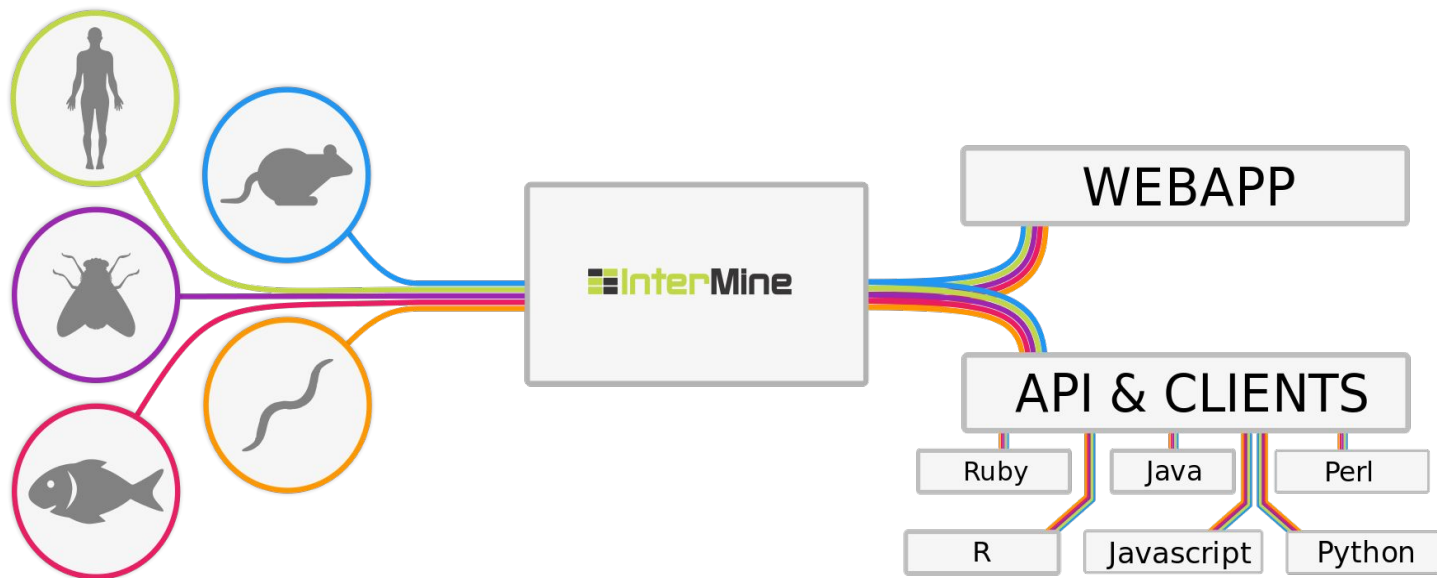

Implementing FAIR identifiers in InterMine with Identifiers.org

Daniela Butano, Software Engineer - daniela@intermine.org

J.Clark-Casey, S.Contrino, J.Heimbach, R.Lyne, J.Sullivan, Y.Yehudi, and G.Micklem

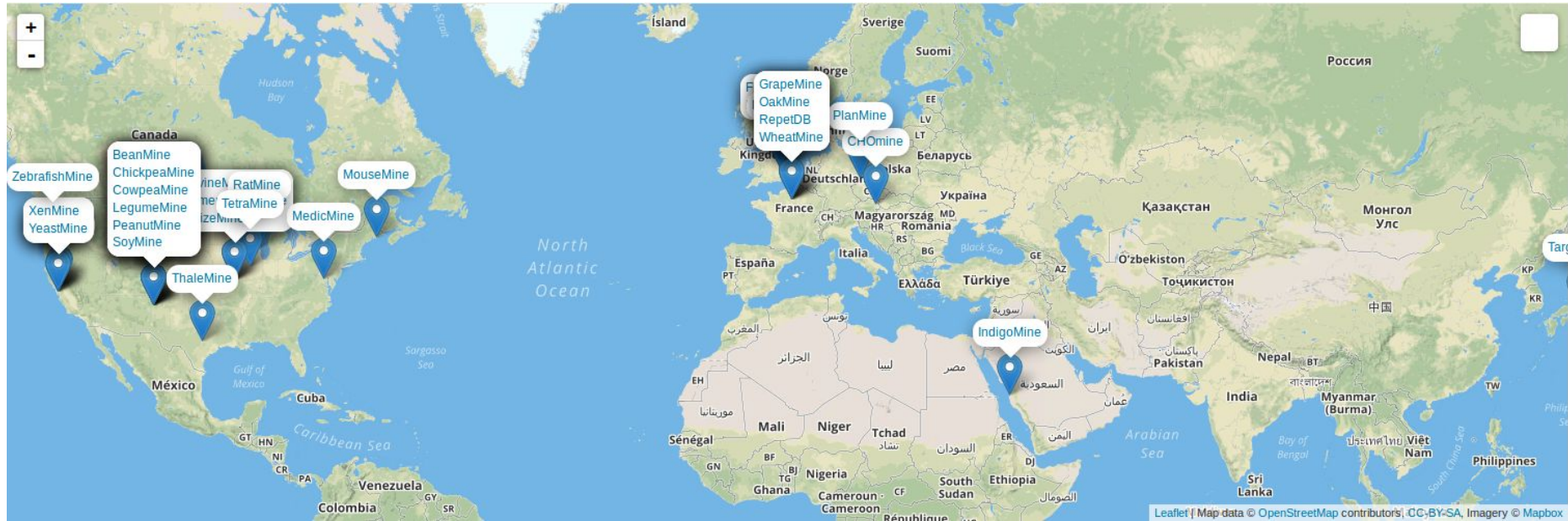
Building a FAIR Bioinformatics environment
October 22-24, 2018, Genoa

