

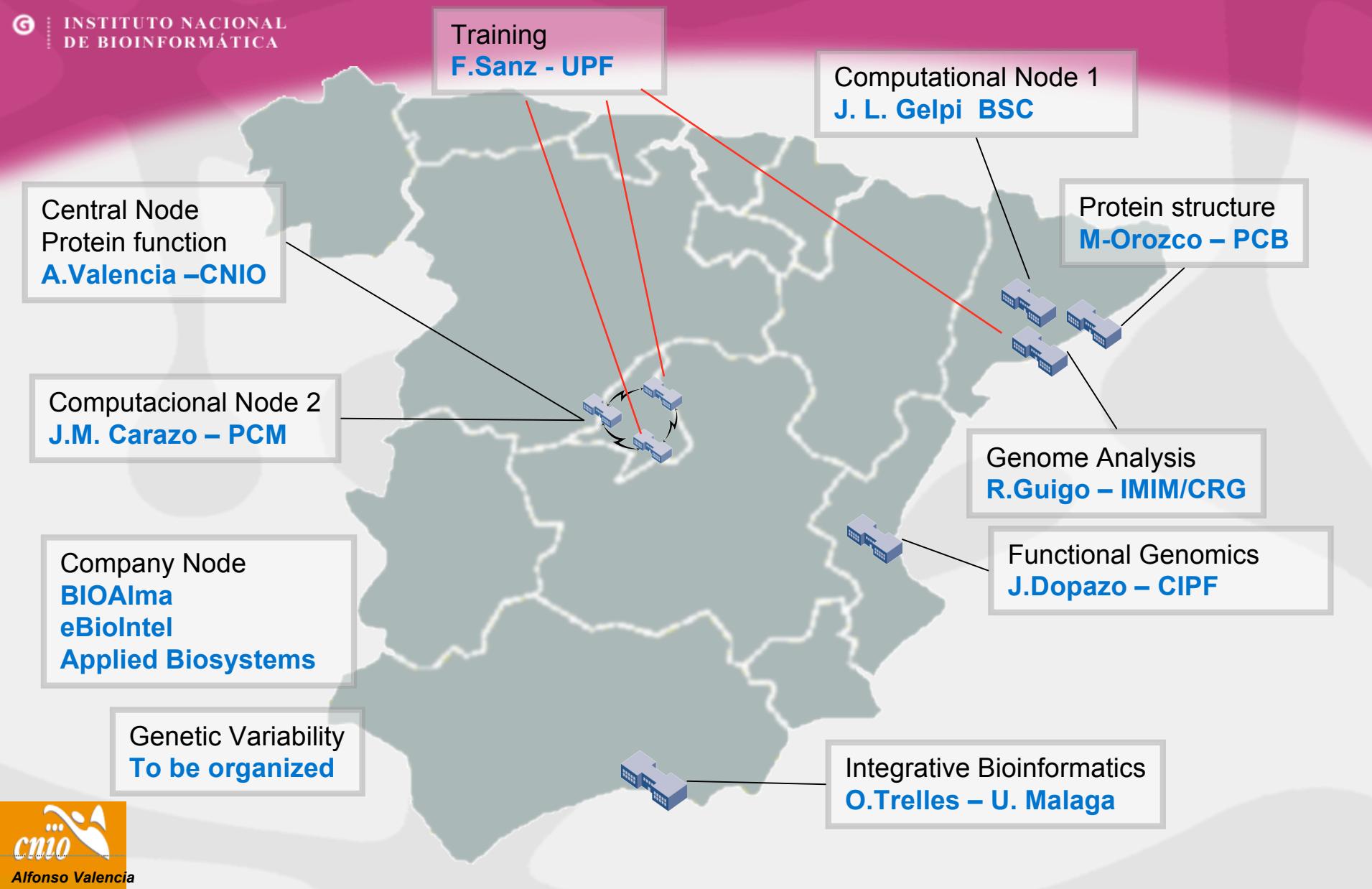
# DAS related developments

Alfonso Valencia

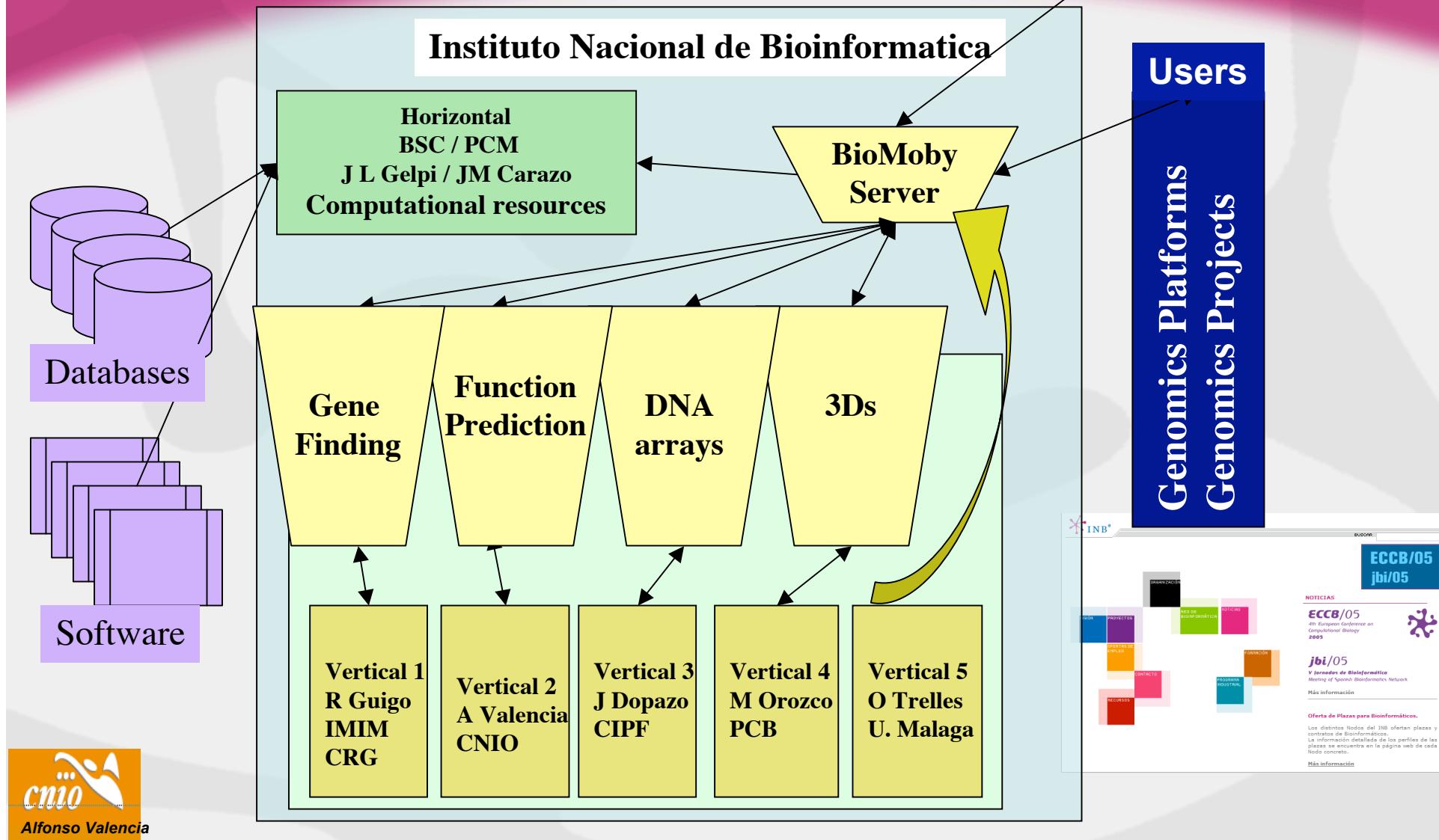
Structural and Computational Biology Programme  
Spanish National Cancer Research Centre  
CNIO, Madrid

