



Swiss Institute of  
Bioinformatics

# Evolutionary study of gene function

Marc Robinson-Rechavi

SIB course Biodiversity bioinformatics 2023



*Unil*  
UNIL | Université de Lausanne  
Département d'écologie  
et évolution



All original material CC0; all other  
material copyright of the original owners



@marcrr@ecoevo.social



[www.bgee.org](http://www.bgee.org)

# Three parts

---

1. Function in model vs non model organisms
2. Function large scale
3. Gene and genome duplication: evolution of function

# Who am I?



PhD: Evolution of rodent genomes, 1997



Postdoc / Lecturer / Postdoc: Zoology, Evo-Devo, Structural biology



Prof bioinformatics, Department of Ecology and Evolution, University of Lausanne, since 2005

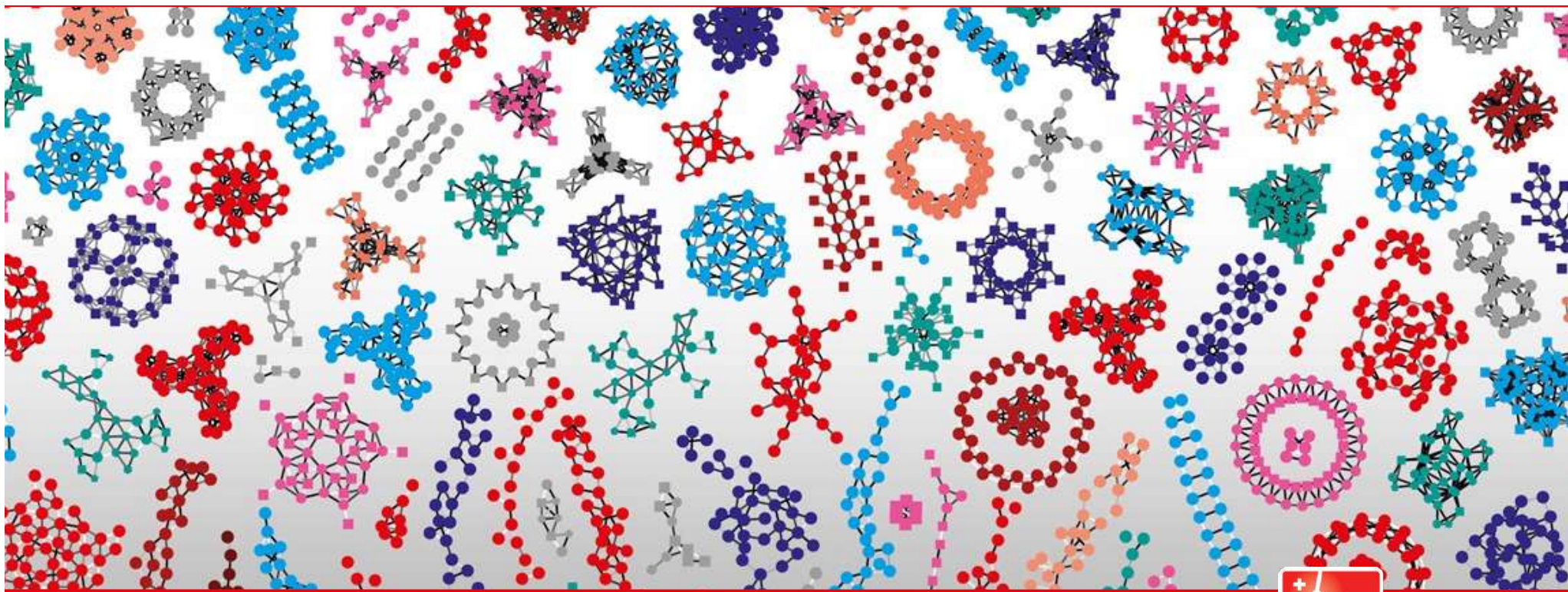


Research group:



- Animal Evo-Devo
- Gene duplication
- Natural selection
- Bgee database





Swiss Institute of  
Bioinformatics

# 1- Function in model vs non model organisms

Marc Robinson-Rechavi

SIB course Biodiversity bioinformatics 2023



*Unil*  
UNIL | Université de Lausanne  
Département d'écologie  
et évolution



@marcrr@ecoevo.social

[www.bgee.org](http://www.bgee.org)

---

# What is function?

---

# Poll

---

What is the function of the heart?

