

Executive Summary: *Xenopus* Transgenic Nomenclature Guidelines

These guidelines cover the naming of transgenic constructs developed and used in *Xenopus* research, and the naming of mutant and transgenic *Xenopus* lines .

- Only those transgenic (Tg) constructs that are supplied to other researchers, or are maintained by a lab or at a stock center, require standardized names following these guidelines.
- Xenbase will record and curate transgenes using these rules.
- Transgenic construct names are italicized following this basic formulae:

Tg(enhancer or promoter:ORF/gene)

- Multiple cassette on the same plasmid are separated by a comma within the parenthesis.

Tg(enhancer or promoter:ORF/gene, enhancer or promoter:ORF/gene)
cassette 1 cassette 2

- Transgenic *Xenopus* lines are named based on their unique combination of Tg constructs (in standard text) and indicate *Xenopus* species and the lab of origin:

XI.Tg(enhancer or promoter:ORF/gene, enhancer or promoter:ORF/gene)^{#ILAR lab code}
species cassette 1 cassette 2 lab code

- *Xenopus* research labs that generate Tg animals are identified using lab codes registered with the [ILAR](#) which maintains a searchable database for model organism laboratories.

Section 1. Provides an overview for Naming Tg Constructs

Section 2. Outlines specific rules for Naming Tg Constructs

Section 3. Provides an overview for Naming Tg *Xenopus* lines and *Xenopus* lines with mutant alleles.

Section 4. Provides guidelines for naming enhancer and gene trap constructs, and lines using these constructs.

Section 5. Outlines rules for naming transposon-induced mutations and inserts.

Section 6. Discusses disambiguation of wild type *Xenopus* strains versus transgenic and mutant *Xenopus* lines.

Section 7. Outlines ILAR lab code requirements.

Section 8. Future plans for using RRID numbers is discussed.

Appendix 1: List of prefixes for promoters/gene of origin from non-*Xenopus* species.

Appendix 2: List of ILAR Lab codes for sources of Tg *Xenopus* lines and mutant alleles.

Please address all comments or questions to the transgenic nomenclature administrator at Xenbase: xenbase@ucalgary.ca

Xenopus Transgenic Nomenclature Guidelines

(draft 10/23/2014)

Numerous transgenic constructs are often described in a single publication but not all are subsequently maintained in Tg lines or archived. Only those transgenic (Tg) constructs that are supplied to other researchers, or are maintained by a lab or at a stock center, require standardized names following these guidelines.

1. Naming Transgenic Constructs- Overview

STANDARD Tg CONSTRUCT NAME = *Tg(promoter:ORF/gene)*

- Brevity and clarity in the transgenic construct name are encouraged.
- The entire transgene/construct name should be *italicized*.
- *Tg* indicates transgene.
- Within the parentheses, the most salient features of the transgene should be described.
- A colon separates the enhancer/promoter from the ORF/gene name.
- Regulatory sequences should be listed to the left of the colon, and coding sequences to the right of the colon.
- No spaces are used between abbreviations, terms and punctuation marks.
- Not all transgenic constructs will have both regulatory and coding elements, and in this case, the colon will not be used.
- In cases where a construct utilizes sequences from a named gene, it should contain the standard lowercase symbol for that *Xenopus* gene. Gene symbols are in all

lowercase (See Xenbase for official gene symbols and gene nomenclature guidelines <http://www.xenbase.org/gene/static/geneNomenclature.jsp>).

- If the gene is from another species, use the 3 letter code for the species, a period (.) and the gene symbol in lower case. For human genes, gene symbol is capitalized. Xl and Xt are reserved for official transgenic line names. Do not use an Xl, Xt, X, or x, to indicate *Xenopus in construct names*. See Appendix 1 for a list of other species abbreviations.
- For those cases where a specific transcript or transcript promoter of a gene is used, the transcript number or name should be used.
- The method of how the Tg was made (eg sperm-mediated, CRISPr/CAS, pTransgenesis) is not captured in the name, but recorded in the description of the construct.
- Fluorescent proteins are either capitalised (eg, GFP), or start with a capital letter (egs., Katushka, Tomato), but are **not** underlined. Prefixes to fluorescent proteins are in lower-case such 'e' for 'enhanced' (egs., eGFP, mCherry).
- The size of the promoter sequence will **not** be included- it will be recorded in the construct description. Constructs utilizing different sizes of coding sequence are distinguished by a serial number, in superscript.
- Names for transgenic and mutant lines of *Xenopus* are based on the names of the Tg constructs. See **Section 3. Naming Transgenic *Xenopus* Lines**.

Xenbase has an independent system for assigning attribution or credit to a lab, a person, and/or an organization for each transgenic construct and Tg line/wt strain, but these unique XB-numbers are not to be used in the transgene names.

Xenbase will record:

- Attribution: XB-LAB-#, XB-PERSON-#, XB-ORG-#
- Literature citation(s).
- Legacy names, recorded as synonyms of the official name.

Note: Examples in green (real and hypothetical) are from Xenopus, blue text indicates zebrafish and pink text examples are from mice. As Xenbase curation proceeds, we will endeavour to find Xenopus examples of all naming cases.

2. Naming Transgenic Construct Names

STANDARD Tg CONSTRUCT NAME = Tg(promoter:ORF/gene)

2.1 USE OF COMMAS ,

a) Multiple cassettes are separated by a comma.