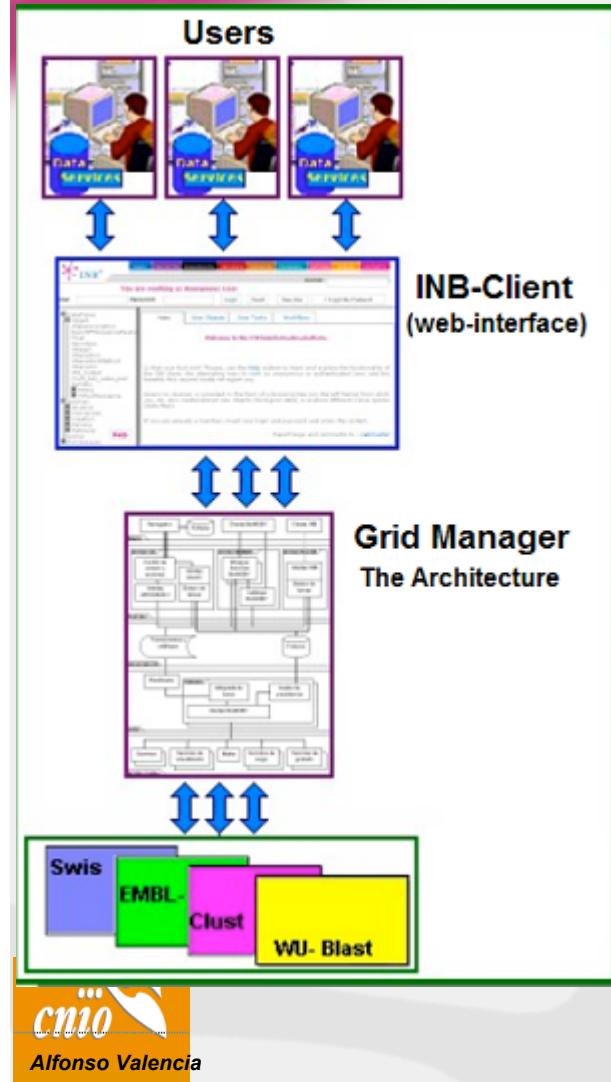
Sanger  
Feb. 2007

# A Virtual Institute

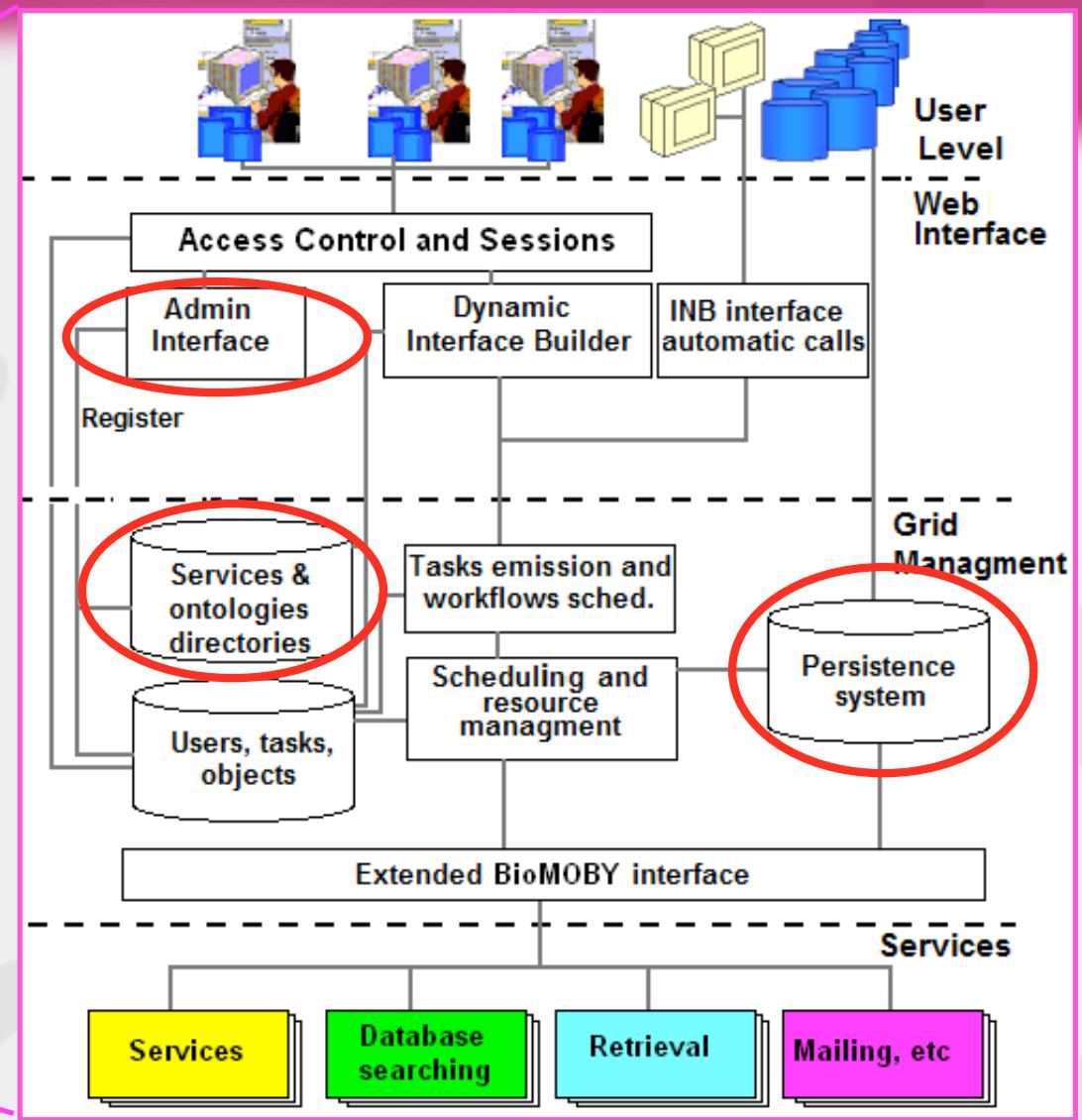


# INB initial resource structure





# INB Moby architecture and Client. > than 200 webservices under control

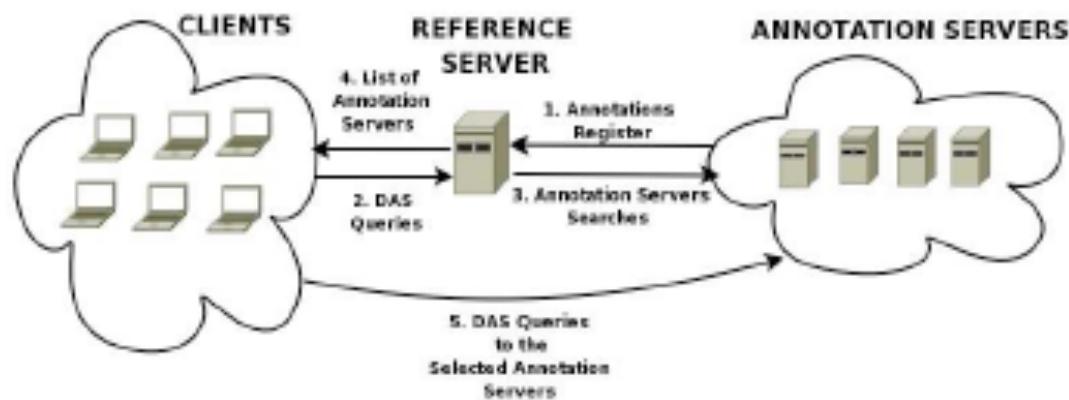


Oswaldo Trelles, UMalaga

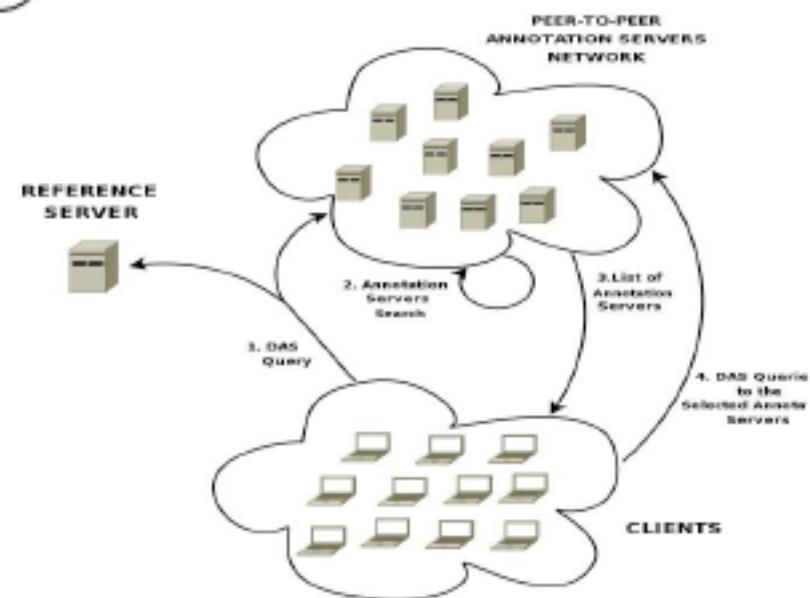
# PEER-TO-PEER SOLUTIONS ON BIOINFORMATICS APPLICATIONS

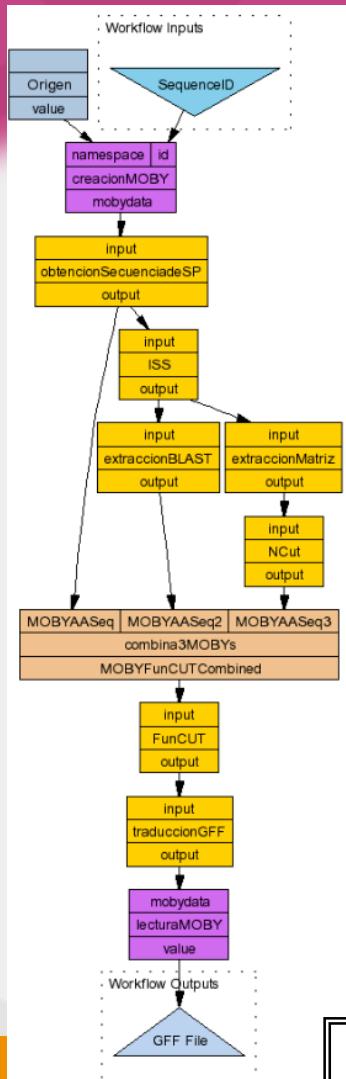