1. to pump blood through the body
  2. to make a noise that Doctors can listen to
  3. to get heart attacks
-

# Function in biology

---

## Selected-effect definition of function:

- **what a structure was selected to do**

## Causal role definition of function:

- **ahistorical ways in which a component contributes to a capacity of a system**

# BRCA1 "Breast cancer type 1 susceptibility protein"

## Function<sup>1</sup>

E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA damage (PubMed:12890688, PubMed:14976165, PubMed:16818604, PubMed:17525340, PubMed:12887909, PubMed:10500182, PubMed:19261748).

It is unclear whether it also mediates the formation of other types of polyubiquitin chains (PubMed:12890688).

The BRCA1-BARD1 heterodimer coordinates a diverse range of cellular pathways such as DNA damage repair, ubiquitination and transcriptional regulation to maintain genomic stability (PubMed:12890688, PubMed:14976165, PubMed:20351172).

Regulates centrosomal microtubule nucleation (PubMed:18056443).

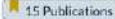
Required for appropriate cell cycle arrests after ionizing irradiation in both the S-phase and the G2 phase of the cell cycle (PubMed:10724175, PubMed:12183412, PubMed:11836499, PubMed:19261748).













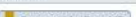









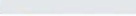



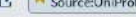

Required for FANCD2 targeting to sites of DNA damage (PubMed:12887909).

Inhibits lipid synthesis by binding to inactive phosphorylated ACACA and preventing its dephosphorylation (PubMed:16326698).

Contributes to homologous recombination repair (HRR) via its direct interaction with PALB2, fine-tunes recombinational repair partly through its modulatory role in the PALB2-dependent loading of BRCA2-RAD51 repair machinery at DNA breaks (PubMed:19369211).

Component of the BRCA1-RBBP8 complex which regulates CHEK1 activation and controls cell cycle G2/M checkpoints on DNA damage via BRCA1-mediated ubiquitination of RBBP8 (PubMed:16818604).

Acts as a transcriptional activator (PubMed:20160719). 

ASPECT	TERM	Source	Publications
Molecular Function	<a href="#">tubulin binding</a> 	 Source:UniProtKB	 1 Publication
Molecular Function	<a href="#">ubiquitin protein ligase binding</a> 	 Source:UniProtKB	 1 Publication
Molecular Function	<a href="#">ubiquitin-protein transferase activity</a> 	 Source:UniProtKB	 3 Publications
Molecular Function	<a href="#">zinc ion binding</a> 	 Source:ProtInc	 1 Publication
Biological Process	<a href="#">cellular response to indole-3-methanol</a> 	 Source:UniProtKB	 1 Publication
Biological Process	<a href="#">cellular response to ionizing radiation</a> 	 Source:ComplexPortal	 1 Publication
Biological Process	<a href="#">cellular response to tumor necrosis factor</a> 	 Source:BHF-UCL	 1 Publication
Biological Process	<a href="#">centrosome cycle</a> 	 Source:Ensembl	
Biological Process	<a href="#">chordate embryonic development</a> 	 Source:GO-Central	 1 Publication
Biological Process	<a href="#">chromosome segregation</a> 	 Source:UniProtKB	 1 Publication
Biological Process	<a href="#">DNA damage response</a> 	 Source:ProtInc	 1 Publication

## Involvement in disease<sup>1</sup>

### Breast cancer (BC)

 22 Publications

**Note** Disease susceptibility is associated with variants affecting the gene represented in this entry. Mutations in BRCA1 are thought to be responsible for 45% of inherited breast cancer. Moreover, BRCA1 carriers have a 4-fold increased risk of colon cancer, whereas male carriers face a 3-fold increased risk of prostate cancer. Cells lacking BRCA1 show defects in DNA repair by homologous recombination

**Description** A common malignancy originating from breast epithelial tissue. Breast neoplasms can be distinguished by their histologic pattern. Invasive ductal carcinoma is by far the most common type. Breast cancer is etiologically and genetically heterogeneous. Important genetic factors have been indicated by familial occurrence and bilateral involvement. Mutations at more than one locus can be involved in different families or even in the same case.



