

# Publishing Interactive Articles: Integrating Journals and Biological Databases

**GSA** A Publication of The Genetics Society of America  
**GENETICS** Search    
Advanced Search  
Home Journal Information Subscriptions & Services Collections Previous Issues Current Issue Future Issues  
Institution: CALIFORNIA INSTITUTE OF TECHNOLOGY | Sign In via User Name/Password

Originally published as Genetics Published Articles Ahead of Print on June 30, 2010.  
Genetics, Vol. 186, 135-145, September 2010, Copyright © 2010  
doi:10.1534/genetics.110.117341

### A Ubiquitin E2 Variant Protein Acts in Axon Termination and Synaptogenesis in *Caenorhabditis elegans*

**Gloriana Trujillo<sup>\*,1</sup>, Katsunori Nakata<sup>1,2,†</sup>, Dong Yan<sup>\*,1</sup>, Ichi N. Maruyama<sup>†</sup> and Yishi Jin<sup>\*,1,3</sup>**

<sup>\*</sup>Neurobiology Section, Division of Biological Sciences, University of California, San Diego, CA 92093 <sup>†</sup>Information Processing Biology Unit, Okinaka Institute of Science and Technology, Okna-Son, Okinaka 304-0412, Japan and <sup>‡</sup>Howard Hughes Medical Institute, La Jolla, CA 92093  
<sup>3</sup>Corresponding author: 9500 Gilman Dr., MC 0366, University of California, San Diego, CA 92093-0366  
E-mail: yjin@ucsd.edu

Manuscript received April 8, 2010. Accepted for publication June 19, 2010.

**THIS ARTICLE**  
Abstract **FREE**  
Full Text (PDF)  
Supporting Information  
All Versions of this Article:  
genetics.110.117341v1  
186/1/135 most recent  
Alert me when this article is cited  
Alert me if a correction is posted

**SERVICES**  
Email this article to a friend  
Similar articles in this journal  
Similar articles in PubMed

Home Genome Synteny Blast / Biat WormMart Markers Genetic Maps Submit Searches  
**WormBase Release WS218** **WormBase**

Find: Any Gene  Search  
 Exact match  Results as XML  Literature Search  Wormbase Suggest

**Web Site Directory**  
About this release [New/Changed Genes](#), [release notes](#)  
General Searches [WormBase Class Browser](#), [Wormbase Query Language Search](#), [AQL Search](#)  
Sequences [C. elegans Genome](#), [C. briggsae Genome](#), [Gene](#), [Blast / Biat](#), [e-PCR](#), [Gene Ontology](#), [Synteny Viewer](#), [Co-Elements \(CisOrtho\)](#)  
Cells and Gene Expression [Cell and Pedigree](#), [Neurons](#), [Expression Pattern](#), [Expression profile](#)  
Genetics, Strains, and Phenotypes [Genetic Interval](#), [Rearrangements](#), [Balancers](#), [Clone](#), [A/sets](#), [SNPs](#), [Markers](#), and [Strains](#), [Strain Report](#), [Phenotypes](#), [RNAi](#)  
Batch Queries [WormMart](#) [about...], [Batch Genes](#), [Batch Sequences](#)  
Downloads and Data Mining [Bulk Downloads](#), [Linking to WormBase and Data Mining...](#)  
Community [Worm Meetings](#), [Worm Community Discussion Forum](#), [Wormbase Wiki](#), [Mailing Lists](#)

**News and Notes**  
- **02 Oct 2010: Unscheduled service outage: all services restored**  
The colocation facility hosting WormBase suffered a catastrophic failure of a critical component today around 8AM ET. System administrators worked throughout the day to restore service. We are happy to report that all services are back online and apologize for any inconvenience.  
- **28 Sep 2010: New release of WormBase: WS218**  
WormBase has been updated to the WS218 release of the database. We've added a new Genome Browser for *Haemaphysalis contortus*, bringing the number of species housed at WormBase to 10. Detailed release notes are available on the WormBase Wiki.  
- **30 Aug 2010: Possible service interruptions, 31 August 2010**  
We're relocating some services to a new hosting facility beginning at 10AM ET (GMT -5), Tuesday August 31, 2010. We plan to maintain systems at the old facility during this transition, but because these upgrades require modifications to the global domain name system records, you may encounter intermittent service interruptions or "server not found" errors. We apologize in advance for any difficulties.

Karen Yook

Genetics Society of America, Dartmouth Journal Services,  
Textpresso and WormBase



# Goal: hyperlink entities in a published paper to their database resource page

The image shows a screenshot of a journal article page from the journal *Genetics*. The page header includes the GSA logo and the journal title "GENETICS". A search bar is present, and navigation links for "Home", "Journal Information", "Subscriptions & Services", "Collections", "Previous Issues", "Current Issue", and "Future Issues" are visible. The article title is "A Ubiquitin E2 Variant Protein Acts in Axon Termination and Synaptogenesis in *Caenorhabditis elegans*". The authors listed are Gloriana Trujillo, Katsunori Nakata, Dong Yan, Ichi N. Maruyama, and Yishi Jin. A sidebar on the right titled "THIS ARTICLE" contains links for "Abstract", "Full Text (PDF)", "Supporting Information", and "All Versions of this Article".

**RESULTS**

**UEV-3 is a Ubc/E2 variant protein:**  
Previous analyses of the suppressors of *rpm-1* loss-of-function (*lf*) mutants revealed five loci, defining *dlk-1*, *mkk-4*, *pmk-3*, *mak-2*, and *catp-1* (NAKATA et al. 2005; YAN et al. 2005). We devised a noncomplementation test scheme and identified four alleles belonging to a new complementation group: *ju593*, *ju587*, *ju638*, and *ju639*. We mapped this suppressor locus to an interval of ~90 kb on the right arm of chromosome I (Figure 1A). We used a combination of RNAi and transgenic expression of cosmid DNAs from the region to locate the gene affected. We found that the 6-kb *Eco105I*-*SpeI* fragment of the cosmid F26H9 contained the rescuing activity for the suppression of *rpm-1(ju44)* by *ju587*. The genomic DNA fragment contains two predicted genes: *rab-5* and *uev-3*. By RT-PCR and 5'-RACE analyses, we determined that *uev-3* transcripts contained an SL2 splice leader, confirming that *rab-5* and *uev-3* form an operon, with *uev-3* as the downstream gene (MATERIALS AND METHODS). DNA sequence analysis from *ju587*, *ju593*, and *ju638* identified single nucleotide alteration at various splice acceptor sites, while *ju639* is a 26-bp deletion from amino acid 277 in the sixth exon (Figure 1B and Table S1). Moreover, we performed RNAi of *uev-3* in a sensitized genetic background and observed partial suppression of *rpm-1(lf)* (Table S1). These analyses are consistent with the suppressor mutations causing loss-of-function in *uev-3*.