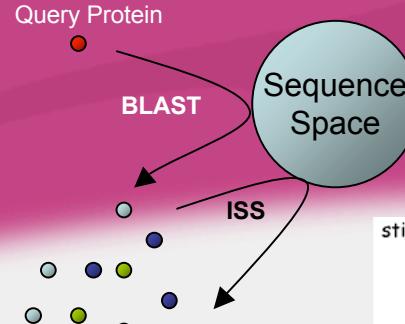
C. Guerrero, R. Cuevas, I. Martinez, A. Azcorra  
Universidad Carlos III Madrid

“ .... possible advantages derived from the usage of peer-to-peer (p2p) technologies on DAS environments”

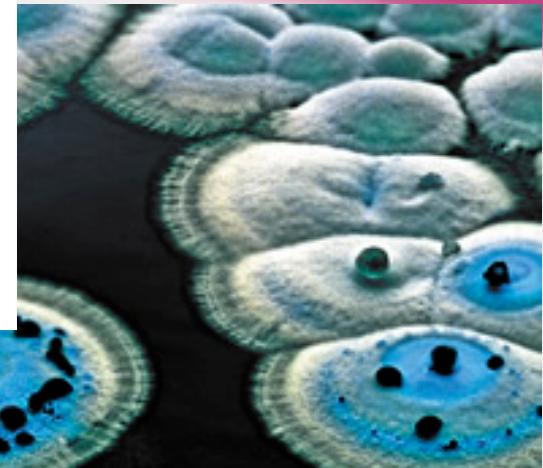
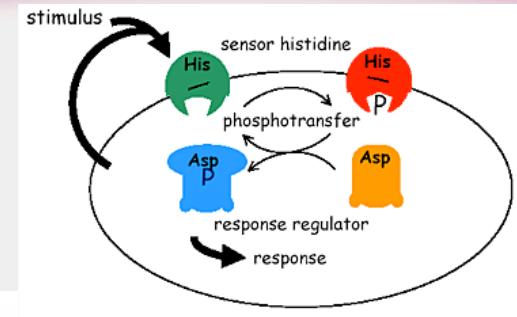
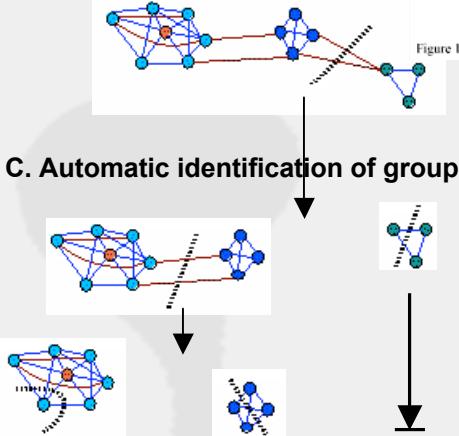
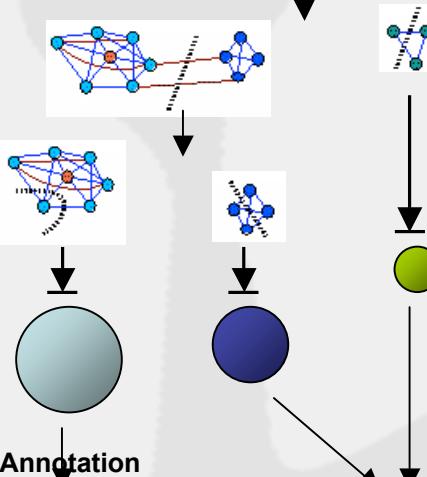
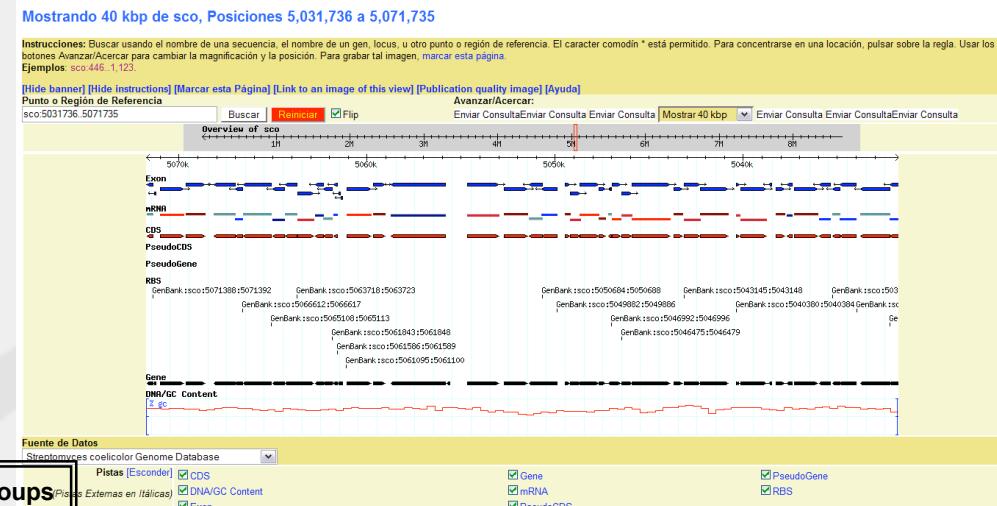