Frequently within a single Tg construct, there are multiple cassettes, each containing regulatory and coding sequences. In these cases, it is necessary to distinguish between what is coding in the first cassette with what is regulatory in the second.

Example 1: *Tg(cryga:GFP,hsp70:nog)*

In this example, the *Xenopus* γ -crystallin promoter drives GFP (cassette 1), and noggin is under control of the heat shock promoter hsp70 (cassette 2).

b) When one promoter is used to drive more than one ORF or gene, a comma is used to separate the ORF/gene names. This includes unidirectional and bidirectional promoters.

Example 2: *Tg(CMV:GFP, RFP)*

In this example, the constitutive simian cytomegalovirus (CMV) IE94 promoter/enhancer drives GFP, but is flanked by FRT sites (for FLP site specific recombination). After FLP recombination, CMV drives RFP. The FRT/FLP sites are not included in the Tg name, and full details are given instead in the construct description.

2.2 USE OF HYPHENS -

Hyphens are used in 3 different circumstances:

1) A hyphen is included in the gene symbol for some enhancers/promoters. This is consistent with the standard for *Xenopus* gene nomenclature guidelines.

Example 3: *Tg(nkx2-5:GFP)*

2) Regulatory fusions or coding sequence fusions (ie., fusion proteins) should be separated by hyphens. Regulatory sequence could be derived from either an enhancer or promoter, and is denoted by the symbol of the regulated gene or gene transcript.

Example 4: *Tg(CMV:eGFP-CAAX)* In this case the construct has a CMV promoter, and a Ras protein CAAX box (with a farnesylation signal) is fused with eGFP to force the protein to be localized in the plasma membrane.

Example 5: *Tg(CMV:Hsa.EB3-eGFP)* In this case the construct has a fusion protein of human (Hsa) EB3 gene and enhanced GFP under the control of a CMV.

3) A hyphen is used to separate the enhancer or promoter from a transposase

using Tg (for transgene), followed in parentheses by the enhancer/promoter and the symbol for the transposase, separated by a hyphen. This distinguishes the construct from a regular Tg construct with the enhancer/promoter and ORF/gene separated by a colon.

See Examples 40-43, Section 5.2, where transposase inserts that used to generate *Xenopus* lines are discussed.

2.3 USE OF FULL STOP/PERIOD

The period (.) in transcript name demarcates which homeologue of a gene is used. This rule is consistent with the standard for *Xenopus* gene nomenclature guidelines, and only applies to *X. laevis*, as it is a pseudo-tetraploid and thus carries a duplicated genome.

Example 6: *Tg(msgn1.b:GFP)* or *Tg(msgn1.S:GFP)*.

In this case the b homeologue *X. laevis* msgn1 gene is the promoter driving expression of GFP. Gene nomenclature will soon change to naming homeologues as L or S (instead of a or b), and in this case the period will be used in the same way

2.4 USE OF SUPERScript NUMBERS (eg, *Tg(gene:ORF)¹*, *Tg(gene:ORF)²*)

Although the length of a sequence will not be used in the name of a Tg construct, constructs that are different from each other by a small number of nucleotides should be given distinct numbers, where a sequential number in superscript after the parentheses will indicate that the constructs differ somewhat in the sequences used in their construction.

Note: only those construct maintain and/or shared between researchers need names, and Tg lines carry the construct name- see section

Example 7: Tg construct 1 = *Tg(col2a1:GFP)¹*
Tg construct 2 = *Tg(col2a1:GFP)^{2.3}*
Tg line = *Xl.Tg((col2a1:GFP)^{2.3Hank}*

In this example, several col2a1 constructs were generated by the Hanken lab, which they consecutively numbered. Only 2 constructs, referred to as '1' and '2.3', worked efficiently and generated maintained *X. laevis* lines (see rules in Section 3 for naming Tg lines). Other constructs used in these experiments are not given official stable names.

2.5 WHEN MORE THAN 1 GENE IS INVOLVED

For those situations where a construct utilizes enhancers or promoter that regulate two or more genes, only one of the genes needs to be represented. Researchers can choose to name the Tg construct with either the gene with the lowest number in the gene family or the gene closest to the promoter. Other gene components of the construct will be detailed in the construct description.

Example 8: *Tg(dlx1aIG:GFP)*

This construct utilizes intergenic (IG) regulatory elements of *dlx1a* and *dlx2a* to drive expression of GFP. In this case the lower numbered gene (*dlx1a*) was listed in the name.

Example 9: *Tg(zic4:GAL4-TA4,UAS:mCherry)*

This construct utilizes an enhancer region of the *zic1* and *zic4* genes to drive expression of GAL4-TA4, with an additional cassette that has UAS driving mCherry expression. In this case the **gene closest to the enhancer** was listed in the name.

2.6 ABBREVIATIONS FOR OTHER SPECIES: THE 3 LETTER RULE (and another use of ‘.’)

For those cases where a gene from a different species is used, the **three letter abbreviation** should be used (eg., *Homo sapien* (Hsa), *Mus musculus* (Mmu), *Salmo salar*(Ssa)) **followed by a period, then the gene symbol**. For human genes, use the standard gene symbol convention of all capital letters. For all other species the gene symbol is all lowercase.

See Appendix 1 lists commonly used species and their abbreviations.

Note: An exception to the 3 letter rule is *Chlamydomonas reinhardtii*. Please use Cr for this organism as the 3-letter abbreviation (Cre) conflicts with the abbreviation for the Cre-Lox system.