**FIGURE 4**

into the nerve ring and forms synapses (Figure 2B). The PLM cell body resides in the tail and sends a projection anterior into the midbody of the animal, terminating posterior to the ALM cell body. PLM cells also extend a synaptic branch to the ventral cord to form synapses onto its partners. In *rpm-1(lf)* mutants, both ALM and PLM axons frequently overextend beyond their normal termination sites and loop posteriorly or into the ventral cord, described as "ALM hook" or "PLM hook" defects, respectively (Figure 2B). Additionally, the PLM synaptic branch is frequently missing. Although low levels of ALM and PLM defects are detected in *uev-3(ju587)* and *uev-3(ju638)* mutants, both mutations significantly suppressed *rpm-1(lf)* (Figure 2B). The degree of suppression of *rpm-1(lf)* by both alleles of *uev-3* is comparable to those observed for the mutants in the DLK-1 MAPK cascade (GRILL et al. 2007). *uev-3(ju639)* has a slightly stronger suppression effect on the mechanosensory neuron phenotypes, so we have designated *ju639* as the canonical mutation of the gene.

# Goal: hyperlink entities in a published paper to their database resource page

GSA A Publication of The Genetics Society of America

# GENETICS

Search  GO

Advanced Search

Home Journal Information Subscriptions & Services Collections Previous Issues Current Issue Future Issues

Institution: CALIFORNIA INSTITUTE OF TECHNOLOGY | Sign In via User Name/Password

Originally published as Genetics Published Articles Ahead of Print on June 30, 2010.

Genetics, Vol. 186, 135-145, September 2010. Copyright © 2010  
doi:10.1534/genetics.110.117341

## A Ubiquitin E2 Variant Protein Acts in Axon Termination and Synaptogenesis in *Caenorhabditis elegans*

Gloriana Trujillo<sup>\*,1</sup>, Katsunori Nakata<sup>1,2,†</sup>, Dong Yan<sup>\*,1</sup>, Ichi N. Maruyama<sup>†</sup> and Yishi Jin<sup>\*,1,3</sup>

THIS ARTICLE

Abstract **FREE**

Full Text (PDF)

Supporting Information

All Versions of this Article:  
genetics.110.117341v1  
186/1/135 most recent

Alert me when this article is cited

### RESULTS

Transgenic expression of cosmid DNAs from the region containing the 6-kb *Eco105I-SpeI* fragment of the cosmid F2620 suppressed the *rpm-1(ju44)* phenotype. The genomic DNA region containing the *rpm-1(ju44)* mutation was determined that *uev-3* transcripts contained an SL2 (MATERIALS AND METHODS). DNA sequence analysis of the *uev-3* gene revealed that *ju587* is a 20-bp deletion from amino acid 277 in the sixth exon (Figure 1B and Table S1). Moreover, we performed RNAi of *uev-3* in a sensitized genetic background and observed partial suppression of *rpm-1(ju)* (Table S1). These analyses are consistent with the suppressor mutations causing loss of function of *uev-3*.

ues, laterally in the midbody region and sends a process into the nerve ring and forms synapses (Figure 2). In *rpm-1(ju)* mutants, both ALM and PLM axons frequently overextend posteriorly into the nerve ring and form "ALM hook" or "PLM hook" defects, respectively (Figure 2). In an animal where a process branches laterally into the midbody of the animal, the ALM and PLM axons frequently overextend into the ventral cord, described as "ALM hook" or "PLM hook" defects, respectively (Figure 2). Although low levels of ALM and PLM defects are detected in *uev-3(ju587)* and *uev-3(ju638)* mutants, both mutations significantly suppressed *rpm-1(ju)* (Figure 2B). The degree of suppression of *rpm-1(ju)* by both alleles of *uev-3* is comparable to those observed for the mutants in the DLK-1 MAPK cascade (GRILL et al. 2007). *uev-3(ju638)* has a slightly stronger suppression effect on the mechanosensory neuron phenotypes, so we have designated *ju638* as the canonical mutation of the gene.

# Goal: hyperlink entities in a published paper to their database resource page

GSA A Publication of The Genetics Society of America

# GENETICS

Search  GO

Advanced Search

Home Journal Information Subscriptions & Services Collections Previous Issues Current Issue Future Issues

Institution: CALIFORNIA INSTITUTE OF TECHNOLOGY | Sign In via User Name/Password

Originally published as Genetics Published Articles Ahead of Print on June 30, 2010.

Genetics, Vol. 186, 135-145, September 2010. Copyright © 2010  
doi:10.1534/genetics.110.117341

## A Ubiquitin E2 Variant Protein Acts in Axon Termination and Synaptogenesis in *Caenorhabditis elegans*

Gloriana Trujillo<sup>\*,1</sup>, Katsunori Nakata<sup>\*,2,†</sup>, Dong Yan<sup>\*,†</sup>, Ichi N. Maruyama<sup>†</sup> and Yishi Jin<sup>\*,1,3</sup>

THIS ARTICLE

Abstract **FREE**

Full Text (PDF)

Supporting Information

All Versions of this Article:  
genetics.110.117341v1  
186/1/135 most recent

Alert me when this article is cited

**RESULTS**

Transgenic expression of cosmid DNAs from the region containing the 6-kb *Eco105I-SpeI* fragment of the cosmid F2620 suppressed the *rpm-1(ju44)* phenotype by *ju587*. The genomic DNA from *ju587* mutants determined that *uev-3* transcripts contained an SL2 domain (MATERIALS AND METHODS). DNA sequence analysis of *uev-3* revealed that *ju587* is a 20-bp deletion from amino acid 277 in the sixth exon (Figure 1B and Table S1). Moreover, we performed RNAi of *uev-3* in a sensitized genetic background and observed partial suppression of *rpm-1(ju)* (Table S1). These analyses are consistent with the suppressor mutations causing loss of function of *uev-3*.

ues, laterally in the midbody region and sends a process into the nerve ring and forms synapses (Figure 2). PLM axons frequently overlie the ALM cell body. PLM axons frequently overlie the ALM hook" or "PLM hook" defects, respectively.

an animal where a process branches anterior into the midbody of the animal, PLM axons frequently overlie the ALM hook" or "PLM hook" defects, respectively.