Current versus mix model

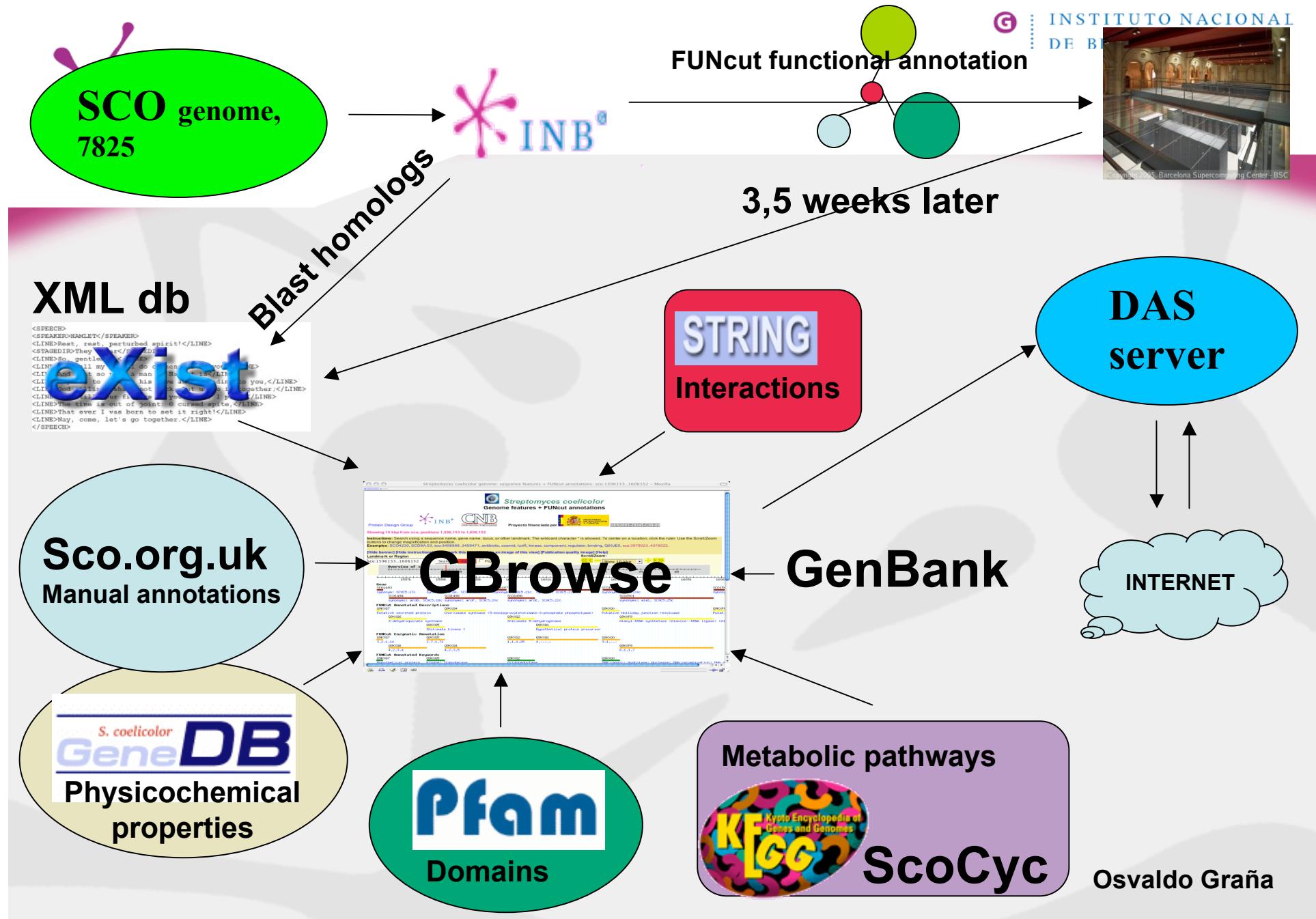



**A. Homology Search**


# Two component system *Streptomyces coelicolor* JF Martin (NV2)


**B. Sequence Space Mapping**

**C. Automatic identification of groups. Clustering.**

**Streptomyces coelicolor Genome Database**

**D. Annotation**  
**Problem Protein Annotation**
**Neighbor Groups Annotation**

# Streptomyces coelicolor computational workflow





# SynBrowse alignment viewer

Comparison between scoSAV and SAVsco

[http://gbrowse.bioinfo.cni.es/cgi-bin/synbrowse?config=synbrowse&Search=1&name=sco:1..22345&qry\\_species=scoSAV&ref\\_species=SAVsco&name=sc](http://gbrowse.bioinfo.cni.es/cgi-bin/synbrowse?config=synbrowse&Search=1&name=sco:1..22345&qry_species=scoSAV&ref_species=SAVsco&name=sc)

Query Species: scoSAV

Refer. Species: SAVsco

Search for: sco:1..22345

Align scoSAV: sco

to SAVsco: sav

Options: Realign Recenter

User Guides

Chromosome: All

Search | Reset

Comparative Alignment | DNA Align | Display Mode | Synteny Region

Search | Reset

Comparative Alignment | DNA Align | Display Mode | Synteny Region

Search | Reset

Comparative Alignment | DNA Align | Display Mode | Synteny Region

Search | Reset

Comparative Alignment | DNA Align | Display Mode | Synteny Region

Tracks:  DNA Align  Gene  Repeat

Streptomyces coelicolor genome alignments

Structural Computational Biology group

INB Centro Nacional de Investigaciones Oncológicas

Project funded by: GEN2003-20245-C09-09

Nucleotide Alignment

sco:1..22345

Gene: SC00001, SC00003, SC00005, SC00007, SC00011, SC00015, SC00019, SC00022, SC00021, SC00020, SC00016, SC00012, SC00014, SC00018, SC00017, SC00013, SC00019, SC00010, SC00009, SC00008, SC00006, SC00005, SC00004, SC00003, SC00001, SC00002, SC00006, SC00007, SC00008, SC00009, SC00010, SC00011, SC00012, SC00013, SC00014, SC00015, SC00016, SC00017, SC00018, SC00019, SC00020, SC00021, SC00022.