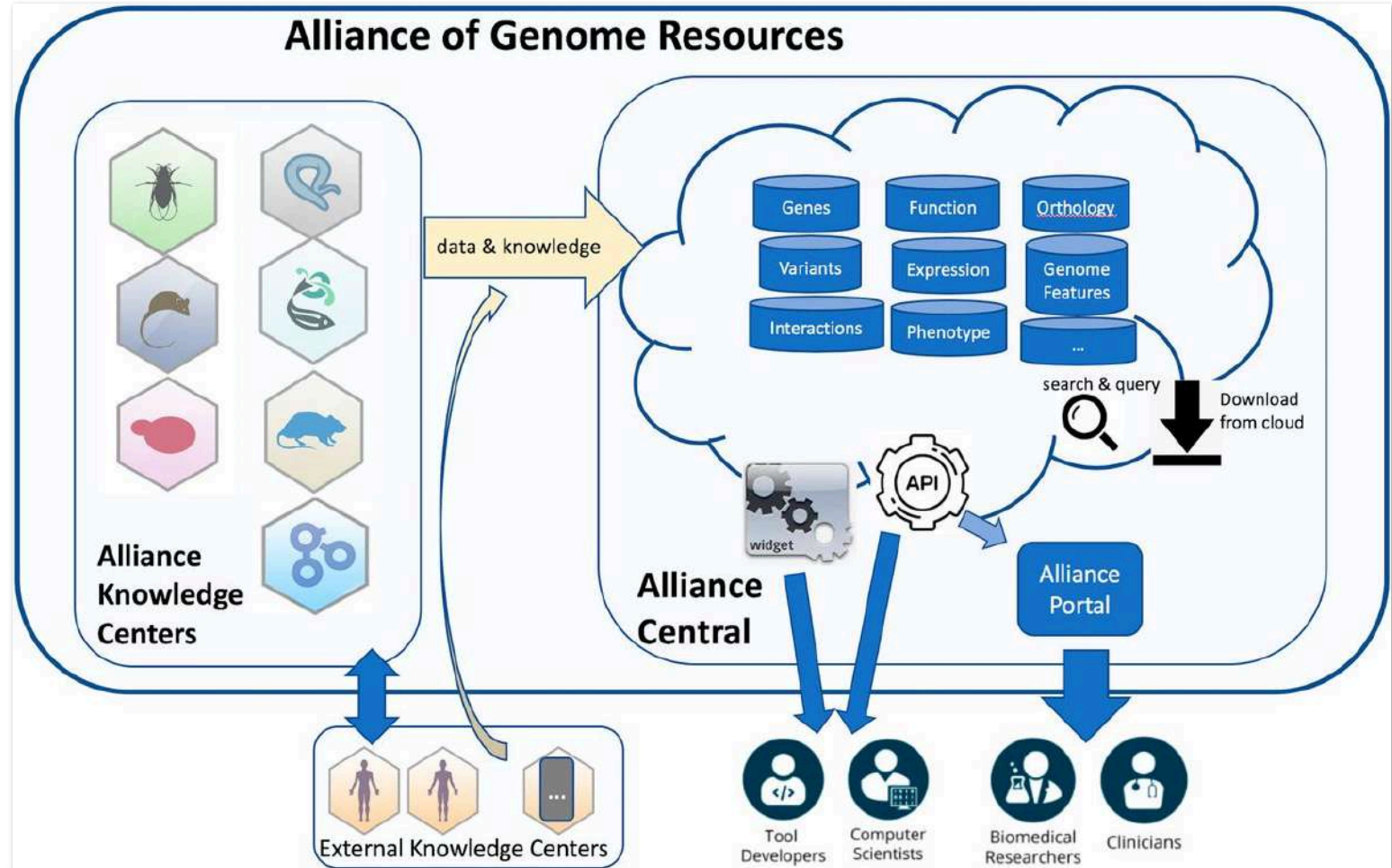
---

# Sources of information on function

---

# Model Organism Databases

Flybase  
Wormbase  
RGD  
ZFIN  
SGD  
MGI  
Gene Ontology



# Annotation

---

Associating a biological object to a feature, based on evidence.

Examples:

- associating a Gene Ontology term (e.g. "centrosome cycle") to a gene based on best Blast hit (**uncurated annotation**)
  - associating a Gene Ontology term (e.g. "tubulin binding") to a gene based on reading an experimental paper (**curated annotation**)
-

# Functional annotation in model organisms

---

## Evidence

- small scale experiments: mutants, in situ...
- large scale experiments: RNA-seq, screens...