# Goal: hyperlink entities in a published paper to their database resource page

GENETICS  
A Publication of The Genetics Society of America

Home Journal Information Subscriptions & Services Collections Previous Issues Current Issue Future Issues

Institution: CALIFORNIA INSTITUTE OF TECHNOLOGY | Sign In via User Name/Password

Originally published as Genetics Published Articles Ahead of Print on June 30, 2010.  
Genetics, Vol. 186, 135-145, September 2010. Copyright © 2010  
doi:10.1534/genetics.110.117341

**A Ubiquitin E2 Variant Protein Acts in Axon Termination and Synaptogenesis in *Caenorhabditis elegans***  
Gloriana Trujillo<sup>\*,1</sup>, Katsunori Nakata<sup>\*,2,†</sup>, Dong Yan<sup>\*,1</sup>, Ichi N. Maruyama<sup>†</sup> and Yishi Jin<sup>\*,1,3</sup>

**RESULTS**

Transgenic expression of cosmid DNAs from the region containing the 6-kb *Eco*105I–*Spe*I fragment of the cosmid F2622, which encodes the *rpm-1(ju44)* allele, suppressed the *rpm-1(ju587)* phenotype. The genomic DNA sequence of the *uev-3* transcripts contained an SL2 motif (MATERIALS AND METHODS). DNA sequence analysis of the *uev-3* gene revealed a 20-bp deletion from amino acid 277 in the sixth exon (Figure 1B and Table S1). Moreover, we performed RNAi of *uev-3* in a sensitized genetic background and observed partial suppression of *rpm-1(ju)* (Table S1). These analyses are consistent with the suppressor mutations causing loss of function of *uev-3* in the midbody region and series of ALM neurons that terminate posteriorly into the nerve ring and forms synapses (Figure 2). In *ju44* mutants, both ALM and PLM axons frequently overexpressed ALM hook or PLM hook defects, respectively (Figure 2). In *ju44* mutants, both ALM and PLM axons frequently overexpressed ALM hook or PLM hook defects, respectively (Figure 2). In *ju44* mutants, both ALM and PLM axons frequently overexpressed ALM hook or PLM hook defects, respectively (Figure 2).

Home Genome Synteny Blast / Blat WormMart Markers Genetic Maps Submit Searches Site Map

Find: ju44

**WormBase**

Variation Summary Gene Summary Locus Summary Genome Browser Tree Display XVL Schema AceDb Image

**Variation report for: ju44**

Specify an allele such as e307, n2559, e345, or pkP6112

[general info][molecular details][location][genetic information][phenotype][history][references][submit new data]

**General Information**  
Variation: ju44  
Variation type: allele

**Molecular Details**  
Nucleotide change: substitution: g / a (wild type / mutant)  
Flanking sequences:  
gttcttcagatcaaacagcggagttcaagttc  
cccaagacacggagtgatcatctcggagta  
Context:  
...cttccagatattgactccatgtctggg C agactgaactccgcttctatctgaacaa... -- Wild type  
...cttccagatattgactccatgtctggg T agactgaactccgcttctatctgaacaa... -- ju44  
Note: Sequence is reported on the plus strand.  
[View expanded context]

Affects CDS: C01B7.6

**Gene Summary for uev-3**

Specify a gene using a gene name (*unc-28*), a predicted gene id (R13A5.5), or a protein ID (CE02711)

[identification][location][function][expression][gene ontology][genetics][homology][reagents][bibliography]

**Identification**  
IDs: Main name Sequence name WB Gene ID  
uev-3 - (Ubiquitin E2 (conjugating enzyme) variant) via Person evidence: Peter Candido F26H9.7 WBGene00006732

**Concise Description:** uev-3 encodes a ubiquitin-conjugating enzyme (UBC or E2) variant that contains the UBC motif, but lacks the critical active-site cysteine residue necessary for catalytic activity; as loss of UEV-3 function via RNA-mediated interference (RNAi) does not result in any abnormalities, the precise role of UEV-3 in *C. elegans* development and/or behavior is not yet known; based on similarity to *Saccharomyces cerevisiae* and human proteins, however, UEV-3 may play a role in cell cycle control or response to stress or DNA damage. [details]

**Species:** *Caenorhabditis elegans*

**Gene**  
Gene Model Status Nucleotides (coding/transcript) Protein Amino Acids

**Summary for Anatomy Ontology Term: ALM (WBbt:0005406)**

The Anatomy Ontology is a controlled vocabulary for cellular anatomy. Type in some text, such as GABAergic neuron, a cell name, such as M1, or an AO accession number, such as WBbt:0005451 to run a specialized search of the Anatomy Ontology. Browse the ontology from the top down by searching for Anatomy Ontology.