Gene: SCJ30.02c, SCJ30.04c, SCJ30.06c, SCJ30.08c, SCJ30.09c, SCJ30.11c, SCJ30.12c, SCJ30.14c, SCJ30.01c, SCJ30.03c, SCJ30.05c, SCJ30.07c, SCJ30.09c, SCJ30.11c, SCJ30.13c, SCJ30.15c, SCJ30.17c, SCJ30.19c, SCJ30.21c, SCJ30.23c, SCJ30.25c, SCJ30.27c, SCJ30.29c, SCJ30.31c, SCJ30.33c, SCJ30.35c, SCJ30.37c, SCJ30.39c, SCJ30.41c, SCJ30.43c, SCJ30.45c, SCJ30.47c, SCJ30.49c, SCJ30.51c, SCJ30.53c, SCJ30.55c, SCJ30.57c, SCJ30.59c, SCJ30.61c, SCJ30.63c, SCJ30.65c, SCJ30.67c, SCJ30.69c, SCJ30.71c, SCJ30.73c, SCJ30.75c, SCJ30.77c, SCJ30.79c, SCJ30.81c, SCJ30.83c, SCJ30.85c, SCJ30.87c, SCJ30.89c, SCJ30.91c, SCJ30.93c, SCJ30.95c, SCJ30.97c, SCJ30.99c, SCJ30.01c, SCJ30.03c, SCJ30.05c, SCJ30.07c, SCJ30.09c, SCJ30.11c, SCJ30.13c, SCJ30.15c, SCJ30.17c, SCJ30.19c, SCJ30.21c, SCJ30.23c, SCJ30.25c, SCJ30.27c, SCJ30.29c, SCJ30.31c, SCJ30.33c, SCJ30.35c, SCJ30.37c, SCJ30.39c, SCJ30.41c, SCJ30.43c, SCJ30.45c, SCJ30.47c, SCJ30.49c, SCJ30.51c, SCJ30.53c, SCJ30.55c, SCJ30.57c, SCJ30.59c, SCJ30.61c, SCJ30.63c, SCJ30.65c, SCJ30.67c, SCJ30.69c, SCJ30.71c, SCJ30.73c, SCJ30.75c, SCJ30.77c, SCJ30.79c, SCJ30.81c, SCJ30.83c, SCJ30.85c, SCJ30.87c, SCJ30.89c, SCJ30.91c, SCJ30.93c, SCJ30.95c, SCJ30.97c, SCJ30.99c.

Nucleotide Alignment

sav:1..29513

Gene: SC00001, SC00003, SC00005, SC00007, SC00011, SC00015, SC00019, SC00022, SC00021, SC00020, SC00016, SC00012, SC00014, SC00018, SC00017, SC00013, SC00019, SC00010, SC00009, SC00008, SC00006, SC00005, SC00004, SC00003, SC00001, SC00002, SC00006, SC00007, SC00008, SC00009, SC00010, SC00011, SC00012, SC00013, SC00014, SC00015, SC00016, SC00017, SC00018, SC00019, SC00020, SC00021, SC00022.

Gene: SCJ30.02c, SCJ30.04c, SCJ30.06c, SCJ30.08c, SCJ30.09c, SCJ30.11c, SCJ30.12c, SCJ30.14c, SCJ30.17c, SCJ30.19c, SCJ30.21c, SCJ30.23c, SCJ30.25c, SCJ30.27c, SCJ30.29c, SCJ30.31c, SCJ30.33c, SCJ30.35c, SCJ30.37c, SCJ30.39c, SCJ30.41c, SCJ30.43c, SCJ30.45c, SCJ30.47c, SCJ30.49c, SCJ30.51c, SCJ30.53c, SCJ30.55c, SCJ30.57c, SCJ30.59c, SCJ30.61c, SCJ30.63c, SCJ30.65c, SCJ30.67c, SCJ30.69c, SCJ30.71c, SCJ30.73c, SCJ30.75c, SCJ30.77c, SCJ30.79c, SCJ30.81c, SCJ30.83c, SCJ30.85c, SCJ30.87c, SCJ30.89c, SCJ30.91c, SCJ30.93c, SCJ30.95c, SCJ30.97c, SCJ30.99c.

Nucleotide Alignment

sco:1..22345

Gene: SC00001, SC00003, SC00005, SC00007, SC00011, SC00015, SC00019, SC00022, SC00021, SC00020, SC00016, SC00012, SC00014, SC00018, SC00017, SC00013, SC00019, SC00010, SC00009, SC00008, SC00006, SC00005, SC00004, SC00003, SC00001, SC00002, SC00006, SC00007, SC00008, SC00009, SC00010, SC00011, SC00012, SC00013, SC00014, SC00015, SC00016, SC00017, SC00018, SC00019, SC00020, SC00021, SC00022.

Gene: SCJ30.02c, SCJ30.04c, SCJ30.06c, SCJ30.08c, SCJ30.09c, SCJ30.11c, SCJ30.12c, SCJ30.14c, SCJ30.17c, SCJ30.19c, SCJ30.21c, SCJ30.23c, SCJ30.25c, SCJ30.27c, SCJ30.29c, SCJ30.31c, SCJ30.33c, SCJ30.35c, SCJ30.37c, SCJ30.39c, SCJ30.41c, SCJ30.43c, SCJ30.45c, SCJ30.47c, SCJ30.49c, SCJ30.51c, SCJ30.53c, SCJ30.55c, SCJ30.57c, SCJ30.59c, SCJ30.61c, SCJ30.63c, SCJ30.65c, SCJ30.67c, SCJ30.69c, SCJ30.71c, SCJ30.73c, SCJ30.75c, SCJ30.77c, SCJ30.79c, SCJ30.81c, SCJ30.83c, SCJ30.85c, SCJ30.87c, SCJ30.89c, SCJ30.91c, SCJ30.93c, SCJ30.95c, SCJ30.97c, SCJ30.99c.

Nucleotide Alignment

sav:1..29513

Gene: SC00001, SC00003, SC00005, SC00007, SC00011, SC00015, SC00019, SC00022, SC00021, SC00020, SC00016, SC00012, SC00014, SC00018, SC00017, SC00013, SC00019, SC00010, SC00009, SC00008, SC00006, SC00005, SC00004, SC00003, SC00001, SC00002, SC00006, SC00007, SC00008, SC00009, SC00010, SC00011, SC00012, SC00013, SC00014, SC00015, SC00016, SC00017, SC00018, SC00019, SC00020, SC00021, SC00022.

