

# Semantic MediaWiki as a platform for lab management and biological annotation

Toni Hermoso Pulido (@toniher)

Bioinformatics Core Facility

Centre for Genomic Regulation (BCN)

<https://biocore.crg.eu>

# Context

## Work in laboratories or core facilities

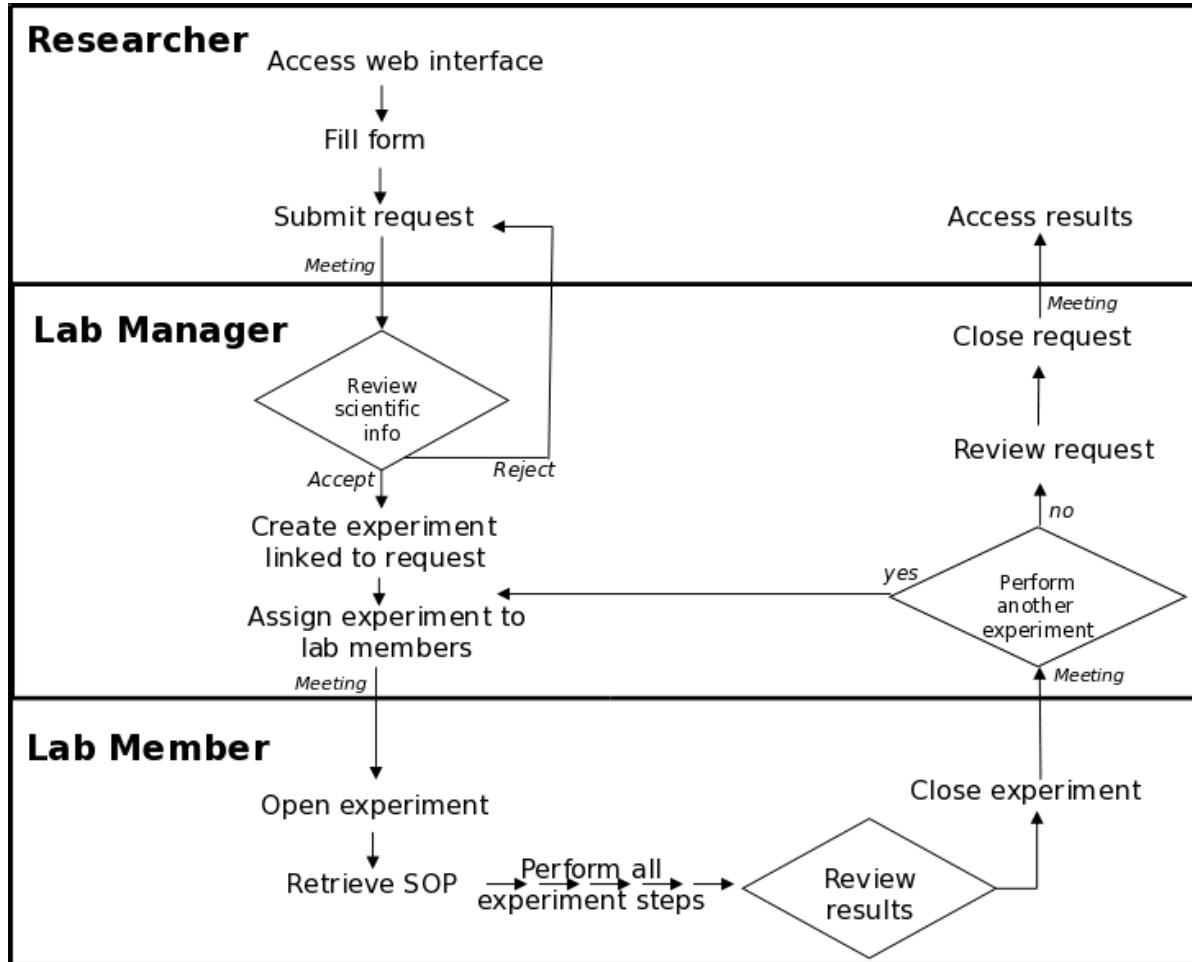


# ProteoWiki

*LIMS: Lab Information Management System*

**Proteomics Unit, CRG**

# ProteoWiki



# ProteoWiki

## Pending requests [\[edit\]](#)

Request number	Creation date	User	Quotation	Assigned to	Sample reception	Type	Budget
20160123248	05 Sep 2016			Cchiva	missing samples		budget
201601751	20 Sep 2016				missing samples		budget
2016011253	26 Sep 2016				missing samples		budget
201602403	26 Sep 2016		61 10	Gespadas			budget
20160104	28 Sep 2016				missing samples		budget

## Active requests [\[edit\]](#)

Request number	Creation date	Delivery Time	User	Quotation	Responsible	Budget	Samples
20160157	12 Apr 2016	5 weeks		61 30	Eborras		
20160303	20 May 2016	8 weeks		€ 36	Eborras		
20160614	28 Jun 2016	8 weeks		61 19	Cchiva		
20160704	05 Jul 2016	4 weeks		616 4	Eborras		
20160603	06 Jul 2016	12 weeks			Cchiva	budget	
2016017	21 Jul 2016	12 weeks			Cchiva	budget	
201601541	22 Jul 2016	3 weeks		61 1	Cchiva		