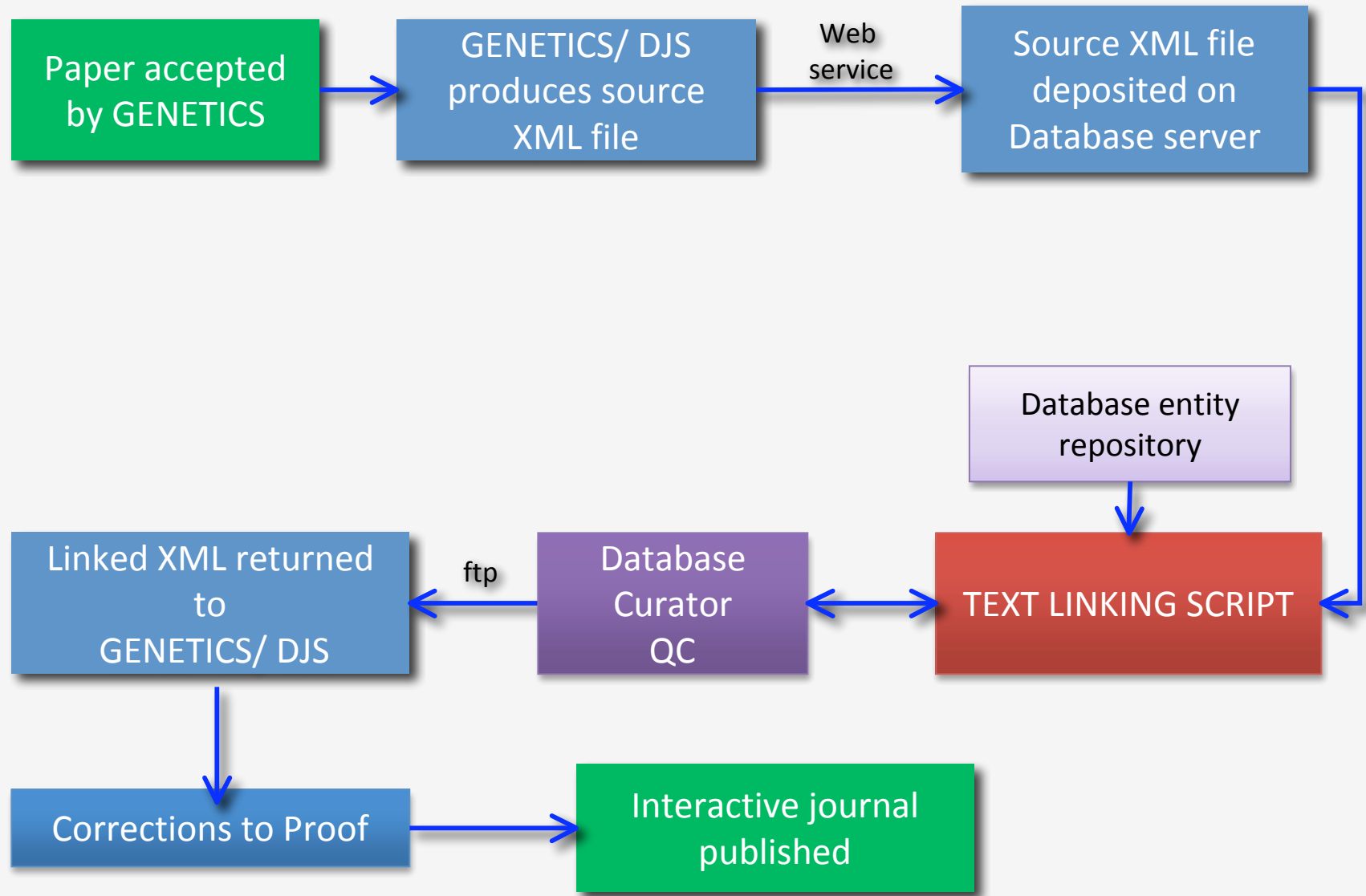
Search text or Anatomy accession number: ALM

[Search All Ontologies]

**Identification**  
Term: ALM (WBbt:0005406)  
Definition: neuron type, a pair of anterior sensory neurons that transduce touch stimuli. (via paper evidence)  
External Resources: WormAtlas Neuralnet

**Ontology Browser**  
There is(are) 1 path(s) for WBbt:0005406 (ALM)  
Go to path:  
1  
path number: 1

# GSA-Database markup pipeline



# What is needed for this to work

- Textpresso -> Arun
- Source file from the journal (XML format)
- List of entities that exist in the database
- Stable database URL constructor
- Database curator for manual quality control (QC)
- Method for ensuring URL stability

# What is needed:

- List of entities that exist in the database, these are mined directly from the database (WormBase)

Entity Class	Examples	# of WormBase entities
Clone	A1D10, C02H12, GAP1_10	204,605
Gene	mec-15, F25B3.5, y110a7a.10	537,027
Variation (alleles)	ad820sd, cb19860	274,649
Transgene	adEx1290, cmls6, jeln2	6,546
Rearrangement	hDp23, stDf47, sT1	709
Protein (to gene pages)	MEC-15, ABF-1	
Strain	BC954, IE35413	28,826
Anatomy	AWC, body wall muscle	6,563
Phenotype	Muv, Lon, Dpy	122
	<b>Total</b>	<b>1,059,047</b>



# What is needed:

- Stable database URL constructor:

For WormBase-

<http://www.wormbase.org/db/get?name=NAME;class=CLASS>

NAME is the entity string (lin-11, e189, gonad)

CLASS is the class the entity belongs to (gene, variation, anatomy)

For Saccharomyces Genome Database (SGD)-

<http://www.yeastgenome.org/cgi-bin/locus.fpl?dbid=SGDID>

SGDID is the SGD identification for the entity

# What is needed:

- Database curator for manual quality control:

- Checks all links made by the linking script

- Checks for ambiguities

- adds them to an exclusion list

- provides the right entity NAME/ID

- Checks for entities that did not get linked

- author writing shortcuts

- science jargon

- new entities not in the database yet

# Manual QC resolves ambiguities that an automated script cannot resolve

promoted and maintained independently of the globe period. THE *Caenorhabditis elegans* gonad derives of four cells that coalesces during embryogenesis and gonad precursors (SGPs) **Z1** and **Z4**, flanking two g **Z3** (Kimble and Hirsh 1979). The SGPs undergo various programs in each sex, involving sexually dimorphic

## Microscopy:

Images were captured using a Zeiss Imager **Z1** microscope MRM camera and processed using Axio Vision Release 4.

Home Genome Synteny Blast / Blat WormMart Markers Genetic Maps Submit Search

Find: Z1 [Any Gene]

WormBase

Anatomy Ontology Tree Display XML Schema

### Summary for Anatomy Ontology Term: Z1 (WBbt:0004577)

The Anatomy Ontology is a controlled vocabulary for cellular anatomy. Type in some text, such as GABAergic neuron, a cell name or an AO accession number, such as WBbt:0005451 to run a specialized search of the Anatomy Ontology. Browse the ontology down by searching for Anatomy Ontology.