Gene: SCJ30.02c, SCJ30.04c, SCJ30.06c, SCJ30.08c, SCJ30.09c, SCJ30.11c, SCJ30.12c, SCJ30.14c, SCJ30.17c, SCJ30.19c, SCJ30.21c, SCJ30.23c, SCJ30.25c, SCJ30.27c, SCJ30.29c, SCJ30.31c, SCJ30.33c, SCJ30.35c, SCJ30.37c, SCJ30.39c, SCJ30.41c, SCJ30.43c, SCJ30.45c, SCJ30.47c, SCJ30.49c, SCJ30.51c, SCJ30.53c, SCJ30.55c, SCJ30.57c, SCJ30.59c, SCJ30.61c, SCJ30.63c, SCJ30.65c, SCJ30.67c, SCJ30.69c, SCJ30.71c, SCJ30.73c, SCJ30.75c, SCJ30.77c, SCJ30.79c, SCJ30.81c, SCJ30.83c, SCJ30.85c, SCJ30.87c, SCJ30.89c, SCJ30.91c, SCJ30.93c, SCJ30.95c, SCJ30.97c, SCJ30.99c.

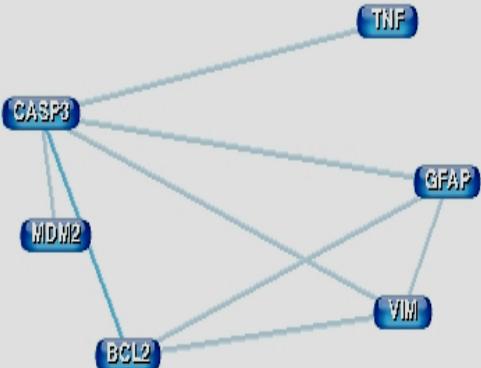
Osvaldo Graña

Search Gene

Clear model

Print version

Edges in the graph  
correspond to genes  
that concur in at least  
1, 2 or 3 phrases.



**MDM2** is cleaved by Caspase 3 (**CPP32**) during apoptosis after aspartic acid-361, generating a 60 kd fragment.

These findings indicate that IR-induced apoptosis involves activation of **CPP32** and that this CrmA-insensitive apoptotic pathway is distinct from those induced by **TNF** and certain other stimuli.

At the end of the experiment, the whole CNS apoptotic markers (BAX, **BCL2** and **CPP32**) a

Fas-induced activation of the ce



## VISIT: iHOP



Synonyms  
CAT1, CCR1, GLC2, HAF3,  
PAS4

Organism  
Saccharomyces cerevisiae



NCBI Protein NP\_010765

The **snf1** mutation also suppresses the glucose repression defects of **reg1**.

The **SIP1** protein co-immunoprecipitated with **SNF1** and was phosphorylated in vitro.

Here we show that **Reg1** interacts with the **Snf1** catalytic domain in the two-hybrid system.

Previous studies showed that **Reg1** regulates the **Snf1** protein kinase in response to glucose.

The **SNF4** protein is physically associated with **SNF1** and positively affects the kinase activity.

The **Sip1** protein is known to undergo phosphorylation when associated in vitro with the **Snf1** protein kinase.

Genetic evidence indicated that the catalytic activity of **Snf1** negatively regulates its interaction with **Reg1**.

The **SNF1** protein kinase and the associated **SNF4** protein are required for release of glucose repression in *Saccharomyces cerevisiae*.

The **SIP1** gene of *Saccharomyces cerevisiae* is a carbon-catabolite-specific negative regulator of GAL gene transcription and acts as a multicopy suppressor of growth defects associated with impaired **Snf1p** protein kinase activity.

We show that different sequences of **Reg1** interact with **Glc7** and **Snf1**.

In two-hybrid assays, one **SNF4** mutation enhances the interaction between **Snf4** and **Snf1**.

Previously, we identified **SIP1** and **SIP2** as proteins that interact with **SNF1** in vivo by the two-hybrid system.

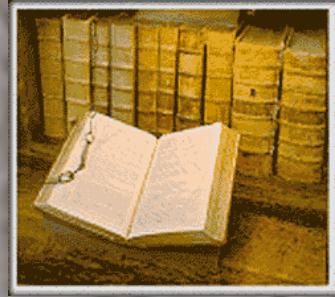
Previous experimental evidence had indicated that **Reg1** might target **Glc7** to nuclear substrates such as the **Snf1** kinase complex.

The catalytic subunits of Arabidopsis SnRKs, AKIN10 and AKIN11, interact with **Snf4** and suppress the **snf1** and **snf4** mutations in yeast.

**Pak1** associates with the **Snf1** kinase in vivo, and the association is greatly enhanced under glucose-limiting conditions when **Snf1** is active.

We show that **SNF4** binds to the **SNF1** regulatory domain in low glucose, whereas its kinase domain of **SNF1** itself

Hoffmann Valencia Nat Genet 2004





information hyperlinked  
over proteins

Search Gene

Show overview [new](#)  
Find in this Page

Filter and options

Gene Model

Help



## Symbol

MYC Myc proto-oncogene protein

## UniProt

P01106, P01106, P01107, Q14026

## IntAct

P01106

## PDB Structure

1A93, 1EE4, 1NKP

## OMIM

190080

## NCBI Gene

4609

## NCBI RefSeq

NP\_002458

## NCBI RefSeq

NM\_002467

## NCBI Accession

AAA36340, AAA59880, AAA59881

## Homologues of MYC ... [new](#)

Interaction information for this gene

Interaction information for this gene

Enhanced PubMed/Google query ... [new](#)

WARNING: Please keep in mind that gene detection is done automatically and can exhibit a certain error. [Read more](#).

c-Myc in [breast cancer](#).

[c-Myc: an [iron](#) oncogene].

Targeting c-myc in [leukemia](#).

[Radioimmunoassay](#) of c-myc protein.

c-myc implicated in [RNA](#) processing.

C-myc expression in [cervical cancer](#).

DNA binding by the Myc oncoproteins.

c-myc amplification in [ovarian cancer](#).

c-myc detection in [bone marrow](#) biopsies.

c-Myc as a therapeutic target in [cancer](#).

[Chromosome 8](#) includes the c-myc oncogene.

[Telomeres, telomerase](#), and myc. An update.

c-myc hypermutation in [Burkitt's lymphoma](#).

Immunohistochemical study of [EGF](#) and c-myc.

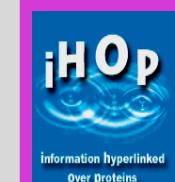
## Name

Synonyms

c-myc, c-Myc, Transcription factor p64, v-myc  
myelocytomatosis viral oncogene homolog (avian)

## Organism

Homo sapiens



Search Gene

Show only  
gene-verb-gene  
relationships in graph

Show official symbols

Save/ Load [new](#)

Send model [new](#)

Clear model

Filter and options

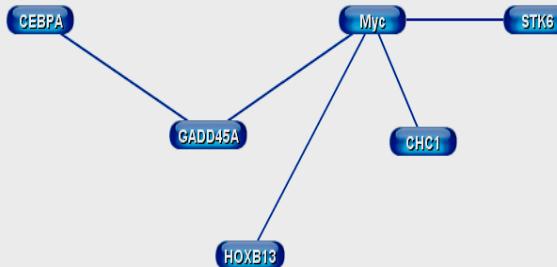
Help

Homologues

Definitions

Enhanced PubMed/Google query

Content & Implementation



[Redo graph layout]

Reciprocal regulation of [gadd45](#) [?] by [C/EBP alpha](#) [?] and [c-Myc](#) [?].