- phenotypes, expression, interactions, diseases, ...
- computational: domains, orthology...

## Annotation to genes or proteins

- Model Organism Databases
- generalist databases, e.g., UniProt
- curated or automatic



# Functional annotation in non model organisms

---

## Evidence

- ~~• small scale experiments: mutants, in situ...~~
- large scale experiments: RNA-seq, ~~screens...~~
- ~~• phenotypes, expression, interactions, diseases, ...~~
- computational: domains, orthology...

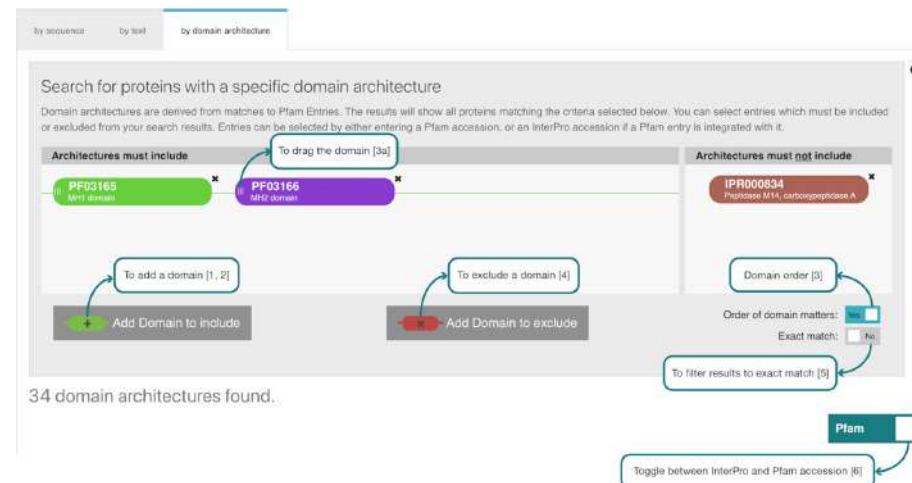
## Annotation to genes or proteins

- ~~• Model Organism Databases~~
- generalist databases, e.g., UniProt
- ~~• curated or automatic~~