Search text or Anatomy accession number: Z1 Search

[Search All Ontologies]

Identification	Term: Z1 (WBbt:0004577) [view: cell details   pedigree] Definition: Somatic gonad precursor cell (via paper evidence) Synonym(s): lineage name: MS.pppaap Remarks: Legacy information from Cell Z1 Remark: Somatic gonad precursor cell Transgene: <a href="#">syls186</a> <a href="#">syls187</a> <a href="#">syls188</a>
Ontology	There is (one) 3 path(s) for WBbt:0004577 (Z1).

# Manual QC is needed to decipher authors' writing shortcuts

Analysis of *zig* gene knockout alleles: Of the eight *zig* genes only one, *zig-1*, was previously analyzed through gene deletion analysis (AURELIO *et al.* 2002). Through generating deletion alleles by either transposon mobilization (*zig-1*) or deletion library screening, conducted by the *C. elegans* knockout consortia ([zig-2](#), [zig-3](#), [zig-5](#), [zig-6](#), [zig-7](#), and [zig-8](#)), we have obtained knockout alleles for each of the so far uncharacterized, seven remaining *zig* genes in *C. elegans*. Each allele is a predicted molecular null allele (Figure 2C) and none affects viability or produces any

All genes are hyperlinked to their respective web page

([zig-2](#), [zig-3](#), [zig-5](#), [zig-6](#), [zig-7](#),

# Manual QC maps informal terminology (jargon) to formal science terms

**SGPs** = somatic gonad precursor cells

daughters is much more pronounced in males t  
pathway cause a “symmetrical sisters” phenoty  
is imposed on the **SGPs** by the global sex deter  
**SGP** sex determination, with *tra-1* feminizing  
2004). **SGP** sex determination is linked to cell  
The later phase of gonadal development involv  
cells enlarge and leader cells (**distal tip cells** in  
in both sexes, leading to the formation of diffe  
(Kimble and Hirsch 1979). Although **SGP** divis

## Summary for Anatomy Ontology Term: somatic gonad precursor (WBbt:0007854)

The Anatomy Ontology is a controlled vocabulary for cellular anatomy. Type in some text, such as *GABAergic neuron*, a cell name, such as *N1*, an accession number, such as *WBbt:0005451* to run a **specialized search of the Anatomy Ontology**. Browse the ontology from the top down **Anatomy Ontology**.

Search text or Anatomy accession number:

[Search All Ontologies]

Identification	Term:	somatic gonad precursor (WBbt:0007854)
	Definition:	any of two cells that generate all somatic tissues of the gonad proper (i.e. ovary or testis ) and genital duct (i.e. vas deferens). (via paper evidence)
	Synonym(s):	SGP somatic gonadal precursor cell
	External Resources:	<a href="#">WormAtlas</a>
	Ontology:	

By adding acronyms and science-field jargon as synonyms to the canonical science term in the database, we capture the current language of our community while making the science more understandable to the wider science community.

# What is needed:

- Method for ensuring URL stability :

A list of entity classes, names, URLs, and webpage status is automatically generated after the final manual QC stage has been done.

This list provides an easy way to test the links for viability in the future.

**WB Paper ID : WBPaper00036368**

**Genetics DOI: 10.1534/genetics.110.115501**

Entity class	Entity name	link	live/silent
Anatomy	anterior gonad arms	<a href="http://www.wormbase.org/db/get?name=anterior_gonad_arm;class=Anatomy_name">http://www.wormbase.org/db/get?name=anterior_gonad_arm;class=Anatomy_name</a>	live
Anatomy	germ cell	<a href="http://www.wormbase.org/db/get?name=germ_cell;class=Anatomy_name">http://www.wormbase.org/db/get?name=germ_cell;class=Anatomy_name</a>	live
Anatomy	germ cells	<a href="http://www.wormbase.org/db/get?name=germ_cell;class=Anatomy_name">http://www.wormbase.org/db/get?name=germ_cell;class=Anatomy_name</a>	live
Anatomy	germ line	<a href="http://www.wormbase.org/db/get?name=germ_line;class=Anatomy_name">http://www.wormbase.org/db/get?name=germ_line;class=Anatomy_name</a>	live
Anatomy	germ lines	<a href="http://www.wormbase.org/db/get?name=germ_line;class=Anatomy_name">http://www.wormbase.org/db/get?name=germ_line;class=Anatomy_name</a>	live
Anatomy	spermatocyte	<a href="http://www.wormbase.org/db/get?name=spermatocyte;class=Anatomy_name">http://www.wormbase.org/db/get?name=spermatocyte;class=Anatomy_name</a>	live
Clone	zip-1	<a href="http://www.wormbase.org/db/get?name=zip-1;class=Clone">http://www.wormbase.org/db/get?name=zip-1;class=Clone</a>	live
Gene/Protein	-2	<a href="http://www.wormbase.org/db/get?name=SYP-2;class=Gene">http://www.wormbase.org/db/get?name=SYP-2;class=Gene</a>	live
Gene/Protein	-3	<a href="http://www.wormbase.org/db/get?name=SYP-3;class=Gene">http://www.wormbase.org/db/get?name=SYP-3;class=Gene</a>	live
Gene/Protein	-4	<a href="http://www.wormbase.org/db/get?name=SYP-4;class=Gene">http://www.wormbase.org/db/get?name=SYP-4;class=Gene</a>	silent
Gene/Protein	F26D2.2	<a href="http://www.wormbase.org/db/get?name=F26D2.2;class=Gene">http://www.wormbase.org/db/get?name=F26D2.2;class=Gene</a>	live
Gene/Protein	HIM-3	<a href="http://www.wormbase.org/db/get?name=HIM-3;class=Gene">http://www.wormbase.org/db/get?name=HIM-3;class=Gene</a>	live
Gene/Protein	HIM-8	<a href="http://www.wormbase.org/db/get?name=HIM-8;class=Gene">http://www.wormbase.org/db/get?name=HIM-8;class=Gene</a>	live
Gene/Protein	RAD-51	<a href="http://www.wormbase.org/db/get?name=RAD-51;class=Gene">http://www.wormbase.org/db/get?name=RAD-51;class=Gene</a>	live
Gene/Protein	SYP-1	<a href="http://www.wormbase.org/db/get?name=SYP-1;class=Gene">http://www.wormbase.org/db/get?name=SYP-1;class=Gene</a>	live