# ProteoWiki

## Form input



Edit Experiment: Experiment:20160406 [redacted]@2016 [redacted]

Request Reference

Requester

Status

Sample	Reduction, Alkylation and Digestion	<b>LC-MS Quantitation</b>	Data Analysis	[edit]
<b>Desalted</b>	<input checked="" type="checkbox"/>	<b>Method of desalation</b>	UltraMicroSpin column ▾	
<b>Label type</b>		Label-free		
<b>Number of states:</b>				
<b>Sample mix ratio (if applicable)</b>				
<b>Instrument</b>		Fusion Lumos ▾		
<b>Fragmentation</b>		HCD ▾		
<b>Amount injected (µg)</b>		% of sample ▾		
		% of sample:	<input type="text" value="10"/>	
<b>Volume injected (µl)</b>		<input type="text" value="1"/>		
<b>LC column</b>		NTCC-360/75-3-123 ▾		
<b>LC solvent</b>				
<b>LC-MS method</b>		Other ▾		
		<b>Other MS method</b>		
		<input type="text" value="DDA_Universal_TopSpeed_HCD_IT_90min"/>		

# Mail communication

- Based on Semantic Tasks extension
- Asking user for action (bring samples to the lab)
- Informing user about request status
- Users can opt out verbose communication

# User satisfaction tracking

- When request closed
- Email sent. User directed to a Special Page form
- Valid for a limited time (e. g., 2 weeks max)
- Only editable a few times (or only once)

# User satisfaction tracking

20160010120021	10 May 2016	10 Jul 2016	01	4	General - No Info
201600108000	01 Jun 2016	13 Jul 2016	01	38	General - No Info
201600103000	20 May 2016	12 Jul 2016	01	38	General - Good
2016001090914	25 May 2016	12 Jul 2016	NA	NA	General - Good
201600102736	02 May 2016	12 Jul 2016	01	91	General - No Info
20151010302	18 Nov 2015	10 Jul 2016	6	94 / 35	General - Good
201600101446	03 May 2016	08 Jul 2016	61	?	General - No Info
201600103924	26 May 2016	28 Jun 2016	01	53	General - No Info
201600103723	20 May 2016	23 Jun 2016	01	51	General - No Info

# Lab operators extra input

- Wiki-way. Flexible. Some info structured, some not
  - Documentation
  - Standard Operation Procedures (SOP)
  - Informal instrument queue

## Queue [\[edit\]](#)

Sort	Request	User	Comment
1	160706_CCES	CCES (La Paz)	10 samples 60min_CID
2	160829_HPFG	HPFG	9 samples 60min_CID 45%
3	QCEU	QCEU	1 samples 30min_CID
4	160908_TFLS	TFLS	11 samples

[Commit](#)

# **Biocore Wiki**

*Task management system*

**Bioinformatics Unit, CRG**

# Biocore Wiki

- SERVICES
- TRAINING
- PROJECTS
- ABOUT US ▾
- WEBSITE ▾
- BIOCORE ▾
- MANAGEMENT ▾
- PAGE TOOLS ▾



Submit a request



## Last Request

Title	Customer	PI	Assumed by	Status	Paid	Order no	Creation Date	Last Modification
[blurred]	[blurred]	[blurred]	[blurred]	Pending			2015-10-21	2015-10-21
[blurred]	[blurred]	[blurred]	Lcozzuto	Accepted			2015-10-15	2015-10-15
[blurred]	[blurred]	[blurred]	Sbonnin	Accepted	No		2015-10-05	2015-10-09
[blurred]	[blurred]	[blurred]	Lcozzuto	Accepted			2015-09-21	2015-09-21
[blurred]	[blurred]	[blurred]	Lcozzuto and Theroso	Accepted			2015-09-04	2015-09-04
[blurred]	[blurred]	[blurred]	Theroso	Accepted			2015-08-21	2015-08-21
[blurred]	[blurred]	[blurred]	Sbonnin	Closed	All	7154630	2015-08-18	2015-09-25

# Biocore Wiki







## Task input

Task

<b>Title:</b>	<input type="text" value="Import of data and discussion"/>
<b>Responsible:</b>	<input type="text" value="Thermoso"/>
<b>Request Reference:</b>	<input type="text" value="20160118134502"/>
<b>Status:</b>	<input type="text" value="Finished"/>
<b>Description:</b>	<input type="text" value="Import of new data and different discussion"/>
<b>Start date:</b>	<input type="text" value="07/01/2016"/> 
<b>End date:</b>	<input type="text" value="15/01/2016"/> 

**IMPORTANT:** Do not add different hours of the SAME kind, sum them up yourself. If you think you need it, it should rather be another task.