Example 10: *Tg(Hsa.UBC:GFP)*

Here the promoter of the Human (*Homo sapiens*, Hsa) Ubiquitin C promoter is driving expression of GFP. Note: Human gene symbols are capitalized.

Example 11 : *Tg(Dre.noto:GFP)*

Here the promoter of the zebrafish (*Danio rerio*, Dre) notochord homeobox (noto) gene is driving expression of GFP.

Example 12: *Tg(tetRE:Mmu.axin1-YFP)*

In this case the construct has a fusion protein of mouse (Mmu) axin1 gene and YFP under the control of a tetracycline response element (tetRE).

2.6 INDICATING MODIFIED CLONE NAMES

Transgenic constructs using modified clones such as BACs and PACs should be named with the clone type inserted between the "Tg" and the parentheses.

Example 13: *TgPAC(tal1:GFP)*.

Here GFP is inserted within or near the coding sequence of *tal1* in the PAC clone with the GenBank# AL592495. The accession number of the clone must be included in the publication/submission, so that it can be associated with the construct. A link to the appropriate clone will be on the Xenbase construct page.

2.7 USING MULTIMERS

Multimers are the numeric multipliers which indicate that more than one copy of a transgenic element is integrated into a transgenic component, eg: two responsive or six repressor elements, which increase the efficiency of that component. Only multimers that are crucial to the construct are included in the official name. Other elements, such as insulators, LoxP FRT and FLP sites, should be listed only in the construct description for the sake of simplicity.

Example 14: *Tg(7xTCF-sia:nls-mCherry)*

This construct utilizes seven TCF responsive elements fused with *Xenopus* siamois (*sia*) minimal promoter to drive expression of nls-mCherry.

2.8 Tg CONSTRUCTS DEVELOPED IN OTHER MODELS.

If a Tg construct was originally developed in a different model (eg., the zebrafish, *Danio rerio*) using a *Xenopus* gene, the zebrafish community would indicate "Xla" as the species of origin for the Tg construct. If a *Xenopus* researcher then uses this same construct, this complete original name should be used, so that the source of the reagents/tools is clear and traceable to other scientists. However, if the 'same' construct was independently generated by a *Xenopus* researcher, the species identifier 'Xla', and the sequence size can be omitted from the construct name. (See section 3.5 and Example 25, dealing with Tg *Xenopus* that use elements/sequence from alternate *Xenopus* species.)

Example 15: *Tg(-5Xla.pou3f4-gata2a:eGFP)*

This zebrafish transgenic line that carries *Xenopus laevis* genes sequence. Note: Zfin includes the insert size in zebrafish Tg curation.

Example 16: *Tg(pou3f4-gata2a:eGFP)*

The essentially same Tg construct in Example 16 from a *Xenopus* lab, omits size of sequence in construct and the species' sequence to which the construct was designed is assumed to be *Xenopus*, thus omits the abbreviation Xla.

2.9 USE OF UNDERSCORES and ARROWS, _ >

When a mutated form of a gene is used in a construct, the **mutations in the gene are noted after an underscore.**

It is critical that an NCBI/GenBank accession number be included in the manuscript's description of the mutant because some genes have multiple entries/sequences. This accession number will be noted on the Xenbase Tg construct page, and relevant sequences will be uploaded.

2.9.1 If the mutation results in an amino acid change/substitution, the following are listed after the underscore: original amino acid residue, amino acid residue number, engineered/mutant amino acid residue.

Example 17: *Tg(pax2a:Rno.fgf8a_S245A)*

In this example, the amino acid mutation is in rat (Rno) fgf8 gene at position 245 where S has been replaced by A.

2.9.2 If the mutation is a nucleotide change/substitution, the following are listed after an underscore: the nucleotide position, the original nucleotide, the symbol '>', and the engineered/mutant nucleotide.

Example 18: *Tg(Hsa.MPZ_1026T>A:eGFP)*

In this example, the nucleotide mutation is in the Human gene promoter MPZ at position 1026 where T has been replaced by A, driving enhancer GFP expression.

2.10 Indicating dominant negative constructs (dn).

Construct that use deleted or truncated forms of a gene to create dominant negative expression are commonly used, and are more succinctly prefixed with 'dn'.

Example 19: *Tg(Rno.elas:^{dn}thra:GFP,cryga:GFP)*

In this example from the Brown Lab, the rat (Rno) elastase promoter is driving the thyroid hormone receptor alpha sequence which lacks 36 bp from its C terminus, and that is fused to GFP by a 15-bp linker. This is a dominant negative form of thra, as use of capitalised TRDN conflicts with the symbol for the human gene 'triadin', and the *Xenopus* convention is to use lowercase official gene symbols, thus we use thra instead of TR. We suggest following the normal use of dn for dominant negative, but putting them in superscript to avoid confusion with gene symbols starting with d' or 'dn'. The hyphen is still used to show it is a fusion protein with GFP. Further details of the construction would be fully explained in the construct description.

The original transgenic *X. laevis* lines made using this construct would then be named $Xl.Tg(Rno.elas:dnthra-GFP,cryga:GFP)^{1Brown}$, see Section 3 below.

Example 20: *Tg(hsp70l:dnHsa.BMPR1A-CFP)*

When the dominant negative form of the gene/protein is from another species, as in this example the Human BMPR1 gene, the dn subscript precedes the species (Hsa) abbreviation.