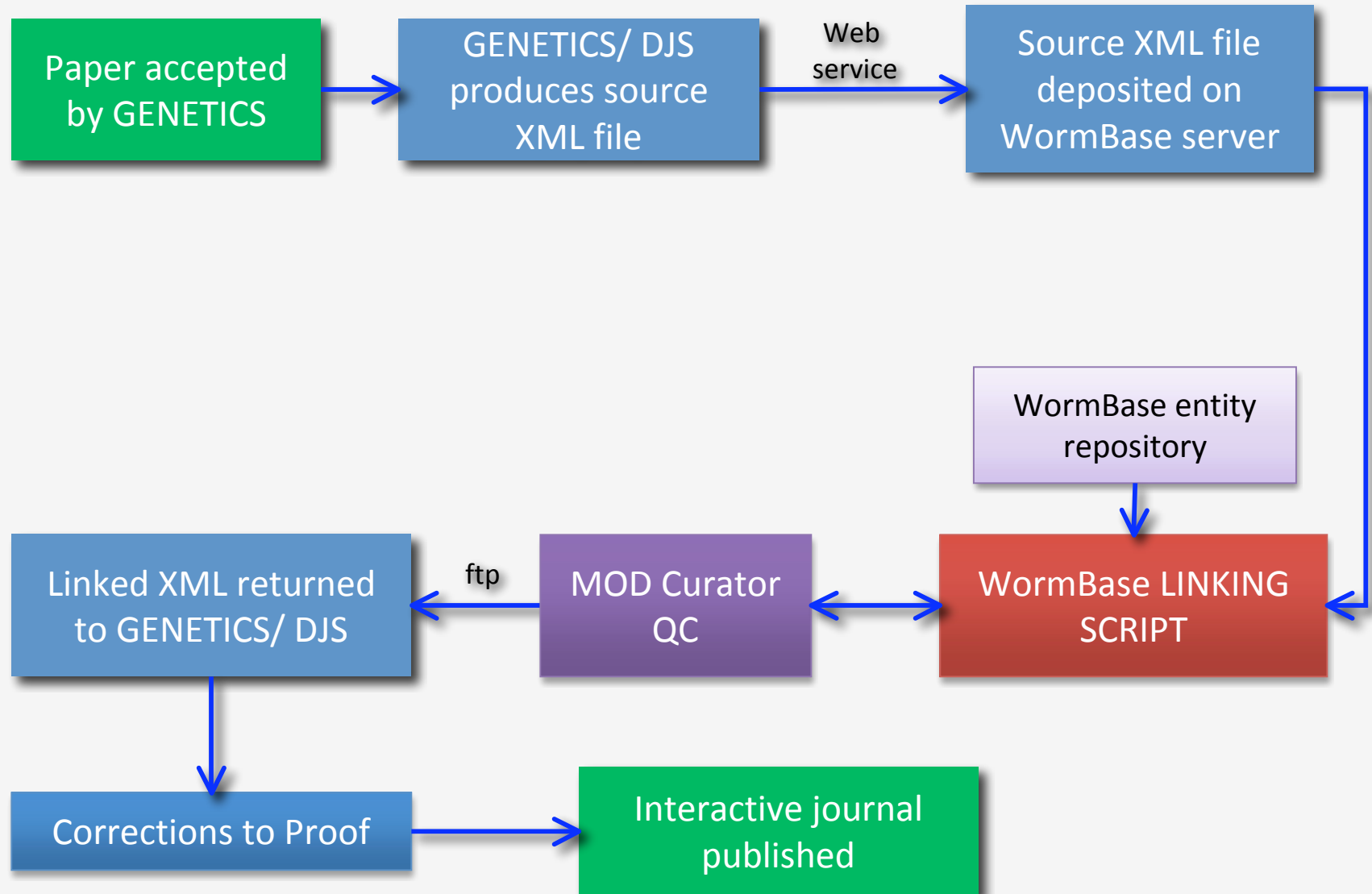
# WormBase creates silent links to entities not yet in the database

GENETICS helps WormBase identify new entities by sending the author an entity declaration form

Manual QC identifies new entities not declared by the author

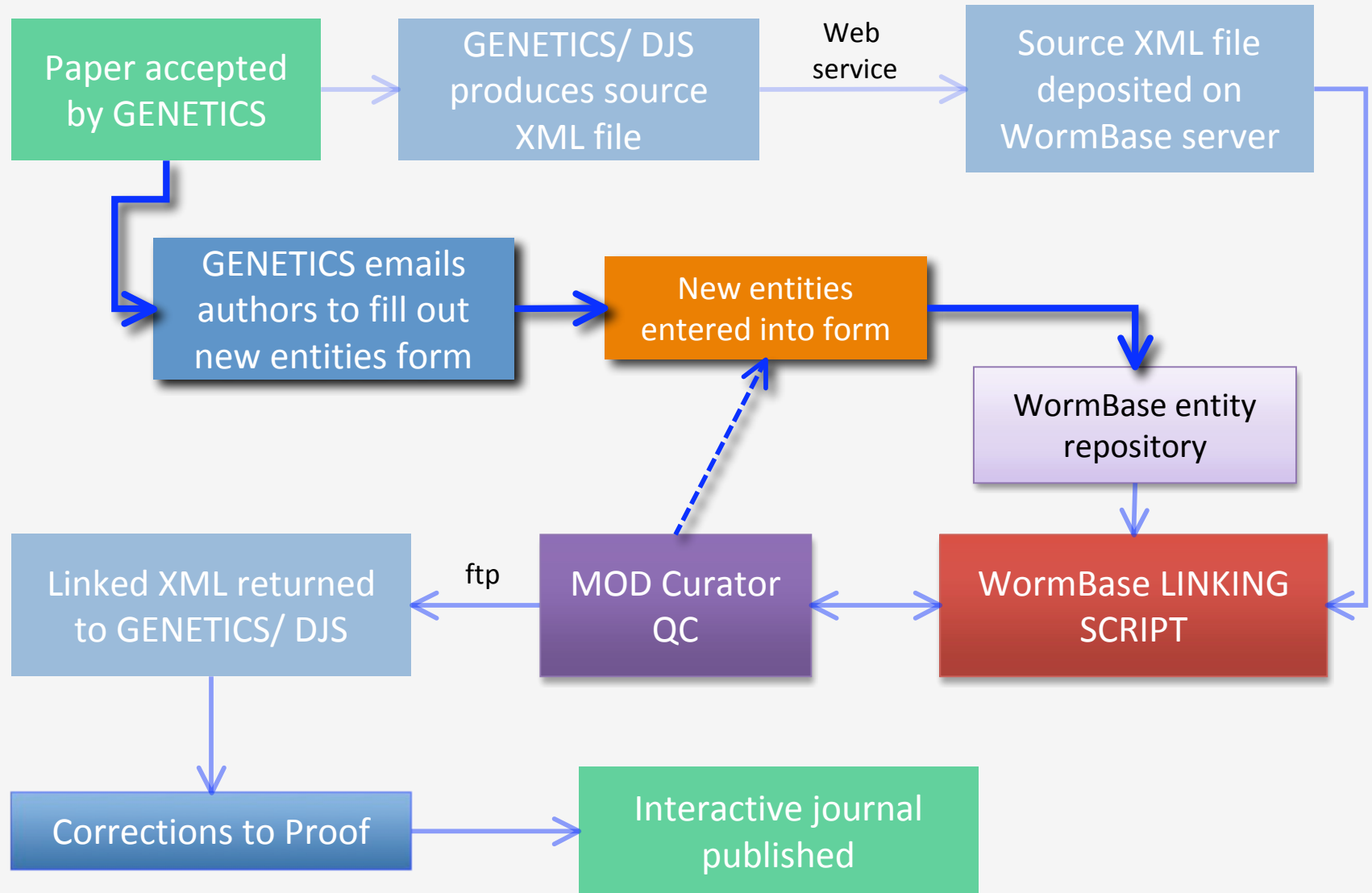
**Silent** links are made using the URL constructors, which will become **live** once the entities are added to the database