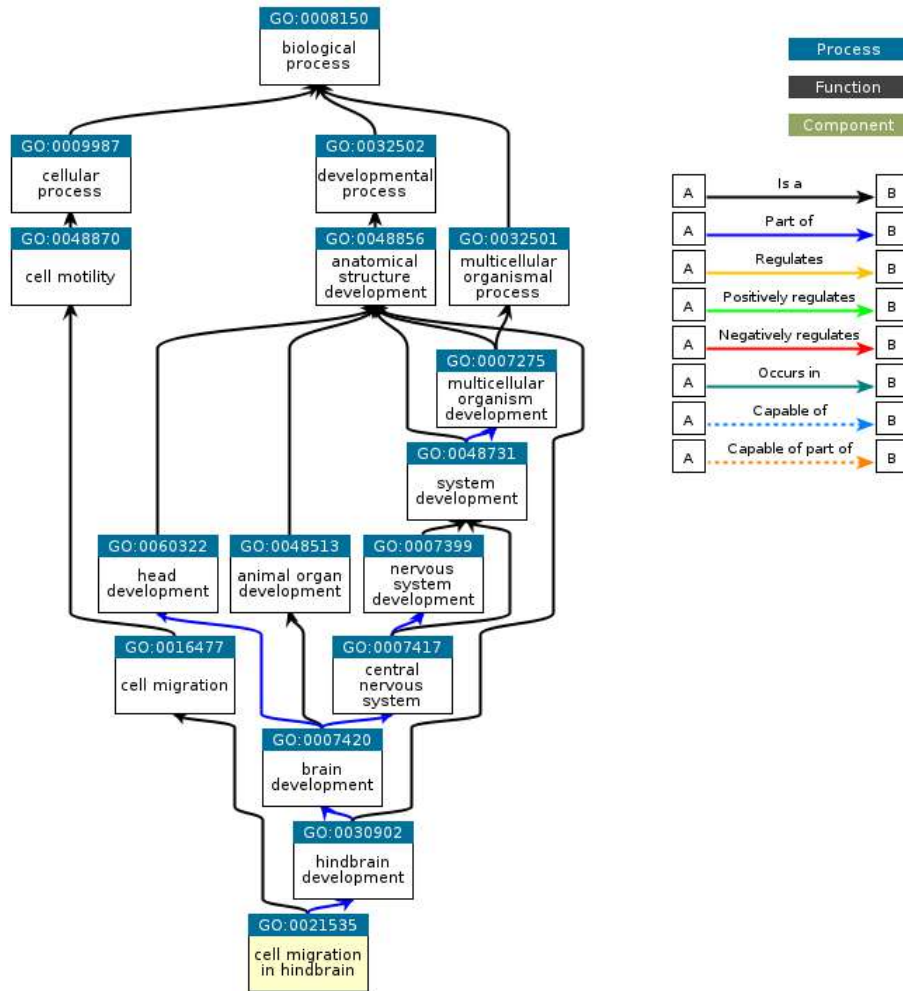


# Computational annotation of non model genomes

- Orthology or similarity network-based methods
  - **many software, fast moving field**
- InterProScan:
  - **curated associations domain – function**
  - **find domains in predicted proteins**



# The Gene Ontology



**O42366 · HXB1A\_DANRE**

Protein<sup>1</sup> Homeobox protein Hox-B1a  
 Gene<sup>1</sup> hoxb1a  
 Status<sup>1</sup> UniProtKB reviewed (Swiss-Prot)  
 Organism<sup>1</sup> Danio rerio (Zebrafish) (Brachydanio rerio)

Amino acids 311 (go to sequence)  
 Protein existence<sup>1</sup> Evidence at transcript level  
 Annotation score<sup>1</sup> 4.0

Entry Variant viewer Feature viewer Publications External links History

BLAST Download Add Add a publication Entry feedback

**Function<sup>1</sup>**  
 Sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis. By Similarity

**GO annotations<sup>1</sup>**  
 Slimming set: generic

all annotations  
 all molecular function  
 all biological process  
 all cellular component