2.11 Indicating constitutively active transgenes (ca)

In cell biology, a constitutively active protein is a protein whose activity is constant and active. Use the 'ca' in superscript, preceding the gene symbol, to avoid confusion with the 100's of gene symbols starting with 'ca'.

Example 21: *Tg(hsp70l:cafgfr1,Dre.cryaa:dsRed)*

This construct allows heat shock-inducible expression of a constitutively active form of fgfr1, and a cassette in which a 0.8 kb dsRed coding sequence is driven by 0.7 kb of the zebrafish alpha-crystallin promoter. Note sequence sizes are omitted for *Xenopus* transgenes.

3. Naming Transgenic *Xenopus* Lines- Overview

Standard Tg line name:

Species.Tg(enhancer/promoter:ORF/gene) ^{#Lab code}

- Transgenic lines of *Xenopus* that are maintained by a lab or stock center require standardized names.
- Standardized names should be used in publications.
- The transgenic line name indicates *Xenopus* species.
- The transgenic line name is derived from the transgenic construct used to generate the line, and follows the same syntax and format.
- Tg lines are in standard text (i.e. not italics).
- ILAR Lab codes follow the Tg line name, in superscript.
- A serial number precedes the ILAR lab code (see Appendix 2) for each unique Tg line, starting with 1. This is a preemptive measure, and any future line will then be number 2 without the need to rename the initial line.

Note: The following additional information is not captured in the name of the Tg lines with mutations or in standard Tg lines but, where available, should be recorded in the description:

- Wt background strain/Pigmentation: eg., J-strain, Ivory Coast, albino
- Sequence of insertion
- Accession number
- Method: eg., CRISPr mediated mutation in exon 1
- Location eg., scaffold number or mutation position
- Legacy name/Synonym
- Location of LoxP and FRT sites
- Image of the construct/design

3.1 Indicating Species.

Transgenic *Xenopus* lines will indicate *Xenopus* species in the official line name, unlike fish, mouse and other models.

The initials Xl or Xt precede the ‘Tg’, and are separated by a period (.).

Xl = *X. laevis*

Xt = *X. tropicalis*

Example 22: Xl.Tg(nkx2-5:GFP)^{1Mead} where Xl = *X. laevis*
Xt.Tg(nkx2-5:GFP)^{1Mead} where Xt = *X. tropicalis*

In this example, the Mead lab generated two Tg lines using the same Tg construct, one in each *Xenopus* species. The wild type strain or background should be recorded in the description of the Tg line, and not in its official name.

3.2 Stable Tg *Xenopus* lines from the same lab.

When new lines of *Xenopus* with unique, stable, and heritable transgenic compositions are generated from another transgenic line, the derived lines should all receive unique allele/line designations. Each unique Tg line name consists of the construct name appended by a sequential (or serial) number and the the ILAR laboratory code.

Example 23: $Xl.Tg(hsp70:GFP)^{1NXR}$ compared with $Xl.Tg(hsp70:GFP)^{2NXR}$

3.3 Stable Tg *Xenopus* lines from different labs.

Independently generated transgenic lines that use the exact same Tg construct are distinguished by the superscript of the ILAR Lab codes that follows the Tg line name.

Example 24: $Xt.Tg(cryga:GFP)^{1Rmng}$ compared with $Xt.Tg(cryga:GFP)^{1EXRC}$

In this example, the first maintained gamma crystallin GFP Tg line of *X. tropicalis* was made by the Grainger lab, and is distinguished from the gamma crystallin GFP Tg line made by the EXRC. Future lines from either lab will be sequentially numbered.

3.4 Hybrid lines and Multiple transgenic constructs in a Tg Line

When lines are produced by using 2 distinct constructs, or by crossing 2 separately derived Tg lines **the new line is described using a semicolon (;) to separate the 2 (or more) Tg components:**

species.Tg(construct 1);(construct 2)^{#Lab}

Example 25: $Xl.Tg(hsp70:Cre);(CMV:eCFP,eYFP)^{1Ryff}$

These *X. laevis* frogs derived from a crossing between two Tg lines from the Ryffel Lab.

3.5 Tg lines with constructs from other *Xenopus* species.

When a Tg line is generated in one *Xenopus* species using a construct with either promoters or enhancers from the other *Xenopus* species, this should be indicated in the Tg name.

Example 26: $Xt.Tg(pax6:Xla.hoxa3)^{1Buschz}$

In this hypothetical example, developed in the Buscholtz lab, the Tg construct is made with the pax6 promoter driving hoxa3 expression, using hoxa3 sequence from *X. laevis*, and the subsequent Tg line is a *X. tropicalis* line.

3.6 Naming Tg *Xenopus* Lines with Mutant Alleles

A mutation is a particular class of variant allele that usually confers a phenotypically identifiable difference to a reference "wild type" phenotype.

Basic rules for naming *Xenopus* Tg lines with mutations in **known genes** (whether or not they confer a phenotype) are described below.