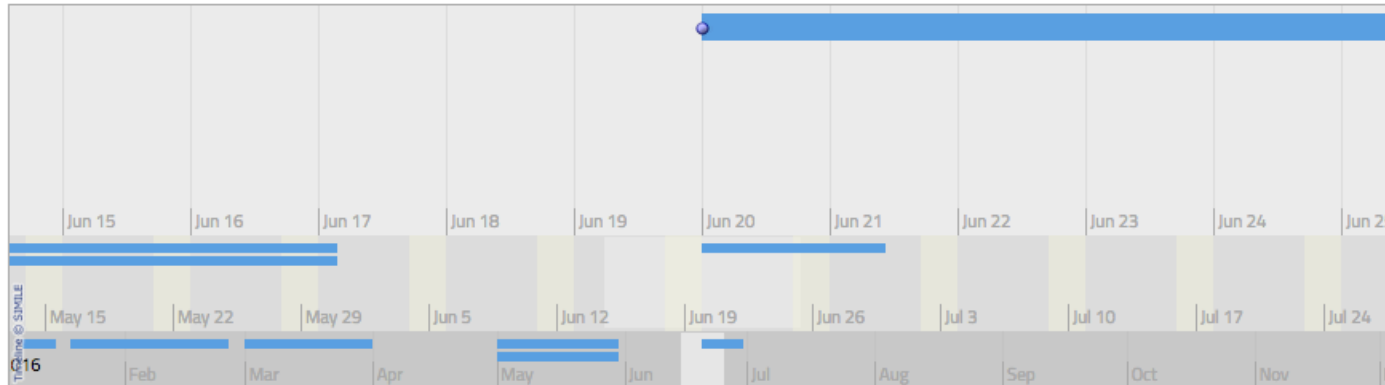
Hours

<b>Hours:</b>	<input type="text" value="2"/>	
<b>Kind:</b>	<input type="text" value="Internal Manual Data Analysis"/>	  
<b>Hours:</b>	<input type="text" value="8"/>	
<b>Kind:</b>	<input type="text" value="Internal Programming/Database development"/>	  

# Biocore Wiki

## Task view

Associated tasks [\[edit\]](#)



Task	Responsible	Start Date	End Date	Status	Details	Description						
Import of data and discussion	User:Thermoso	7 January 2016	15 January 2016	Finished	<table border="1"> <tr> <td>Internal Programming/Database development</td> <td>8</td> </tr> <tr> <td>Internal Automated Data Analysis</td> <td>10</td> </tr> <tr> <td>Internal Manual Data Analysis</td> <td>2</td> </tr> </table>	Internal Programming/Database development	8	Internal Automated Data Analysis	10	Internal Manual Data Analysis	2	Import of new data and different discussion
Internal Programming/Database development	8											
Internal Automated Data Analysis	10											
Internal Manual Data Analysis	2											
Importing data and discussion	User:Thermoso	19 January 2016	26 February 2016	Finished	<table border="1"> <tr> <td>Internal Programming/Database development</td> <td>2</td> </tr> </table>	Internal Programming/Database development	2	More importing data and mail discussion about contents				
Internal Programming/Database development	2											

# Biocore Wiki

## Hour & costs list

Finished Hours [\[edit\]](#)

Type	Total Hours	Price
Manual Data Analysis	0	0 EUR
Automated Data Analysis	0	0 EUR
Programming/Database development	0	0 EUR
Database maintenance support	0	0 EUR
<b>Total (for customers)</b>	<b>0</b>	<b>0 EUR</b>
Internal Manual Data Analysis	18	252.72 EUR
Internal Automated Data Analysis	12	33.72 EUR
Internal Programming/Database development	50	702 EUR
Internal Database maintenance support	2	28.08 EUR
<b>Total (internal)</b>	<b>82</b>	<b>1016.52 EUR</b>
One Cent - Manual Data Analysis	0	0 EUR
One Cent - Automated Data Analysis	0	0 EUR
One Cent - Programming/Database development	0	0 EUR
One Cent - Database maintenance support	0	0 EUR
<b>Total (one cent)</b>	<b>0</b>	<b>0 EUR</b>
<b>Total (internal + customers + one cent)</b>	<b>82</b>	<b>1016.52 EUR</b>

# **Example of biological data Content Management System (CMS)**

**VastDB, Manuel Irimia's lab (CRG)**

# Biological data CMS

## VastDB



VastDB

An atlas of alternative splicing profiles  
in vertebrate cell and tissue types

### Search

By gene

Human

Search

By coordinates

Mouse (mm9)

Search

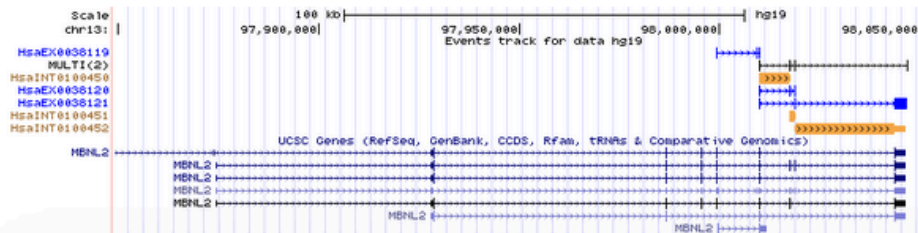
Human (hg19)