ASPECT	TERM
Cellular Component	nucleus <sup>1</sup> <a href="#">Source:GO_Central</a> <sup>1</sup> <a href="#">Publication</a>
Molecular Function	DNA-binding transcription factor activity, RNA polymerase II-specific <sup>1</sup> <a href="#">Source:GO_Central</a> <sup>1</sup> <a href="#">Publication</a>
Molecular Function	RNA polymerase II cis-regulatory region sequence-specific DNA binding <sup>1</sup> <a href="#">Source:GO_Central</a> <sup>1</sup> <a href="#">Publication</a>
Molecular Function	sequence-specific DNA binding <sup>2</sup> <a href="#">Source:ZFIN</a> <sup>2</sup> <a href="#">Publications</a>
Biological Process	cell migration in hindbrain <sup>1</sup> <a href="#">Source:ZFIN</a> <sup>1</sup> <a href="#">Publication</a>
Biological Process	facial nerve development <sup>2</sup> <a href="#">Source:ZFIN</a> <sup>2</sup> <a href="#">Publications</a>
Biological Process	facial nucleus development <sup>1</sup> <a href="#">Source:ZFIN</a> <sup>1</sup> <a href="#">Publication</a>
Biological Process	hindbrain maturation <sup>1</sup> <a href="#">Source:ZFIN</a> <sup>1</sup> <a href="#">Publication</a>
Biological Process	regulation of transcription by RNA polymerase II <sup>1</sup> <a href="#">Source:GO_Central</a> <sup>1</sup> <a href="#">Publication</a>
Biological Process	rhombomere 4 development <sup>1</sup> <a href="#">Source:ZFIN</a> <sup>1</sup> <a href="#">Publication</a>
Biological Process	rhombomere 4 morphogenesis <sup>1</sup> <a href="#">Source:ZFIN</a> <sup>1</sup> <a href="#">Publication</a>

QuickGO - <https://www.ebi.ac.uk/QuickGO>

# Gene Ontology: 3 ontologies

---

## Molecular function

- what does the gene product do?

## Biological process

- in which process does it act?

## Cellular component

- where does it act?
-



# The GO describes selected effect functions

---

The GO does not cover

- pathological roles ("lymphome")
  - experimental conditions ("in vitro")
  - evolutionary relations ("orthologue of")
  - gene product type ("protein")
-

# Discussion

---

You have a new genome, what's the next experiment?

---

# Gene expression as an indicator of function

---

## RNA-seq:

- cheap and straightforward to apply to different species, conditions, or individuals

## Provides:

- where and when a gene is expressed
- how highly it is expressed
- which genes are co-expressed
- link between phylogenomics and Evo-Devo

## Doesn't provide:

- specific aspects of the phenotype
  - information for housekeeping genes
-

---

**Bgee**

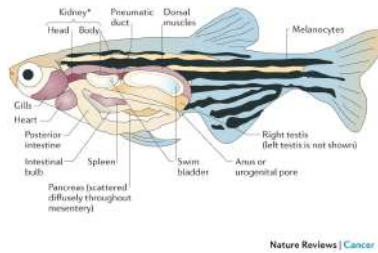
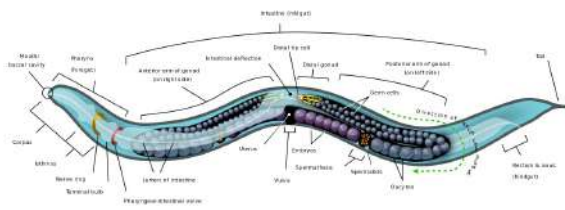
---