Standard format:

Xenopus species.Tg.(species.gene symbol^{mutant type#})#ILAR lab code

- Species designated as Xl = *X. laevis* and Xt = *X. tropicalis*, followed by period (.)
- Tg for transgene
- In parentheses
 - Official species abbreviation for gene origin if not *Xenopus*
 - Include any aspects of Tg constructs following rules above
 - Official gene symbol (eg., tubb2b, not NBT)
 - In superscript -m#/ del# /tm#. Note: the number can be a serial number (line or founder number) OR if the position of the deletion, use the underscore (_) following rule 2.3
- Also in superscript, outside the parentheses, # ILAR Lab code

Designation for mutant type are as follows:

- 'm' stands for mutant/mutation
- 'del' is for deletion
- 'tm' stands for targeted mutation
- 'ins' stands for insertion

Example 27: **Xl.Tg(CMV:eCFP,Hsa.HNF1^{del_328})^{1Ryff}**

In this Tg *X. laevis* line from the Ryffel Lab, CMV is driving enhanced CFP, and the mutated form of the human HNF1 gene, which has a deletion at nucleotide position 328. The larvae show blue fluorescence upon Cre recombinase action. Locations of LoxP and FRT sites are not included in the name.

3.7 If a new allelic mutation of a gene, known only by a mutant phenotype, is caused by a transgenic insertion, the symbol of this mutation should use the symbol of the transgene (not 'ins') as superscript, thus the basic format is:

species.gene^{Tg#Lab}

Example 28: **Xt.acta1**^{Tg3Amaya}

In this hypothetical example in *X. tropicalis*, a transgenic insertion into the acta1 gene was generated by the Amaya lab, and although a phenotype can be described, the location of the insertion on the loci is unknown.

3.8 Non-transgenic *Xenopus* Lines with Mutant alleles/Mutations in Known Genes

Basic format: species.gene symbol^{mutant type# ILAR lab code}

- Official species abbreviation if not *Xenopus* (eg., Hsa for human genes), followed by period (.).
- Standard case (not italics)
- Official gene symbol (eg., use tubb2b, not NBT)
- In superscript
 - m# / del# /ins# /tm#, where # is the serial number
 - ILAR Lab code

Note: the serial number can be a line or founder number OR the position of the deletions.

In some cases, such as when homologous recombination is used to target a gene, a readily identified phenotype may not result, even though the gene may be rendered non-functional. In such cases, the targeted genes are nevertheless referred to as mutant alleles.

Example 29: **Xt.myod1**^{tm1.3Rmh/+}

This is an hypothetical *X. tropicalis* line with targeted mutation 1(tm1) in the myogenic differentiation 1 gene (*myod1*), established from founder individual no. 3 (this tm1.3), is from the Harland laboratory at UC Berkeley (Rmh). Note there is no promoter, and the /+ denotes that it is a heterozygous line.

3.8.1 If the mutation is shown to be a deletion of all or part of a single gene, the superscript *del* is used in place of *m* or *tm*.

Example 30: **Xt.shh^{del1Rmh}**

In this example, the deletion occurs in the sonic hedgehog gene (*shh*) of a *X. tropicalis* line, made by the Harland Lab. Specifics of the location of the deletion are recorded in the lines' description, as are any phenotypic traits.

Example 31: **Xl.tyr^{tm1NXR1}**

In this example, the a CRISPR-induced tyrosinase mutant (a targeted mutation) was generated by the NXR to create mosaic *X. laevis*.

3.8.2 If the mutation is known to have occurred on a particular *X. laevis* homeologue, that can be specified by preceding the superscript with the homeologue symbol and a hyphen.

Example 32: **Xl.soxx17a-b^{m1Zorn}**

In this example the Zorn Lab confirmed that the mutation was in the *X. laevis* 'b' homeologue of *soxx17a*.

3.9 Non-transgenic *Xenopus* Lines with Mutant alleles/Mutations in Unknown or Known Genomic Locations.

Basic rules for naming mutant *Xenopus* lines carrying mutations in unknown locations where they confer a phenotype:

species.phenotype name/symbol #ILAR lab code

Examples 33: **Xt. kaleidoscope^{1Zimml}**

X. tropicalis tadpoles carrying the mutation conferring *kaleidoscope* phenotype from the Zimmerman lab.

3.10 Non-transgenic *Xenopus* Lines with Mutant alleles/Mutations in Known Genomic Locations.

Basic rules for naming mutant *Xenopus* lines carrying mutations in known locations where they confer a phenotype:

species.gene symbol^{phenotype name/symbol} #ILAR lab code

Examples 34: *Xt.atp7a*^{kaleidoscope1Zimml}

From the mutant line in Example 32, once the location of the mutation was identified as in the *atp7a* copper transporter gene, then it would be renamed using the official gene symbol and the phenotype name in the superscript.

4.0 Enhancer traps (Et) & Gene traps (Gt): Overview

An effective scheme used to generate reporter lines is enhancer or gene trapping, whereby a transgenic cassette is randomly inserted into the genome and co-opts the expression pattern of a nearby gene. (Transposon systems are commonly used to introduce these cassettes. See Section 5 below, for guidelines to name transgenes that are generated with transposon methods). Enhancer trap and gene trap constructs use the same nomenclature convention as described above for transgenic constructs, substituting Et or Gt as necessary. Likewise, *Xenopus* Lines generated using Et and Gt methods, use these symbols to replace Tg.

One common use of enhancer traps is in creating Cre driver lines. Enhancer traps of this type can be created to include a minimal promoter, introns, a Cre recombinase cassette (sometimes fused with another element such as ERT2), and polyA sites from different sources. The engineered regions used to make these constructs, (eg., UAS, LoxP FRT & FLP sites) have been incorporated into enhancer and gene trap names in other models, but it is recommended here to omit them from the Et and Gt line names, and that these engineered regions be fully described and recorded in the description of the Et and Gt constructs name, as they would be for Tg constructs. A diagram of the construct, showing the location of for eg., 1-Scel, LoxP, FRT & FLP sites, can be uploaded on Xenbase, along with the construct record.