# GSA-WormBase markup pipeline





# GSA-WB markup pipeline includes author participation for new entities



# Genetics sends authors an entity declaration form for WormBase

<a href="#">Home</a>	<a href="#">Genome</a>	<a href="#">Synteny</a>	<a href="#">Blast / Blat</a>	<a href="#">WormMart</a>	<a href="#">Markers</a>	<a href="#">Genetic Maps</a>	<a href="#">Submit</a>	<a href="#">Searches</a>	<a href="#">Site Map</a>
----------------------	------------------------	-------------------------	------------------------------	--------------------------	-------------------------	------------------------------	------------------------	--------------------------	--------------------------

**WormBase Release WS218**

**WormBase**

**Congratulations on your paper being accepted by GENETICS for publication!**

GENETICS is working with [textpresso](http://www.textpresso.org) ([www.textpresso.org](http://www.textpresso.org)) and [WormBase](http://www.wormbase.org) ([www.wormbase.org](http://www.wormbase.org)) to create links between genetic and genomic objects (genes, alleles, reagents, etc.) that are in the online full text and PDF formats of your paper to the appropriate page on WormBase. Existing WormBase objects that are in your paper will automatically be linked. *However, if you want any new objects discovered / described in your online paper to be linked to WormBase then you need to tell us what they are.*

Please use the form provided below to enter the names of these objects. This information will be sent to WormBase where these objects will be curated and made available in the next possible release of our database. Follow the examples carefully as your data will be processed automatically. If you would rather upload a file, please contact [kyook@caltech.edu](mailto:kyook@caltech.edu).

Thank you for your help.

Best Wishes,  
WormBase  
<http://www.wormbase.org>  
[wormbase-help@wormbase.org](mailto:wormbase-help@wormbase.org)

Click the "?" to find out more about the data type.  
Click the "Add information" to open or close the data entry box.  
This paper has already had data submitted, loading it now.

**Enter new objects that you want linked in your paper :**

- Genes cloned or mapped in this paper (example: *odf-2*, *mett-10*). ? [Add information.](#)
- Alleles (example: *dh115*, *oz36*, *u53*). ? [Add information.](#)
- Strains (only those being sent to the CGC) (example: YY166, YY173, YY216). ? [Add information.](#)
- Rearrangements (example: *hT96*, *cD157*). ? [Add information.](#)
- Transgenes (example: *amts4*, *amts5*). ? [Add information.](#)
- Cell / anatomy terms and synonyms (example: arcade ring, sex myoblasts). ? [Add information.](#)

# Genetics sends authors an entity declaration form for WormBase

Home Genome Synteny Blast / Blat WormMart Markers Genetic Maps Submit Searches Site Map

WormBase Release WS218

WormBase

Congratulations on your paper being accepted by GENETICS for publication!

GENETICS is working with textpresso ([www.textpresso.org](http://www.textpresso.org)) and WormBase ([www.wormbase.org](http://www.wormbase.org)) to create links between genetic and genomic objects (genes, alleles, reagents, etc.) that are in the online full text and PDF formats of your paper to the appropriate page on WormBase. Existing WormBase objects that are in your paper will automatically be linked. *However, if you want any new objects discovered / described in your online paper to be linked to WormBase then you need to tell us what they are.*

Please use the form provided below to enter the names of these objects. This information will be sent to WormBase where these objects will be curated and made available in the next possible release of our database. Follow the examples carefully as your data will be processed automatically. If you would rather upload a file, please contact [kyook@caltech.edu](mailto:kyook@caltech.edu).

Thank you for your contribution.

Best Wishes,  
WormBase  
<http://www.wormbase.org>

**50% (11/22) author response**

Click the "?" to find out more about the data type.  
Click the "Add information" to open or close the data entry box.  
This paper has already had data submitted, loading it now.