Human (hg38)

Mouse (mm9)

Mouse (mm10)

Chicken (galGal3)



### News

- **22/06/2016:** [vast-tools v1.1.0](#) has been released!
- **18/12/2014:** [vast-tools](#) paper has been published ([link](#))

[New version of VastDB](#)

Published on 12 August 2016, 17:22 by

WikiSysop – [lorem](#)

Lorem Ipsum is simply dummy text of

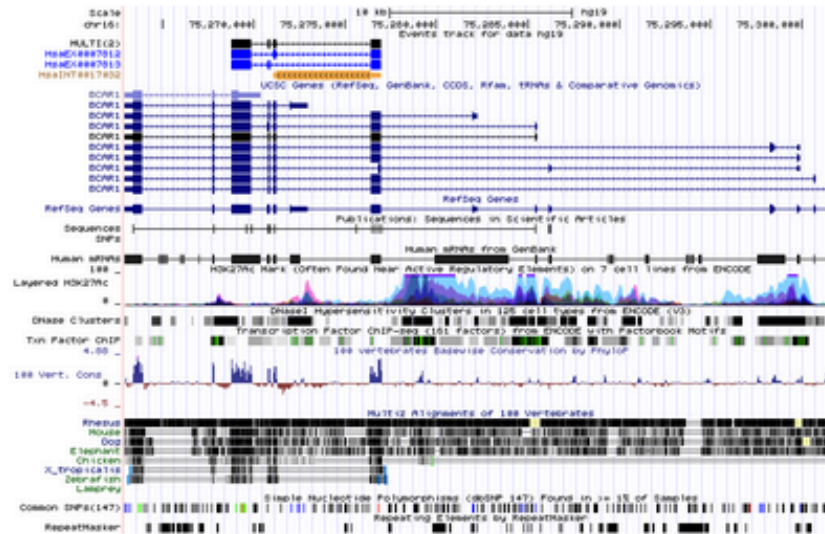
# Biological data CMS

## VastDB

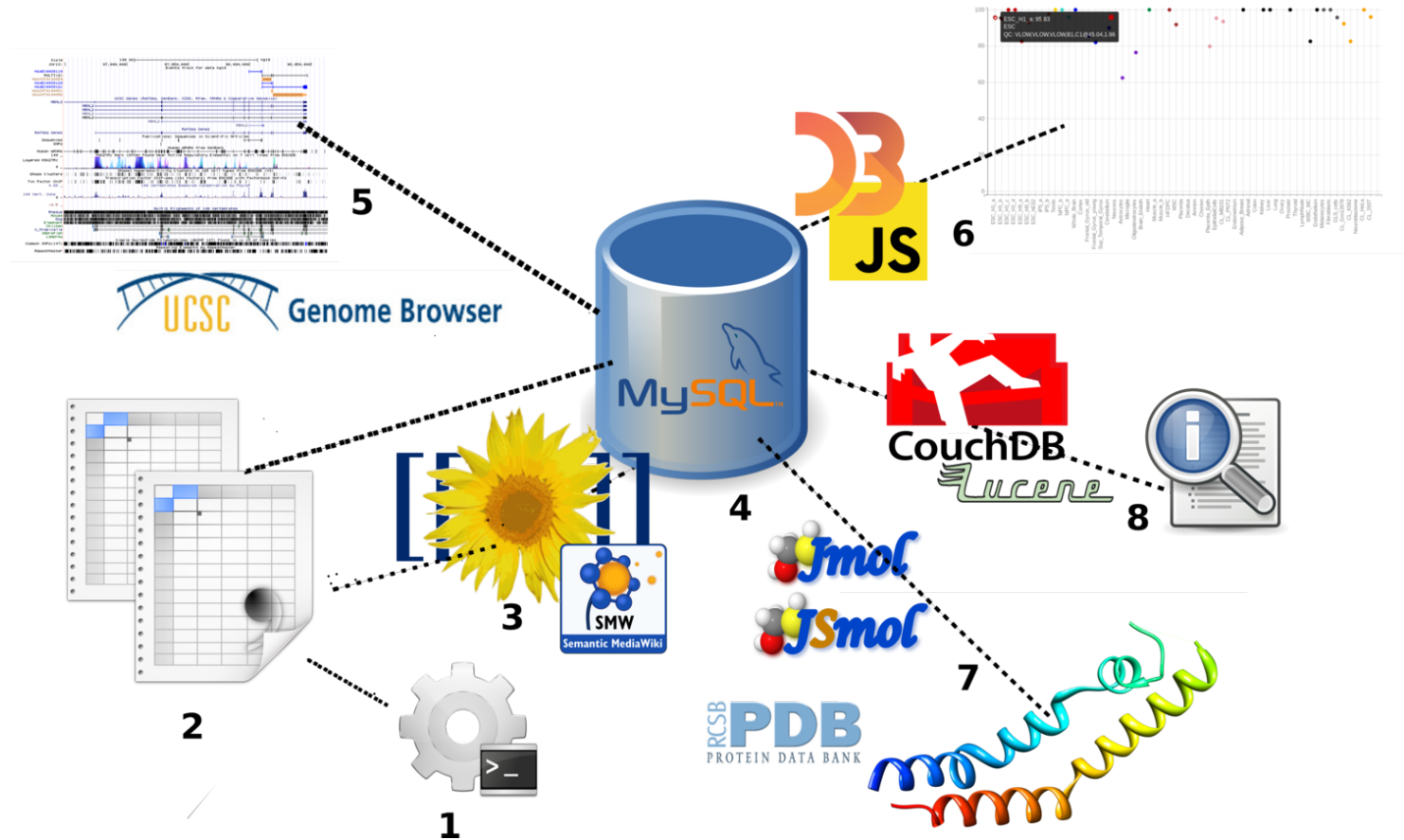
ENSG00000050820

Biotype	protein_coding
Gene Symbol	BCAR1
Description	breast cancer anti-estrogen resistance 1 [Source:HGNC Symbol;Acc:971]
Coordinates	chr16:75262929-75301951-
Assembly	hg19

### GENOMIC BROWSER



# VastDB overview



# Different data handling in MediaWiki as a CMS

- User import via specific extensions
- Using modified External data extension
- Extensions accessing file system
  - Mirror of PDB structures

# Semantic Data Import

## Data from CSV input

Expression

**Gene:ENSG00000050820@Genome:hg19**

Tissue	CRPKMs	Rawcount
Adipose Breast@Genome:hg19	9.77	3612
Adrenal@Genome:hg19	15.28	2369
Amnion@Genome:hg19	78.16	3866
Whole Brain@Genome:hg19	8.46	1428
Cortex@Genome:hg19	20.96	21864