4.1 Enhancer Traps

Basic nomenclature format for enhancer traps: *Et(cassette1, cassette2)*

Basic nomenclature for Et *Xenopus* lines: species.Et(cassette)^{#ILAR lab code}

Example 35: Enhancer trap name: *Et(eef1a1:GFP)*

Line name: *Xt.Et(eef1a1:GFP)*^{16FMead}

In this example, the enhancer trap integrated site *eef1a1* promoter is driving ubiquitous GFP expression. The *Xenopus* line was bred from the founder female 16F, from the Mead lab.

4.2 Gene Traps/Gene Trap Loci

Gene trap experiments can produce cell lines in which integration into a putative gene is selected by virtue of its expression in a cell line. The trapped gene is usually (though not necessarily) mutated by the integration. The site of integration can be characterized by a number of means, including cloning or extension of cDNA products. The loci of integration of a series of gene trap lines, once characterized as potentially unique, can be named and symbolized as members of a series.

Basic nomenclature format for gene traps: $Gt(\textit{vector designation})^{\#ILAR Lab code}$

If the trapped gene is novel/unknown, vector names can be used.

Example 36: $Gt(ST629)^{1Walfd}$

For example, the gene trapped by the ST629 vector is unknown, from the Wallingford lab (Walfd).

If the trapped gene is known, then the gene trap designation becomes an allele of the gene into which it was inserted, once that gene is identified.

Basic nomenclature for gene trapped alleles: $gene\ symbol^{Gt(\textit{vector content})\#ILAR Lab code}$

Example 37: $ntn1^{G(ST629)Walfd}$

In this hypothetical example studying the $Gt(ST629)^{1Walfd}$, if it is found to disrupt the netrin1 (*ntn1*) gene, after which the full allele designation and official name for this gene trap mutation changes to that of an allele of *ntn1* in subsequent publications.

Gene traps with mutations are symbolized in a similar way to targeted mutations.

Example 38: $Gt(ROSA26)Sor^{tm1(sb11)Ryffl}$

The symbol consists of:

- The gene trap into which the transposase was integrated, in this case $Gt(ROSA26)Sor$
- In the superscript:

- *tm* denoting targeted mutation
- A serial number of the targeted mutation
- In parentheses, a lowercase transposase symbol, in this case sb11 for the Sleeping Beauty 11 transposase
- The ILAR Lab Code of the originating lab - in this case Ryffel's lab in Germany.

5.0. Transposon-Induced Mutations & Inserts -Overview

Standard nomenclature for transposable-element inserts, transposase transgenes, and resulting transposed insertion alleles, based on mouse and zebrafish standards, are given below.

5.1 Transgenic Transposable Element Concatamers (use of /)

The transgenic transposable element concatamers are identified with a standard prefix Tg (for transgenic) and Tn (for transposon). The class of transposable element may be included in parentheses. The general format of the symbol is:

TgTn(transposon class:vector1/vector2)#ILAR Lab code

Example 39: *TgTn(sb-T2/GT2/tTA)1Nxr*

The name consists of:

- Tg denoting transgenic
- Tn denoting transposon
- In parentheses, a lowercase abbreviation of the transposon class (in this case sb for Sleeping Beauty), followed by a hyphen and then the vector designations, each separated by forward slash (/).
- The laboratory's line or founder designation or a serial number
- The ILAR Lab code of the originating lab

5.2 Transposase inserts used to generate *Xenopus* lines

Transposases can be engineered into the genome via transgenesis or specific gene targeting. In these cases either the relevant nomenclature for transgenes or targeted mutations is used, respectively.

For a transgene, use the standard prefix Tg (for transgene), followed in parentheses by the enhancer/promoter and the symbol for the transposase, separated by a hyphen. Note that this distinguishes the construct from a regular Tg construct with the enhancer/promoter and ORF/gene separated by a colon.

The basic format of the symbol is:

Xl.Tg(**insertion locus**/promoter-transposase)#ILAR Lab code

The symbol consists of:

- *Xenopus* species symbol (Xl or Xt)
- Tg denoting transgene
- In parentheses, the official gene symbol for the locus/promoter, followed by a hyphen and a lowercase transposase symbol

In the superscript:

- The laboratory's line or founder designation or a serial number
- The #ILAR Lab code

Example 40: Xt.Tg(acta1-sb10)^{7F^{Mead}}

In this case, the Sleeping Beauty 10 transposase, abbreviated here as sb10, was engineered into the acta1 gene, founder female #7, as 7F, by the Mead lab.

Example 41. Xt.Tg(insertion locus-tol2, eef1a1:eGFP)^{10M^{Mead/+}}

In this example, the Mead lab generated heterozygous Tol2 transposon animals derived from the “10M” female founder, that harbors a single Tol2 insertion mapping to scaffold 8 of the *X. tropicalis* JGI Genome v. 7.1. This transposon also contains an EF1 α enhancer driving ubiquitous expression of eGFP in the embryo.

Note that although this line was subsequently supplied to the Khokha lab, but it will retain its ILAR lab code describing its lab of origin in Khokha lab papers.