[C/EBPalpha](#) regulation of the growth-arrest-associated gene [gadd45](#).

[HOXB13](#) also suppressed the activity of natural [c-myc](#) promoter.

[c-myc](#) activates [RCC1](#) gene expression through E-box elements.

[c-Myc](#) in [breast cancer](#).

Quantitative reverse transcription polymerase chain reaction verified mRNA changes in this group, which included [gadd45a](#), [c-myc](#), cyclin D1 and [cdk4](#), [pcna](#), cyclin G, Rb, and [E2F5](#).

In epidermal [cells cultured](#) with [IFN-gamma](#) supplementation, we also show by RT-PCR that there is an upregulation of the genes [c-myc](#), [p53](#), [gadd45](#), dsRNA-activated protein kinase, and 2'-oligo(A)-dependent RNase, which have all been implicated in apoptosis in other cell types.

Ectopic expression of [Aurora-A](#) up-regulates [c-Myc](#).

[Myc](#) represses the growth arrest gene [gadd45](#).

[c-Myc](#) as a therapeutic target in [cancer](#).

If you find iHOP useful please cite as "Hoffmann, R., Valencia, A. A gene network for navigating the literature. *Nature Genetics* 36, 664 (2004)".

Find in this Page



Hoffmann Valencia Nat Genet 2004

Hoffmann Valencia Bioinformatics 2005

Fernandez-Gonzalez et al., submitted



Myc induces [TRRAP](#) recruitment and [histone](#) hyperacetylation at specific Myc-activated genes in vivo.

In addition, Myc recruited [TRRAP](#) to [chromatin](#), consistent with a role for this cofactor in [histone](#) acetylation.

The essential cofactor [TRRAP](#) recruits the histone acetyltransferase [HGCN5](#) to [c-Myc](#).

[MYC](#) associates with [TRRAP](#), a subunit of distinct macromolecular complexes that contain the HATs GCN5/PCAF or [TIP60](#).

[FIR](#) repressed a c-myc reporter via the FUSE.

[Myc](#) represses the growth arrest gene [gadd45](#).

[c-Raf](#) kinase binds to N-terminal domain of [c-Myc](#).

Ectopic expression of [Aurora-A](#) up-regulates [c-Myc](#).

Thus, [thrombospondin-1](#) is a bona fide target of [Myc](#).

[Epo](#) rapidly up-regulated [Myc](#) protein in BaF3-EpoR cells.

[Myc](#) represses transcription of the growth arrest gene [gas1](#).

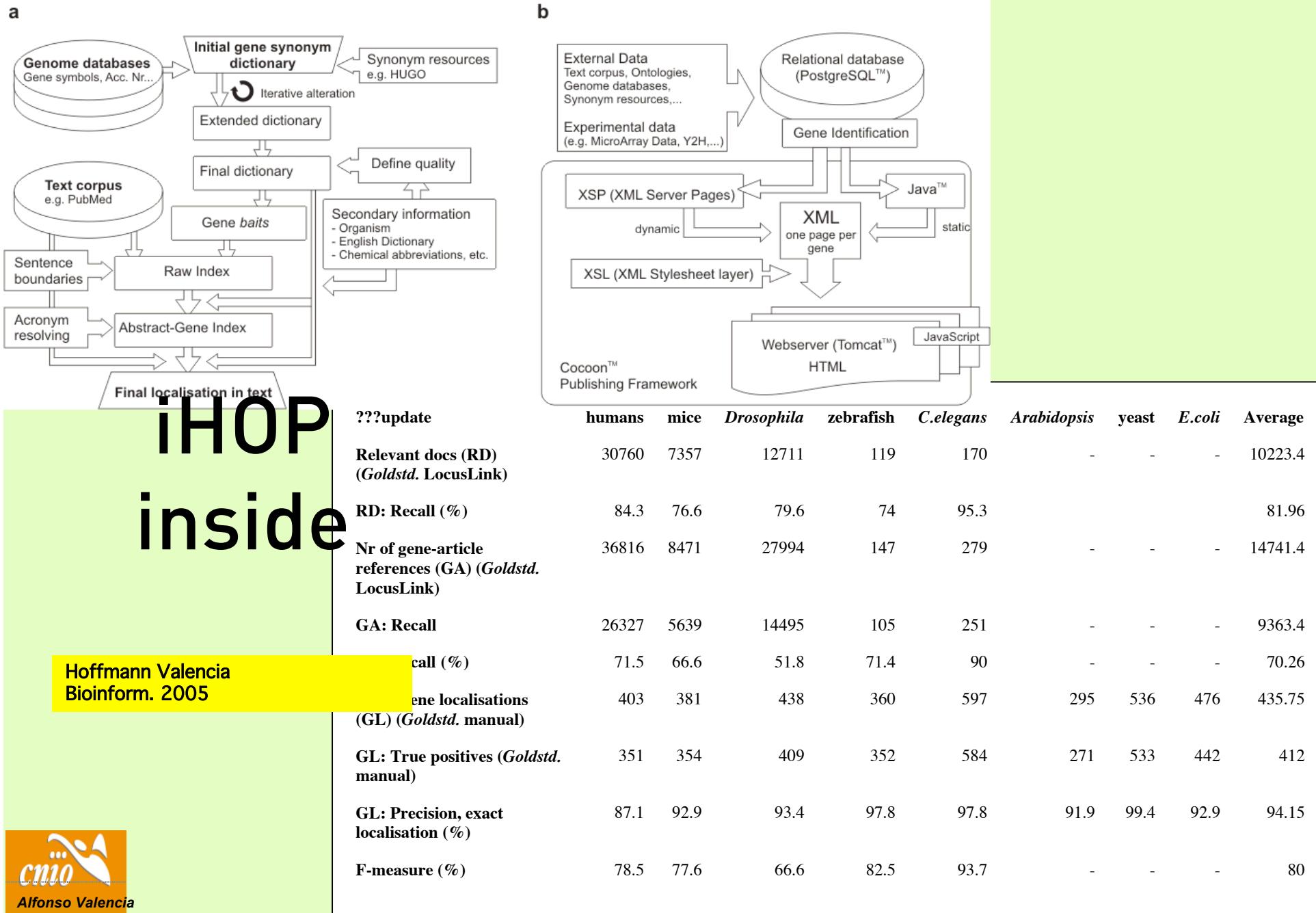
Dominant-negative [TAK1](#) induces [c-Myc](#) and G(0) exit in liver.

[c-myc](#) activates [RCC1](#) gene expression through E-box elements.

Apoptotic cell death induced by [c-myc](#) is inhibited by [bcl-2](#).

[GM-CSF](#) [?] had no effect on the half-life of [c-myc](#) messenger RNA.