Frontal Gyrus old@Genome:hg19	12.55	5864
Frontal Gyrus young@Genome:hg19	7.75	3112
Sup Temporal Gyrus@Genome:hg19	23.15	10542
Cerebellum@Genome:hg19	63.58	20821
Chorion@Genome:hg19	50.12	3607
CL 293T@Genome:hg19	16.06	2634
CL Gm12878@Genome:hg19	4.81	1893
CL HeLa@Genome:hg19	23.80	5082
CL K562@Genome:hg19	11.40	5217
CL MB231@Genome:hg19	59.78	18117
Neuroblastome@Genome:hg19	11.48	930
CL PNT2@Genome:hg19	19.49	2824
Colon@Genome:hg19	8.38	1688
Decidua@Genome:hg19	63.59	3209
Endometrium@Genome:hg19	25.63	4240

Output view handled with  
[handsontable.com](https://handsontable.com)

# Semantic Data Import

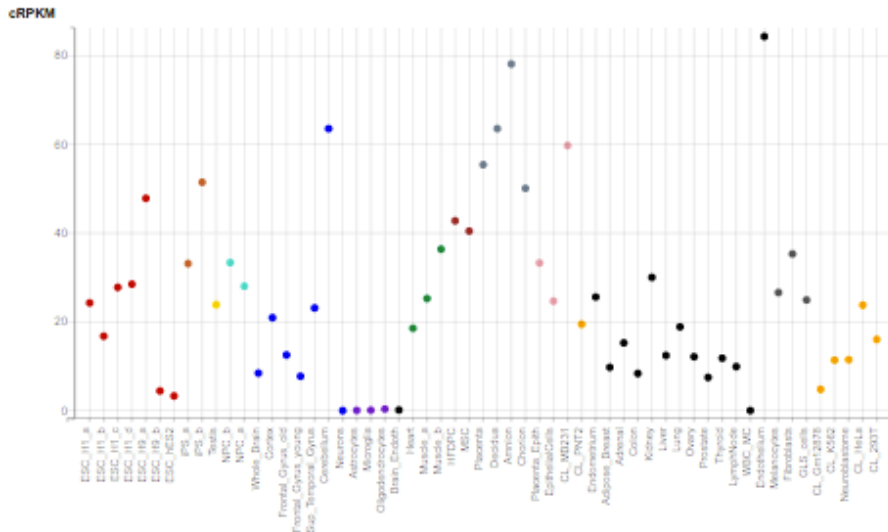
## EXPRESSION LEVELS

[Show All] [Show None] [Select All] [Select None]

CLines Cells  Fibroblasts  GLS\_cells  Melanocytes  ESC

Epithelial  CL\_MB231  EpithelialCells  Placenta\_Epith  Glia Muscle NPC Neural PREC Placenta

Testis Tissue iPSC



Output view  
handled with  
**Rickshaw** (D3.js)

[Export as image...](#)

### Associated events

Event	Type	Length	Coordinates	Strand
HsaEX0007812	Exon Skipping	162	chr16:75,268,787-75,276,988	-
HsaEX0007813	Exon Skipping	117	chr16:75,268,787-75,276,988	-