Example 42: Xt.Tg(kcl1-tol2, eef1a1:eGFP)^{U2644H6^{Khokha}}

In this example from the Khokha lab, the integration site in high expression offspring from double transgenic founders, *X. tropicalis* line was demonstrated as the gene kcl1. the superscript number records the founder number (U2644) and the H6 represents the offspring serial number.

Example 43: Xt.Tg(Dre.zp3-tol2,cryga:eGFP)^{Khokha}

In this example from the Khoka lab, zebrafish (*Dre*) zona pellucida glycoprotein 3 (*zp3*) promoter drives egg specific expression of Tol2 transposase. Gamma crystallin promoter (*cryga*) drives expression of eGFP in lens of the eye as a reporter for transgene insertion.

For a targeted knock-in of the transposase, use the standard format for a targeted mutation, *i.e.*, the symbol of the targeted gene with a superscripted allele symbol beginning with the prefix *tm*, and a serial number. The contents of the parentheses will usually be the symbol for the transposase with which it is associated. The general format of the symbol is:

gene^{tm#(transposase)#ILAR Lab code}

Example 44: *tmem260*^{tm2(Tol2)H1Khokha}

In this hypothetical example, the Khoka lab used a target mutation technique to get a

5.3 Transposed Insertion Alleles

Transposed insertion alleles follow the rules for naming all other alleles.

In general a transposable element concatamer marker will already be established, as above.

The new allele, then, will be a superscripted form of the concatamer symbol.

Note that all such alleles that are "derived from" a transposable element concatamer carry the original number with a decimal point and serial number identifying the specific allele.

The basic format is: gene^{Tn(transposon_class_abbreviation-vector)#ILAR Lab code}

Example 44: *foxd3*^{Tn(sb-T2/GT2/tTA)1.1Vlemx}

In this case, the symbol consists of:

- The gene into which the transposable element was integrated/transposed (eg., *foxd3*)
- In the superscript:
 - *Tn* denoting transposon
 - In parentheses, a lowercase abbreviation of the transposon class, followed by a hyphen and the vector designation (in example 39, 'sb' for Sleeping Beauty , the transposon class and T2 vector respectively).
 - A serial number, in which the primary number corresponds to that given to the transposable element concatamer from which it arose, followed by a

decimal point and a serial number designating its number within the series of derivative insertion alleles.

- The ILAR Lab code of the lab originating the transposable element line.

5.4 If a newly transposed insertion occurs in an unknown site or intergenic region, the form will be:

Tn(transposon class abbreviation-vector)# ILAR Lab code

This format is used to symbolize the "genomic mutation" **without being superscripted** to a gene symbol, similar to the way a random transgene inserted into a non-gene site is designated.

Example 43: Tn(sb-T2/GT2/tTA)1.1Vlemx

6.0 Wild-type *Xenopus* Strains

The term 'lines' is reserved for transgenic lines or mutant lines (ie., animals carrying specific transgenes or mutant alleles).

Wild-type (wt) *Xenopus* should only be referred to as 'strains'. As it is assumed wild type *Xenopus* carry genetic variation characteristic of the geographic location from which captive laboratory colonies were established, the wt strain used to establish Tg lines, and those used in control/other experiments, should be reported.

The NXR, EXRC, and the NBRP stock centers provide up-to-date information about available wt strains of both *Xenopus* species, and can provide details of the geographic origin, pigmentation (albino, pigmented, colour morph), breeding (inbred/outbred), history/generation number (F5 etc) and source of each wt strain.

7.0. ILAR Lab Codes for *Xenopus* lines.

The source or lab of origin of *Xenopus* transgenic and mutant lines needs to be made obvious to the research community. The Institute for Laboratory Animal research maintains an online registry for unique lab codes. www.dels.nas.edu.ilar

A list of *Xenopus* labs generating and/or providing transgenic lines to the stock centers, and their proposed ILAR lab codes are listed in **Appendix 2**. As there are a manageable number of *Xenopus* labs worldwide that are generating transgenics, we suggest longer ILAR Codes, which can be the entire surname (eg., Amaya) or an abbreviated version of a

longer surname, (eg., Kellr, Buchz, Mazab). This will make it (almost) immediately obvious as to the lab source of each Tg line.

Proposal:

- *Xenopus* labs supplying or maintaining Tg lines will be registered at the ILAR.
- Lab codes should not conflict with gene symbols and other common abbreviations.
- *Xenopus* labs will be named after the surname or initials of the laboratory head/principal investigator, usually with 3-6 letters, and no numerals.
- Each ILAR lab code needs to be unique across all model organism labs- not just *Xenopus* research labs
- Lab code format: first letter uppercase, followed by all lowercase letters. All lab codes, when registered, are automatically written with only the first letter capitalised- so EXRC becomes Exrc on the ILAR site. Company names and Stock Centres can however be referred to in all capitals in
- Lab code will be in superscript for mutant alleles, and follow Tg line names as the source.
- Lab code starting with **the letter 'D' should be avoided**, because dominant and semi-dominant alleles have a 'd' in the first position of the line designation to distinguish them from recessive alleles, and these designation become superscripts, along with the lab code for mutant alleles.
- Where more than one Tg line or mutant has been generated, the Lab code will be prefixed by a sequential numeral (eg., 1NXR and 2NXR).
- Institutes and organizations (eg., NXR, NASCO, EXRC) can be registered. Institutional names are often used where patents or material transfer agreements (MTAs) are in effect.