InterMine Overview



Model organism images Designed by Freepik and distributed by Fla

InterMine around the world



URI: Uniform Resource Identifiers

From FAIR principles:

F1: (meta)data are assigned a globally unique and persistent identifier

A1: (meta)data are retrievable by their identifier using a standardized communication protocol



Assigning URIs to entities:

<http://purl.uniprot.org/uniprot/P12883>

<http://identifiers.org/biosample/SAMN00808411>

<http://identifiers.org/ncbigene/5468>

Current InterMine URIs change
at every build

Gene : PARG *Homo sapiens*

Name ⁰ **peroxisome proliferator activated receptor gamma** Cytological Location **3p25.2**
 Brief Description **peroxisome proliferator activated receptor gamma**
 synonyms: PPARgamma, PARG2, GLM1, X90563, NM_005037, HGNC:9236, PARG2, [Show more](#)
 identifiers: 5468, ENSG00000132170, PARG
 Region: **gene** ⁰ Length: 183570 FASTA...
 Location: **3:12287485-12471054** Cyto location: **3p25.2**

25 Pathways Reactome, KEGG 4 Diseases OMIM 30 Mouse Alleles (MGI) mouse alleles 127 Gene Ontology

Tissue Expression **15** ↓ **14** ↑ Genes Expression **23** ↓ **21** ↑

SMART

Quick Links: [Summary](#) [Function](#) [Genomics](#) [Proteins](#) [SNPs](#) [Disease](#) [Homology](#) [Interactions](#) [Gene Ontology](#) [Other](#)

Curated comments from UniProt

Type	Comment	Show proteins
disease	Defects in PARG can lead to type 2 insulin-resistant diabetes and hypertension. PARG mutations may be associated with colon cancer.	
disease	MIM:137800; Glioma 1; GLM1; Gliomas are benign or malignant central nervous system neoplasms derived from glial cells. They comprise astrocytomas and glioblastoma multiforme that are derived from astrocytes, oligodendrogliomas derived from oligodendrocytes and ependymomas derived from ependymocytes. Disease susceptibility may be associated with variations affecting the gene represented in this entry. Polymorphic PARG alleles have been found to be significantly over-represented among a cohort of American patients with sporadic glioblastoma multiforme suggesting a possible contribution to disease susceptibility.	
disease	MIM:601665; Obesity; OBESITY: A condition characterized by an increase of body weight beyond the limitation of skeletal and physical requirements, as the result of excessive accumulation of body fat. Disease susceptibility may be associated with variations affecting the gene represented in this entry.	
disease	MIM:604367; Lipodystrophy, familial partial, 3; FPLD3: A form of lipodystrophy characterized by marked loss of subcutaneous fat from the extremities. Facial adipose tissue may be increased, decreased or normal. Affected individuals show an increased preponderance of insulin resistance, diabetes mellitus and dyslipidemia. The disease is caused by mutations affecting the gene represented in this entry.	
function	(Microbial infection) Upon treatment with <i>M.tuberculosis</i> or its lipoprotein LpH, phosphorylation of MAPK p38 and IL-6 production are modulated, probably via this protein.	
function	Nuclear receptor that binds peroxisome proliferators such as hypolipidemic drugs and fatty acids. Once activated by a ligand, the nuclear receptor binds to DNA specific PPAR response elements (PPRE) and modulates the transcription of its target genes, such as acyl-CoA oxidase. It therefore controls the peroxisomal beta-oxidation pathway of fatty acids; Key regulator of adipocyte differentiation and glucose homeostasis. ARF6 acts as a key regulator of the tissue-specific adipocyte P2 (ap2) enhancer. Acts as a critical regulator of gut homeostasis by suppressing NF-kappa-B-mediated proinflammatory responses. Plays a role in the regulation of cardiovascular circadian rhythms by regulating the transcription of ARNTL/BMAL1 in the blood vessels (By similarity).	
function	Nuclear receptor that binds peroxisome proliferators such as hypolipidemic drugs and fatty acids. Once activated by a ligand, the nuclear receptor binds to DNA specific PPAR response elements (PPRE) and modulates the transcription of its target genes, such as acyl-CoA oxidase. It therefore controls the peroxisomal beta-oxidation pathway of fatty acids; Key regulator of adipocyte differentiation and glucose homeostasis. ARF6 acts as a key regulator of the tissue-specific adipocyte P2 (ap2) enhancer. Acts as a critical regulator of gut homeostasis by suppressing NF-kappa-B-mediated proinflammatory responses. Plays a role in the regulation of cardiovascular circadian rhythms by regulating the transcription of ARNTL/BMAL1 in the blood vessels.	
similarity	Belongs to the nuclear hormone receptor family. NRI subfamily.	
subcellular location	Redistributed from the nucleus to the cytosol through a MAP2K1/MEK1-dependent manner. NOCT enhances its nuclear translocation.	
tissue specificity	Highest expression in adipose tissue. Lower in skeletal muscle, spleen, heart and liver. Also detectable in placenta, lung and ovary.	