# CouchDB + Lucene

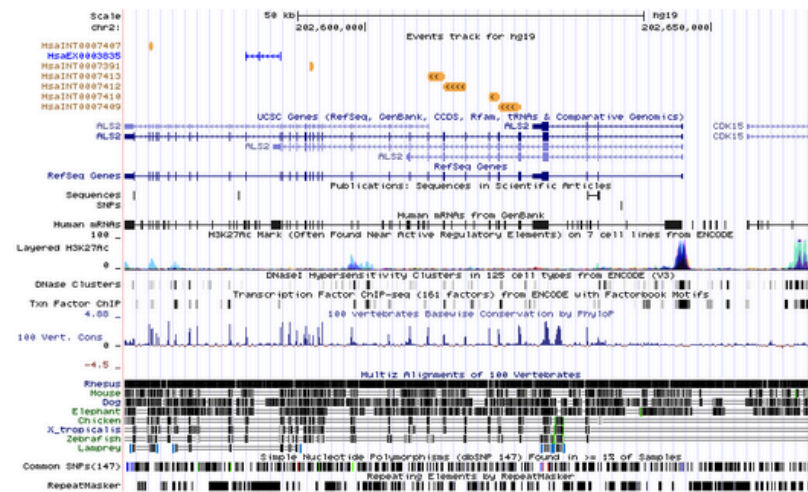
## Making search faster

- CouchDB: NoSQL Document DBMS
- Lucene: Information retrieve library.  
ElasticSearch or Solr based on it
- Mapping SMW Templates to JSON documents
- Indexing for coordinates and full-text search
- It might be ported to ElasticSearch

# CouchDB + Lucene

## Coordinate search

### Coordinates



chr2:202565277-202665277

Genome:hg19


9

Id	Chr	Start	End	Assembly
<a href="#">ENSG0000003393</a>	chr2	202565277	202645912	hg19
<a href="#">HsaEX003835</a>	chr2	202582800	202587843	hg19
<a href="#">HsaINT007391</a>	chr2	202591935	202592498	hg19

# CouchDB + Lucene

## Full-text search

VastDB

MENU 

### Genes

60 [Next](#)

Gene id	Gene symbol	Assembly
<a href="#">Gene:ENSG00000164304@Genome:hg19</a>	CAGE1	<a href="#">Genome:hg19</a>
<a href="#">Gene:ENSG00000169635@Genome:hg19</a>	HIC2	<a href="#">Genome:hg19</a>
<a href="#">Gene:ENSG00000177374@Genome:hg19</a>	HIC1	<a href="#">Genome:hg19</a>
<a href="#">Gene:ENSG00000008226@Genome:hg19</a>	DLEC1	<a href="#">Genome:hg19</a>
<a href="#">Gene:ENSG00000050820@Genome:hg19</a>	BCAR1	<a href="#">Genome:hg19</a>
<a href="#">Gene:ENSG00000054282@Genome:hg19</a>	SDCCAG8	<a href="#">Genome:hg19</a>
<a href="#">Gene:ENSG00000012048@Genome:hg19</a>	BRCA1	<a href="#">Genome:hg19</a>

# Genome Annotation Wiki framework

AnnoWiki

# Genome Annotation

## AnnoWiki

### Edit Gene: Gene:MPN009



**Id:**

**Genome\_id:**

**Coding:**

**Name:**

**Start:**

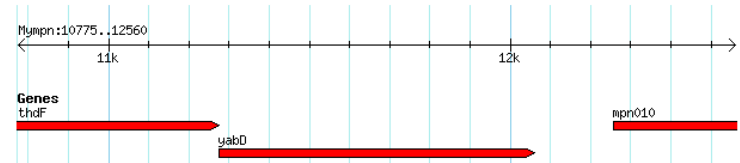
**End:**

**Strand:**

**Sequence:**

<b>coding</b>	0
<b>name</b>	yabD
<b>start</b>	11275
<b>end</b>	12060
<b>strand</b>	+
<b>genome_id</b>	<a href="#">Mympn</a>
<b>Sequence</b>	<pre> ATGGAATACTTTGATGCGCACTGTCACCTAAACTGTGAACCCCTACTGGAACAGCAGCAA AAGAGTTTGGCCAACCTTAGGTTAATTGGGTTAAAGGCCAATGTTGTGGGCACTAATTTA ACTAACTCCCAAATTGCGGTGAATTTAGCTAAACAGCACCCCTGATTACTCAAGGCAGGG GTGGGTATTCACCCTAATGATGTCCAGCTTTTTGATTTAAAGCAGCGCAAGCAACACTA AAAAAATTAGTAAGTACGCATCGCTCATTCAATGTTGATTGGTGAATATGGCTTTGAT TACCACTACACTAAGGACTACATTACCAACAAGACAGTTCTTTTTGATGCAGTTCCAA CTAGCAGAACAGTACCAGTTAGTGCACATGTTGCATGTCCTGTGATGTGCACGAAAGGATA TACGAAGTTTGAAGGTTAAAACTAAACAACAGTGGTGTTCCTACTGTTTTAGTGAA GACCCAATACTGTTTTAAAGCTGTAACCTTAAGAGAAGTAGGACTAAAGGTTATTTTT TCGATTCGGGGATTGTGACCTTTAAGAATGCGAAAACTGCAAGCCGCTTTATCAGTT ATTC AACAGAGCTACTTTTTAAGTGAACCGATTGCGCTTATTTAGCACCAGTACCTTTT CGGGCAAAAATAATTGACCGAATGCGTTGTTATACAGTGCAACGATTGCTGACATT AAGCAAGTTCCTTTAGCTGAAATAAAACAAGCGATTGTGCATAACGCAAGAAGTATTT TGGTAG </pre>