8.0 Future Plans: Using Research Resource Identifiers (RRIDs)

The Resource Identification Initiative (#RII) (Group Leader: Maryann Martone, UC San Diego, USA) aims to help researchers sufficiently cite the key resources used to produce the scientific findings reported in the biomedical literature. Transgenic constructs, antibodies and morpholinos, as well as model organism Tg lines and wild type strains will all be allocated RRIDs in the near future. **These numbers are machine readable, and promote the reproducibility of data.** Xenbase and the stock centers will attach these numbers to our records for these resources in the future. Authors are strongly encouraged to cite RRID numbers, when available, in the methods of their publications.

Please see the main Resource Identification Initiative website for more information.
https://www.force11.org/Resource_identification_initiative

Appendix 1: Prefixes for promoters/gene of origin from non-*Xenopus* species.

● <i>Aequorea victoria</i>	Avi	(crystal jelly)
● <i>Anopheles gambiae</i>	Aga	(mosquito)
● <i>Arabidopsis thaliana</i>	Ath	(cress)
● <i>Caenorhabditis elegans</i>	Cel	(worm)
● <i>Chlamydomonas reinhardtii</i> *	Cr	(green algae)
● <i>Danio rerio</i>	Dre	(zebrafish)
● <i>Drosophila melanogaster</i>	Dme	(fly)
● <i>Escherichia coli</i>	Eco	(bacteria)
● <i>Gallus gallus</i>	Gga	(chicken)
● <i>Homo sapien</i>	Hsa	(human)
● <i>Mus musculus</i>	Mmu	(mouse)
● <i>Oryctolagus cuniculus</i>	Ocu	(rabbit)
● <i>Oryza sativa</i>	Osa	(rice)
● <i>Oryzias latipes</i>	Ola	(Medaka fish)
● <i>Rattus norvegicus</i>	Rno	(rat)
● <i>Salmo salar</i>	Ssa	(salmon)

(* exception to 3-letter rule, to avoid confusion with CRE)

In other model/research organisms where *Xenopus* genes, or when sequence from the alternate *Xenopus* species are utilized in transgenic constructs, the 3 letter abbreviations indicating *Xenopus* species of origin are:

- | | |
|-----------------------------|-----|
| ● <i>Xenopus laevis</i> | Xla |
| ● <i>Xenopus tropicalis</i> | Xtr |

Appendix 2: ILAR Lab codes for the source of transgenic and mutant *Xenopus* lines.

- Amaya: Enrique Amaya, Manchester, UK (registered 10/8/14)
- Brown: Don Brown, Carnegie Institute, USA (registered 10/8/14)
- Buchz: Dan Buchholz, Univ. Cincinnati (registered 10/8/14)
- Exrc: European Xenopus Resource Center, Portsmouth, UK (registered 10/9/14)

- Gurdn: John Gurdon (registered 10/9/14)
- Green: Jeremy Green, London, (registered 10/9/14)
- Grngr: Robert Grainger (registered 10/9/14)
- Hank: James Hanken, Harvard (registered 9/3/14)
- Horb: Marko Horb, NXR (registered 10/9/14)
- Jmws: Jonathan M. W. Slack (registered 10/20/14)
- Jpsj: Jean-Pierre Saint-Jeannet, New York University. (registered 10/17/14)
- Kellr: Ray Keller, Univ. Virginia (registered 10/9/14)
- Khok: Mustafa Khokha, Yale School of Medicine (registered 10/9/14)
- Knox: Barry Knox, SUNY (registered 10/9/14)
- Krieg: Paul A Krieg, UA (registered 10/9/14)
- Mayor: Roberto Mayor, University College London (registered 10/9/14)
- Mazab: Andre Mazabraud, Orsey, France (registered 10/9/14)
- Mead: Paul Mead, St Jude Research Hospital. (registered 10/17/14)
- Mohun: Tim Mohun, NIMR, UK (registered 10/9/14)
- Nasco: NASCO Inc., Fort Atkinson, WI , USA (registered 10/17/14)
- Nxr: National Xenopus Resource, MBL, USA (registered 10/17/14)
- Papal: Nancy Papalopulu, Manchester UK ((registered 10/17/14)
- Patnt: Roger Patient, Oxford (registered 10/20/14)
- Pec: Peter Carmeliet (already registered at ILAR- tho institution has changed)
- Rmh: Richard Harland, UC Berkeley (registered at ILAR 2002-07-15)
- Ryffl: Gerhart Ryffel (registered 10/17/14)
- Senk: Sabine Senkel (ex Ryffel lab) (registered 10/17/14)
- Tamur: Koji Tamura, Sendai, Japan (registered 10/17/14)
- Ueno: Naoto Ueno, Okazaki, Japan (registered 10/17/14)
- Walfd: John Wallingford, Univ. Texas at Austin. (registered 10/17/14)
- Vlemx: Kris Vleminckx (Vlem in use at ILAR) (registered 10/17/14)
- Zorn: Aaron Zorn, CCHMC (registered at ILAR)
- Zimml: Lyle Zimmerman, UK (registered 10/17/14)
- Zuber: Michael Zuber, Upstate Med. University, NY. (registered 9/2/14)

Please address all comments or questions to the transgenic nomenclature administrator at Xenbase: xenbase@ucalgary.ca

References:

The zebrafish transgenic nomenclature rules can be seen here:

<https://wiki.zfin.org/display/general/ZFIN+Zebrafish+Nomenclature+Guidelines>

The mouse transgenic nomenclature rules can be seen here: