and cleave specific DNA sequences.
DNA	DNA sequence, usually a transgenic construct, injected into embryos to create heritable transgenic insertions

Table 8.

Data for Citation of a Data Submission

REQUIRED	OPTIONAL
Author list	PubMed ID
Publication title	
Abstract describing data and methods	

Table 9.

Data for Submitting Transgenic Constructs to ZFIN

REQUIRED	OPTIONAL
Construct Name	Link to Alternate Resource
Promoter Gene Symbol	Construct Accession
Promoter Accession	Construct Map Image Name
Coding Sequence Gene Symbol	Note
Coding Sequence Accession	
Engineered Region Name	
Citation	

Table 10.

Data for Submitting Morpholinos, TALENs, and CRISPRs to ZFIN

REQUIRED	OPTIONAL
MO/TALEN/CRISPR Name	Link to Alternate Resource
Target Sequence 1	Note
Target Sequence 2 (TALEN Only)	Citation PMID
Target Gene Symbol	
Target Gene Accession	
Data Load Citation	

Table 11.

Data for Submitting Expression Data to ZFIN

REQUIRED	OPTIONAL
Expressed Gene Symbol	Image/Movie File Name
Expressed Gene Accession	Antibody Name (for Immunoassays)
Genotype	Probe GenBank Accession # (for Hybridization Assays)
Morpholinos, TALENs, CRISPRs	Antibody Name (for Immunoassays)
Anatomical Structure	
Developmental Stage	
Experimental Conditions	
Citation	
Assay Type	

Table 12.

Experimental Conditions in ZFIN

Condition	Description
standard	Experimental condition that is the standard environment for zebrafish husbandry, as described in The Zebrafish Book. In general the standard environment utilizes contaminant free tank water, heated to 28.5°C, with the fish fed a normal contaminant free diet, with standard osmolarity, pH, and normal light cycle of 14hr light/10hr dark.
generic control	Experimental condition that is used as a reference point to compare with results of treated zebrafish. Generic experimental controls often use sham injections, injections of vehicle, injections of control MOs, etc. This environment is used for non-standard conditions used in control treatments.
chemical	Experimental condition in which the fish is treated in tank water, or by injection or consumption, with a chemical substance. The ChEBI ID for the chemical should be included in the data submission.
pH, acidic	Experimental condition in which the pH of the water is lower than the pH of the controlled conditions.
pH, basic	Experimental condition in which the pH of the water is higher than the pH of the controlled conditions
electric field	Experimental condition in which an electric field is applied to the fish, fish cells, or organs as compared to control conditions.
gravity	Experimental condition in which the fish is exposed to forces that simulate low or high gravity as compared to earth's gravity.
hyperoxia	Experimental condition in which the oxygen (O ₂) concentration is higher than the one in controlled conditions.
hypoxia	Experimental condition in which the oxygen (O ₂) concentration is lower than the one in controlled conditions.
light	Experimental condition in which the intensity, wavelength, and/or duration of illumination is (are) different from the one in controlled conditions.
magnetic field	Experimental condition in which the fish is exposed to a magnetic field as compared to control conditions. A magnetic field is a region in which the force of magnetism is applied.
mechanical stress	Experimental condition in which an external force is applied to the fish or part of the fish.
radiation	Experimental condition in which the fish is exposed to ionizing and/or non-ionizing radiation. The radiation could be ionizing such as gamma rays, alpha particles, UV, X-ray and non-ionizing such as infrared, microwaves etc.
bacterial infection	Experimental condition in which fish have been infected with bacteria. This infection can be done by addition of bacteria in the water or by injection of bacteria, (for example in the brain ventricle, in the caudal vein, in the yolk sac), or ingestion, or other means.
cancer	Experimental condition in which cancer cells are introduced to the fish via injection of tumor cells.
fungal infection	Experimental condition in which fish have been infected with a fungus.
germ free	Experimental condition in which fish were raised in the absence of bacteria
high calorie diet	Experimental condition in which fish are fed a high calorie diet as compared to the normal diet.
low calorie diet	Experimental condition in which fish are fed a low calorie diet as compared to the normal diet
organ culture	Experimental condition in which an organ is dissected/isolated/collected from the fish and placed in culture. The analysis of the experiment is done on this organ in culture.
primary cell culture	Experimental condition in which an embryo or adult fish is dissociated to a single cell suspension. The analysis is made on this cell culture.
regeneration/healing	Experimental condition in which fish's organ (e.g. heart) or anatomical structure (e.g. fin) was wounded or amputated.
starvation	Experimental condition in which fish were deprived of food.
Salinity, hypertonic	Experimental condition in which the salt concentration is higher than the one in controlled conditions.
Salinity, hypotonic	Experimental condition in which the salt concentration is lower than the one in controlled conditions.
temperature, cold shock	Experimental condition in which fish are subjected for a short period of time to temperature lower than the controlled temperature. The standard controlled temperature (according to The Zebrafish Book) is 28.5°C
temperature, heat shock	Experimental condition in which fish are subjected for a short period of time to temperature higher than the controlled temperature. The standard controlled temperature (according to The Zebrafish Book) is 28.5°C