Tissue Expression (ArrayExpress)

lung transplant
prostate
cerebellum
testis



Key Controls Help
Expression
Ureoulation

Lists

This Gene is in 4 lists:
 PL_DiabetesGenes (68)
 PL_BHF_UCL_cardiovascGenes (4103)
 PL_monogenGlucTolerance_ORahilly09 (20)
 PL_DiabetesGWAS_pval-4 (711)

Links to other Mines

FlyMine
D. melanogaster
Eip75B ⁰

MouseMine
M. musculus
Pparg ⁰

YeastMine No results

RatMine
R. norvegicus
Pparg ⁰

ZebrafishMine
D. rerio
ZDB-GENE:990415-213 ⁰

External Links

ensembl ⁰
 BioGPS ⁰
 ArrayExpress Atlas ⁰
 Clinvar ⁰
 NCBI entrez ⁰
 HuGe ⁰
 BioGRID ⁰

New FAIR identifiers for InterMine


Identified two possible solutions

- ✓ No InterMine generated IDs
 - ✓ Based on external IDs: LUI = Local Unique Identifier issued by the data source provider
1. Using Identifiers.org prefixes:
`humanmine.org/humanmine/ncbigene:5468`
 2. Using InterMine core model class names:
`humanmine.org/humanmine/gene:5468`

Schema A: using Identifiers.org prefixes and LUI

humanmine.org/humanmine/**ncbigene:5468**

- ✓ URI is guaranteed unique
- ✓ No embedded type
- ✓ Automatic links to other sources



Prefixes in Identifiers.org
biosample
ensembl
kegg.pathway
ncbigene
pubmed
reactome
uniprot

Schema B: using core model class names and LUI

humanmine.org/humanmine/gene:5468

- ✓ No configuration required
- ✓ Human-readable
- ✓ Straightforward URI resolution

InterMine core.xml

```
<class name="BioEntity" is-interface="true" extends="Annotatable">
  <attribute name="secondaryIdentifier" type="java.lang.String"/>
  <attribute name="symbol" type="java.lang.String"/>
  <attribute name="name" type="java.lang.String"/>
  ....
</class>
<class name="SequenceFeature" extends="BioEntity" is-interface="true">
  <attribute name="score" type="java.lang.Double"/>
  <attribute name="length" type="java.lang.Integer"/>
  ....
</class>
<class name="Gene" extends="SequenceFeature" is-interface="true">
  <attribute name="description" type="java.lang.String"/>
  <attribute name="briefDescription" type="java.lang.String"/>
  ....
</class>
```

Schema B

It requires less human intervention

Identifiers.org as Permanent URL provider

Registering HumanMine in Identifiers.org with URI *http://identifiers.org/humanmine*

<http://identifiers.org/humanmine:protein:D2KUA6>



<http://humanmine.org/humanmine/protein:D2KUA6>

InterMine Team

Gos Micklem Lab

-  Daniela Butano
-  Sergio Contrino
-  Julie Sullivan
-  Yo Yehudi
-  Rachel Lyne



-  www.intermine.org
-  info@intermine.org
-  github.com/intermine
-  twitter.com/intermineorg



Thank you