<https://www.bgee.org>

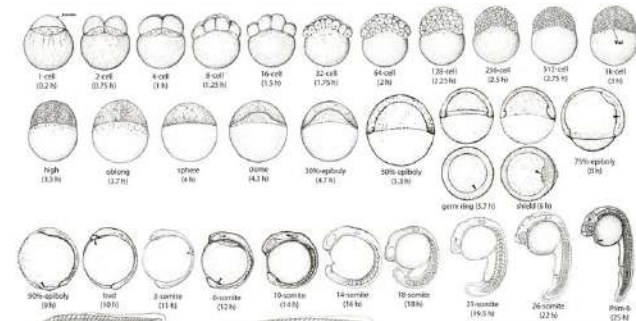
# Bgee condition annotations

Anatomical entity  
+ cell type

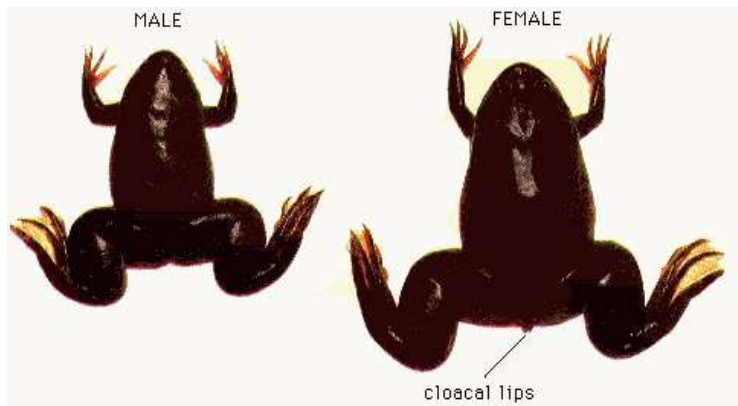
Developmental and life stage



Nature Reviews | Cancer



Sex



Strain



# Curation: wild-type healthy only

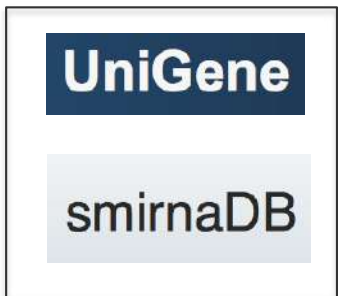
---

Informative on selected function of genes

Evolutionarily relevant (comparisons between species)

Reference for biomedical studies

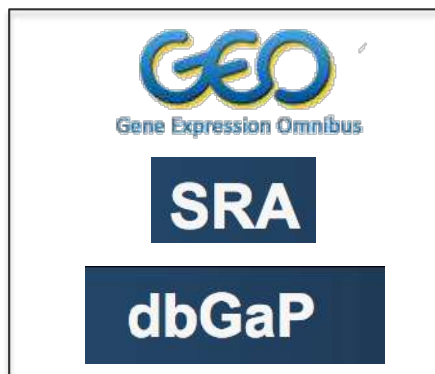
---



EST data



Affymetrix data



Bulk RNA-Seq and  
scRNA-Seq data



*In situ*  
hybridization data

Quality control and condition filtering

Raw data processing

Data standardization and mapping to ontologies

Integrate all data types in Bgee

# Data integration in Bgee

---

Calls present / absent from each data type

➤ integrated calls

Expression level or evidence rank for each data type

➤ integrated expression scores

---



# Apoc1 gene (liver specific inhibitor of lipoprotein)



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0001114</a> <sup>Ⓜ</sup> right lobe of liver	99.94	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>
<a href="#">UBERON:0035825</a> <sup>Ⓜ</sup> left adrenal gland cortex	99.84	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>
<a href="#">UBERON:0035827</a> <sup>Ⓜ</sup> right adrenal gland cortex	99.80	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0001115</a> <sup>Ⓜ</sup> left lobe of liver	99.92	0.002	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0002107</a> <sup>Ⓜ</sup> liver	99.96	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0002107</a> <sup>Ⓜ</sup> liver	99.64	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0001235</a> <sup>Ⓜ</sup> adrenal cortex	99.73	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>
<a href="#">UBERON:0000992</a> <sup>Ⓜ</sup> ovary	99.64	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>
<a href="#">UBERON:0002107</a> <sup>Ⓜ</sup> liver	99.54	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0002107</a> <sup>Ⓜ</sup> liver	99.76	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0002107</a> <sup>Ⓜ</sup> liver	99.97	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0002107</a> <sup>Ⓜ</sup> liver	99.97	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>



Anatomical entity	Expression score	FDR	Link to source data	Sources
<a href="#">UBERON:0002107</a> <sup>Ⓜ</sup> liver	100.00	$\leq 1.00e-14$	<a href="#">See source data</a>	<a href="#">R</a> <a href="#">S</a> <a href="#">C</a> <a href="#">A</a> <a href="#">I</a> <a href="#">E</a>

---

Diverse genomes: 👍

Function of diverse genes: 😄💧

---