Condition	Description
temperature, stable	Experimental condition in which fish are raised in temperature different (lower or higher) than the controlled temperature. The standard controlled temperature (according to The Zebrafish Book) is 28.5°C

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 13.

Data Used to Create New Antibody Records

Data	Description
Host Organism	Organism from which the antibody was made.
Immunogen Organism	Species from which the immunogen was obtained. If it is a peptide based on a sequence from a particular organism list that organism.
Antibody Type	List whether antibody is polyclonal or monoclonal.
Antibody Isotype	List isotype if known. (optional)
Source	If the antibody was purchased from a commercial supplier, list the supplier.
Catalog Number	If the antibody is from a commercial supplier please provide the catalog number. (optional)
Name	Include clone names if known. (optional)
Note	Include sequence of peptide or accession number of sequenced used to produce antibody if the antibody was custom made. Also include any usage notes here. (optional)
Target Gene	Provide ZDB-GENE ID or sequence accession number for the target gene if known (optional).
Citation for Original Source	If this is a previously published antibody please provide a reference PubMed ID. Otherwise the antibody will be attributed to the data load publication.

Table 14.

Data for Submitting EQ Phenotypes to ZFIN

REQUIRED	OPTIONAL
Genotype	Image or Movie File Name
Morpholinos, TALENs, CRISPRs	
Developmental Stage	
Experimental Conditions	
Phenotype Entity	
Phenotype Quality	
Tag	
Citation	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 15.

Required track hub configuration information

Label	Description
track	The track file name
shortLabel	A brief (17 character) label to describe the track in the genome browser. Visible to the left of the track in the genome browser. Example: 4 day methylome
longLabel	A longer label (76 character) to describe the track in the genome browser. Visible above the track in the genome browser. Provide enough detail to uniquely identify the track. Example: Howe et al. 2015 male 4 day methylome
type	States the track file format (bigWig, bigBed, etc.)

Table 16.

Sources of valid values for data submissions

Data Type	Source of Valid Values
Anatomy	Zebrafish Anatomy Ontology
Developmental Stage	Zebrafish Developmental Stages
Human Disease	Human Disease Ontology
Biological Processes	Gene Ontology
Experimental Condition	Constrained list of experimental conditions (table 12)
Mutagen	Constrained list of mutagens (table 6)
Subject	Constrained list of subjects (table 7)
Phenotype Entity	Gene Ontology or Zebrafish Anatomy Ontology
Phenotype Quality	Phenotypic Trait Ontology
Genetic Background	The list of standard lines at ZFIN

Table 17.

Data submission files for each data type

Data Type Being Submitted	Data Sheets to Submit	Other Files
Mutants	Mutants/Transgenics Genotypes Citations	
Transgenics	Constructs Mutants/Transgenics Genotypes Citations	Construct Image
Phenotype	Mutants/Transgenics Genotypes Citations Constructs Phenotypes	Media Files
Expression	Mutants/Transgenics Genotypes Citations Constructs Expression	Media Files
Morpholinos, TALENs, CRISPRs	MO/TAL/CRSP Citations	
Genome Browser Tracks	TrackInfo Citations	Track File
Disease Models	Mutants/Transgenics Genotypes Disease Models Citation	
Antibodies	Antibodies	

Following these guidelines will help ensure a smooth experience for researchers who wish to submit data to ZFIN.