**Enter new objects that you want linked in your paper :**

- Genes cloned or mapped in this paper (example: *odf-2*, *mett-10*). ? [Add information.](#)
- Alleles (example: *dh115*, *oz36*, *u53*). ? [Add information.](#)
- Strains (only those being sent to the CGC) (example: YY166, YY173, YY216). ? [Add information.](#)
- Rearrangements (example: *hT96*, *cD157*). ? [Add information.](#)
- Transgenes (example: *aml54*, *aml55*). ? [Add information.](#)
- Cell / anatomy terms and synonyms (example: arcade ring, sex myoblasts). ? [Add information.](#)

Submit

# Where we stand at the moment

*Caenorhabditis elegans* papers marked up- **42**

**35** published starting in August 2009

**7** waiting to be published

**5** more papers accepted, still to mark up

**10** entity classes marked up

**All** have needed manual QC

(ambiguities, new objects, copy edit errors)

*Saccharomyces cerevisiae* papers marked up- **23**

**2** entity classes marked up

*Drosophila melanogaster* papers will be ready to go live soon

# Current challenges

## Linking to multiple MODs in a single paper

### *C. elegans* eIF1 carrying an equivalent yeast *suil* mutation allows translation initiation at a similar subset of non-AUG codons:

In contrast to what we observed with the *elf-1* mutants described above, non-AUG initiation by yeast eIF1 *suil* mutants occurs at UUG but not with other non-AUG codons (Huang *et al.* 1997). To address whether eIF1 carrying these *suil* mutations are able to confer a similar phenotype in our reporter system, we constructed mutations in the *C. elegans* eIF1 gene (Figure 2A) corresponding to all known *suil* alleles (*suil-1*, *suil-4*, and *suil-17*) isolated in the *His4* suppressor screens (Yoon and Donahue 1992) and *mof2-1*, which was first isolated as a mutant with an altered frameshift efficiency but later found to have a reduction in start codon recognition fidelity as well (Cui *et al.* 1998). Since *suil* mutations are dominant or codominant in yeast (Yoon and Donahue 1992), it is possible that these mutations also behave dominantly in *C. elegans*, allowing the detection of their defects in our transgenic assay. These mutants were initially assayed with non-AUG GFP reporters carrying GUG or UUG codons. The *C. elegans* eIF1 transgene carrying the *suil-4* mutation promoted GFP expression from both GUG and UUG reporters unlike transgenes carrying the wild-type eIF1 and the mutant eIF1 genes containing the *suil-1*, *suil-17*, or *mof2-1* mutations (data not shown). These results indicate that only the *suil-4* mutant was able to confer non-AUG translation in the *C. elegans* reporter system, which is in contrast to that observed in yeast where all these mutations allowed translation to start at the UUG codon. We further assayed the ability of the *suil-4* mutant to allow GFP expression from reporters containing other non-AUG start codons. GFP expression was observed from additional reporters that changed a single base either at the first or the third base position of the AUG codon (Table 1, lines 1, 3 and

Home Genome Synteny Blast / Biat WormMart Markers Genetic Maps Submit Searches Site Map

Find: WBGene00020868 [Anything]

**WormBase**

Gene Summary Locus Summary Sequence Summary Protein Summary EST Alignments Genome Browser Genetic Map Nearby Genes Bibliography Tree Display XML Schema Aceb Image

### Gene Summary for *elf-1*

Specify a gene using a gene name (*unc-26*), a predicted gene id (R13A5.9), or a protein ID (CE02711) *et al.*

[Identification][location][function][expression][gene ontology][genetics][homology][reagents][bibliography]

IDs:	Main name	Sequence name	Other name(s)	WB Gene ID
	<i>elf-1</i> - (Eukaryotic Initiation Factor)	T27F7.3	PIG-B (via accession evidence: EMBL:AJ431373)	WBGene00020868

Concise Description: none available

NCBI KOGs\*: Translation initiation factor 1 (eIF-1/SUI1) [KOG1770]; [OMpre\_WH004978]

Species: *Caenorhabditis elegans*

SGD Search

Site Map | Search Options | Help | Home | [Icons]

Community Info Submit Data BLAST Primers PatMatch Gene/Seq Resources Advanced Search Community Wiki

### SUI1/YNL244C Summary

Summary Locus History Literature Gene Ontology Phenotype Interactions Expression Protein Wiki

#### SUI1 BASIC INFORMATION

Standard Name	SUI1 <sup>1</sup>
Systematic Name	YNL244C
Alias	MOF2 <sup>2</sup> , RFR1
Feature Type	ORF, Verified
Description	Translation initiation factor eIF1; component of a complex involved in recognition of the initiator codon; modulates translation accuracy at the initiation phase [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100]

#### SUI1 RESOURCES

Click on map for expanded view

SGD ORF map GBrowse

126000 to 121000 Chr22V

5' UPTC GAGT 3' YNL244W YNL243W

5' GAGT 3' YNL245C YNL244C

• Literature [Literature Guide] [View]

# Current challenges

## Linking fly genes to FlyBase

Many fly gene names are difficult to distinguish from standard everyday words, e.g., *we*, *for*, *a*  
Information-rich **XML-tagged** files help identify these genes

“Therefore, we tested members of major pathways that control DC such as **Jun-related antigen** (**Jra**), **puckered** (**puc**), and **Src homology 2 ankyrin repeat tyrosine kinase** (**shark**) in the DJNK signaling pathway and **thickveins** (**tkv**), **schnurri** (**shn**), and **zipper** (**zip**) in the TGF- $\beta$  signaling pathway. Furthermore, we tested additional genes involved in DC such as **u-shaped** (**ush**), **Epidermal growth factor receptor** (**Egfr**), **Protein kinase related to protein kinase N** (**Pkn**), **scab** (**scb**), and **Rho1**....”

# Benefits of published text markup

## For the community:

- Less time to get to the correct database page
- Transparency of science jargon
- Term ambiguities are dealt with before publication

## For the journal:

- Becomes a live portal to science information

## For the author:

- Extra pair of eyes for proofing their document
- Data enter the database faster

## For the database:

- Increase depth and accuracy of data in the database
- Get information before publication
- Resolve conflicts in term use before publication

# Future challenges

- Extending this pipeline to other MODs / Databases
- Extending this pipeline to other journals
  - standardized XML format can help
- Increasing the number of Databases and entity classes linked in a multi-organism paper
- Fixing broken links
- Re-marking up publications



# Acknowledgements



Arun Rangarajan  
Karen Yook  
Juancarlos Chan  
Hans-Michael Muller  
Paul W. Sternberg (HHMI)



Tracey DePellegrin-Connelly  
Ruth Isaacson  
Tim Schedl (Wash.U. School of Medicine)



Stephen Haenel  
Lolly Otis  
Sharon Faelten



Saccharomyces Genome Database  
Marek S. Skrzypek, J. Michael Cherry (Stanford University School of Medicine)



William M. Gelbart (Harvard University)  
Thom Kaufman, Kathy Matthews (Indiana University)  
Raymund Stefanicsik, Steven Marygold (Cambridge University, UK)