### GENOMIC BROWSER



# Import and export formats

- FASTA files (sequences)
- GFF or GTF (feature, relationship, location)
- Others: chromosome sizes, etc.
- Raw text files
- When convenient external tools:
  - NCBI-Blast
  - SAMTools
  - etc.

# Import and export formats

Special

## Wikify your biological data

Create a new genome

Upload a file

### Upload a biological data file

Upload

Navega... No s'ha seleccionat cap fitxer.

File type

Genes (GFF)

Genome

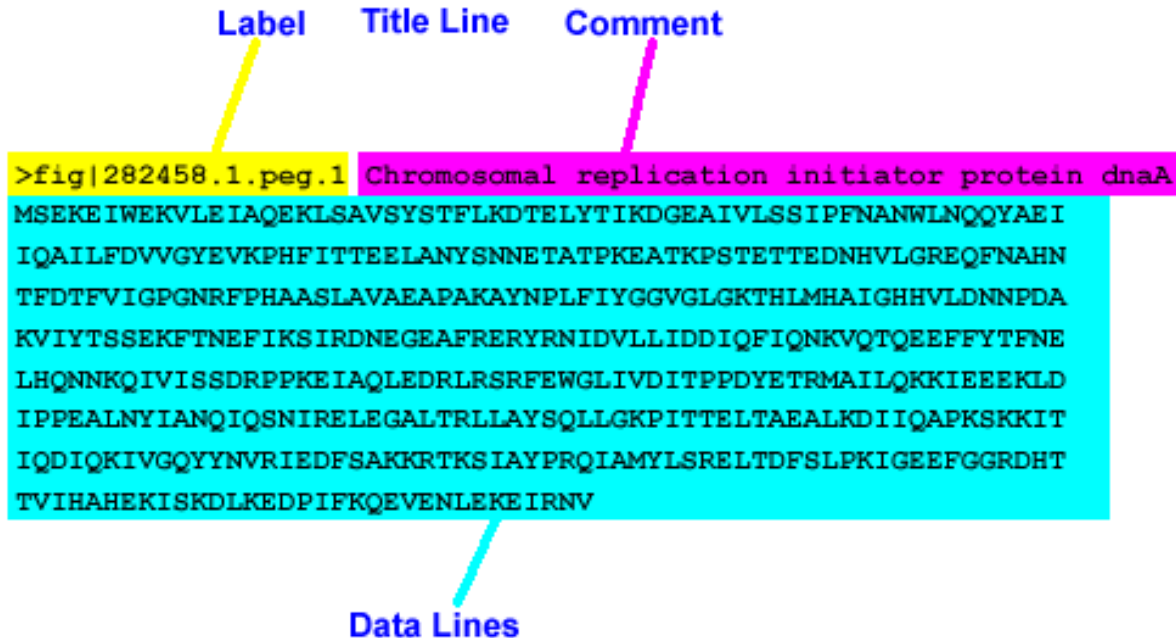
Genome:hg19

Upload and process

Undo changes

# Import and export formats

## FASTA



<http://www.nmpdr.org/FIG/wiki/view.cgi/FIG/FastaFormat>

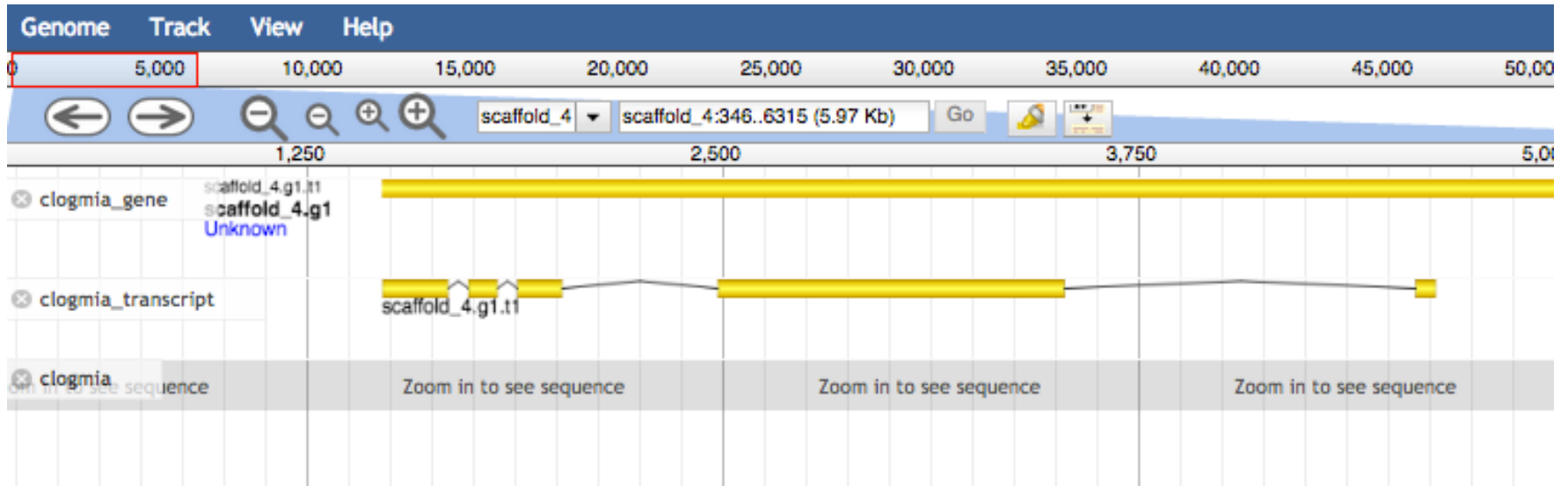
# Import and export formats

## GFF

```
##gff-version 3
##sequence-region ctg123 1 1497228
ctg123 . gene 1000 9000 . + . ID=gene00001;Name=EDEN
ctg123 . TF_binding_site 1000 1012 . + . ID=tfbs00001;Parent=gene00001
ctg123 . mRNA 1050 9000 . + . ID=mRNA00001;Parent=gene00001;Name=EDEN.1
```

[https://bioinf.comav.upv.es/courses/sequence\\_analysis/snp\\_calling.html](https://bioinf.comav.upv.es/courses/sequence_analysis/snp_calling.html)

# Integrating a genome browser



JBrowse

# Integrating a genome browser

Gene:scaffold 4.g1@Genome:clogmia

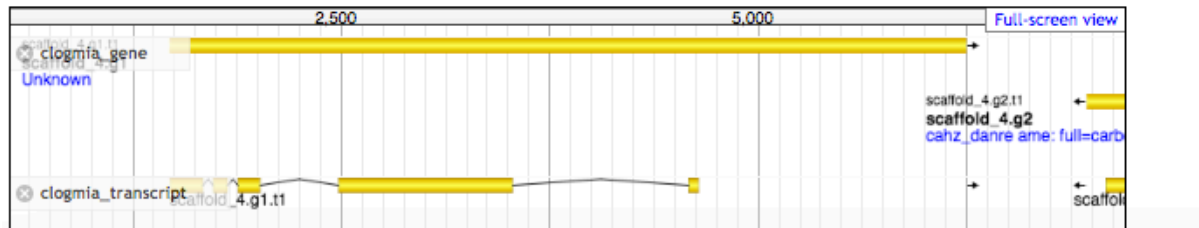
## Description

### type

name	Unnamed
location	scaffold_4
start	1477
end	6251
strand	+
genome_id	Genome:clogmia
id	scaffold_4.g1@Genome:clogmia
ref_id	Contigs:scaffold_4@Genome:clogmia
source	AUGUSTUS

## Children [\[edit\]](#)

Transcript:scaffold 4.g1.t1@Genome:clogmia



# Linking pages, conceptual hierarchies

- By using specific properties
- SMWParent extension
  - Quick retrieval of linked elements
    - Parent, ancestors
    - Children, descendants
    - Number of hops
    - Filter by another property value

# Linking pages, conceptual hierarchies

end	6251
strand	+
genome_id	<a href="#">Genome:clogmia</a>
id	<a href="#">scaffold_4.g1.t1@Genome:clogmia</a>
ref_id	<a href="#">Gene:scaffold_4.g1@Genome:clogmia</a>

## Parent [\[edit\]](#)

[Gene:scaffold 4.g1@Genome:clogmia](#)

## Children [\[edit\]](#)

[Exon:scaffold 4.g1.t1.exon.1477@Genome:clogmia](#),[Exon:scaffold 4.g1.t1.exon.1739@Genome:clogmia](#),[Exon:scaffold 4.g1.t1.exon.1887@Genome:clogmia](#),[Exon:scaffold 4.g1.t1.exon.2491@Genome:clogmia](#),[Exon:scaffold 4.g1.t1.exon.4587@Genome:clogmia](#)

Category: [Transcript](#)

# Acknowledgements



## **Biocore Wiki**

Carlos Company  
Julia Ponomarenko  
Luca Cozzuto  
Sarah Bonnin  
Guglielmo Roma  
et al.

## **ProteoWiki**

Eduard Sabidó  
Francesco Mancuso  
Cristina Chiva  
Eva Borràs  
Guadalupe Espadas  
et al.

## **VastDB**

Manuel Irimia  
Javier Tapial  
Luca Cozzuto

## **AnnoWiki**

Luca Cozzuto  
Carlos Company

... and all involved open-source community

# Questions?

[@toniher](#)