**Figure 4**



```

<-<iHOPInteractionList>
- <iHOPsymbolInfo>
- <iHOPsymbol symbol="BRCA1" id="36781" organism="Homo sapiens" organismId="1" name="breast cancer 1, early onset">
  <synonym>BRCA1</synonym>
  <synonym>Breast cancer type 1 susceptibility protein</synonym>
  <synonym>IRIS</synonym>
  <synonym>PSCP</synonym>
  <synonym>RING finger protein 53</synonym>
  <synonym>RNF53</synonym>
  <homolog symbol="Bica1" id="119729" organism="Mus musculus"/>
  <dblink dbNamespace="NCBI_GENE_ID" dbId="672"/>
  <dblink dbNamespace="NCBI_GENE_ID" dbId="672"/>
  <dblink dbNamespace="NCBI_REFSEQ_TRANSCRIPT" dbId="NM_007296"/>
  <dblink dbNamespace="NCBI_REFSEQ_TRANSCRIPT" dbId="NM_007302"/>
  <dblink dbNamespace="INTACT_ID" dbId="P33398"/>
  <dblink dbNamespace="UNIPROT_AC" dbId="Q92897"/>
  <dblink dbNamespace="UNIPROT_AC" dbId="Q64FK3"/>
  <dblink dbNamespace="NCBI_GENBANK_AC" dbId="AAM18219"/>
  <dblink dbNamespace="NCBI_GENBANK_AC" dbId="BC062429"/>
  <dblink dbNamespace="UNIPROT_PRIMARY_AC" dbId="P33398"/>
  <dblink dbNamespace="NCBI_OMIM_MIM" dbId="114490"/>
  <dblink dbNamespace="NCBI_OMIM_MIM" dbId="113705"/>
  <dblink dbNamespace="PROTEIN_DATABANK_PDB_ID" dbId="1OQA"/>
  <dblink dbNamespace="PROTEIN_DATABANK_PDB_ID" dbId="1T15"/>
  <dblink dbNamespace="NCBI_REFSEQ_PROTEIN" dbId="NP_009237"/>
  <dblink dbNamespace="NCBI_REFSEQ_PROTEIN" dbId="NP_009236"/>
</iHOPsymbol>
+<evidenceList></evidenceList>
</iHOPsymbolInfo>
- <iHOPsentence sentenceNumber="4" pmid="960997" iHOPpmid="1471529" journalImpact="1.6850">
  Examining the entire coding sequences as well as exon-intron boundaries of both genes by
  - <iHOPatom>
    <MeSHLink term="Polymerase Chain Reaction" iHOPmeshId="233" meshId="E.393.620.500"/>
    polymerase chain reaction
  </iHOPatom>
  (
  - <iHOPatom>
    <MeSHLink term="Polymerase Chain Reaction" iHOPmeshId="233" meshId="E.393.620.500"/>
    PCR
  </iHOPatom>
  ) single-strand conformation polymorphism (SSCP) and multiplex-SSCP analysis, we identified possible disease-causing alterations in
  - <iHOPatom>
    <evidence symbol="BRCA1" id="36781" organism="Homo sapiens" organismId="1" confidence="37"/>
    BRCA1
  </iHOPatom>
  among

```

Symbol	Name	Synonyms	Organism
BRCA1	breast cancer 1, early onset	BRCA1, Breast cancer type 1 susceptibility protein, IRIS, PSCP, RING finger protein 53, RNF53	Homo sapiens

UniProt P38398, Q92897, Q64FK3  
 IntAct P38398  
 PDB Structure 1OQA, 1T15  
 OMIM 114480, 113705  
 NCBI Gene 672  
 NCBI RefSeq NP\_009237, NP\_009236  
 NCBI RefSeq NM\_007296, NM\_007302  
 NCBI UniGene 672  
 NCBI Accession AAM18219, BC062429

Homologues of BRCA1 ... new

Definitions for BRCA1 [i] ...

Enhanced PubMed/Google query ... new

WARNING: Please keep in mind that gene detection is done automatically and can exhibit a certain error. [Read more](#) about synonym ambiguity and the iHOP confidence value ★★★.

Find in this Page [i]

Examining the entire coding sequences as well as exon-intron boundaries of both genes by **polymerase chain reaction (PCR)** single-strand conformation polymorphism (SSCP) and multiplex-SSCP analysis, we identified possible disease-causing alterations in **BRCA1** ★ among **affected** members of 15 families and in **BRCA2** ★ in another 14 families.

**BRCA1** ★ may also **enhance** chemosensitivity and repair of DNA damage through **binding** to and coactivation of **p53** ★.

The protein encoded by **BRCA1** ★ **interacts** in vivo with the **BRCA1-associated RING domain (BAI)** protein.

Developmental expression of **Brc2a** [?] ★ suggests that it is associated with normal development and differentiation in multiple tissues.

The **RAD51** ★ 135G allele was detected in **BRCA1** ★ carriers.

However, many other mutations have been found in **BRCA1** ★ carriers.

The **BRCA1** ★ 135G allele was detected in **BRCA1** ★ carriers.

However, many other mutations have been found in **BRCA1** ★ carriers.

The **BRCA1** ★ 135G allele was detected in **BRCA1** ★ carriers.

However, many other mutations have been found in **BRCA1** ★ carriers.

The **BRCA1** ★ 135G allele was detected in **BRCA1** ★ carriers.

However, many other mutations have been found in **BRCA1** ★ carriers.

The **BRCA1** ★ 135G allele was detected in **BRCA1** ★ carriers.

However, many other mutations have been found in **BRCA1** ★ carriers.

The **BRCA1** ★ 135G allele was detected in **BRCA1** ★ carriers.

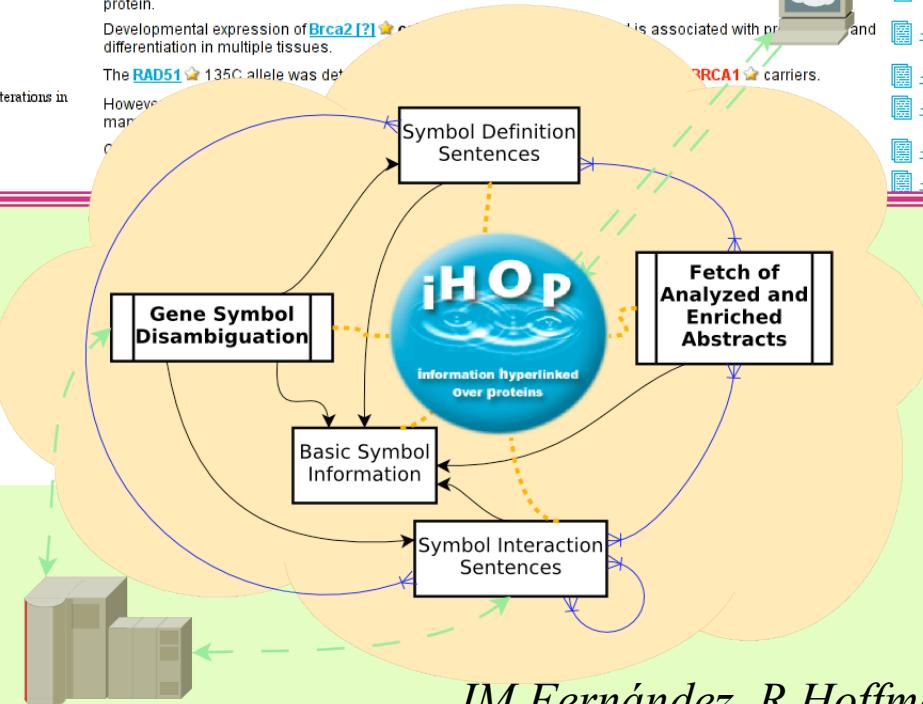
However, many other mutations have been found in **BRCA1** ★ carriers.

The **BRCA1** ★ 135G allele was detected in **BRCA1** ★ carriers.

However, many other mutations have been found in **BRCA1** ★ carriers.

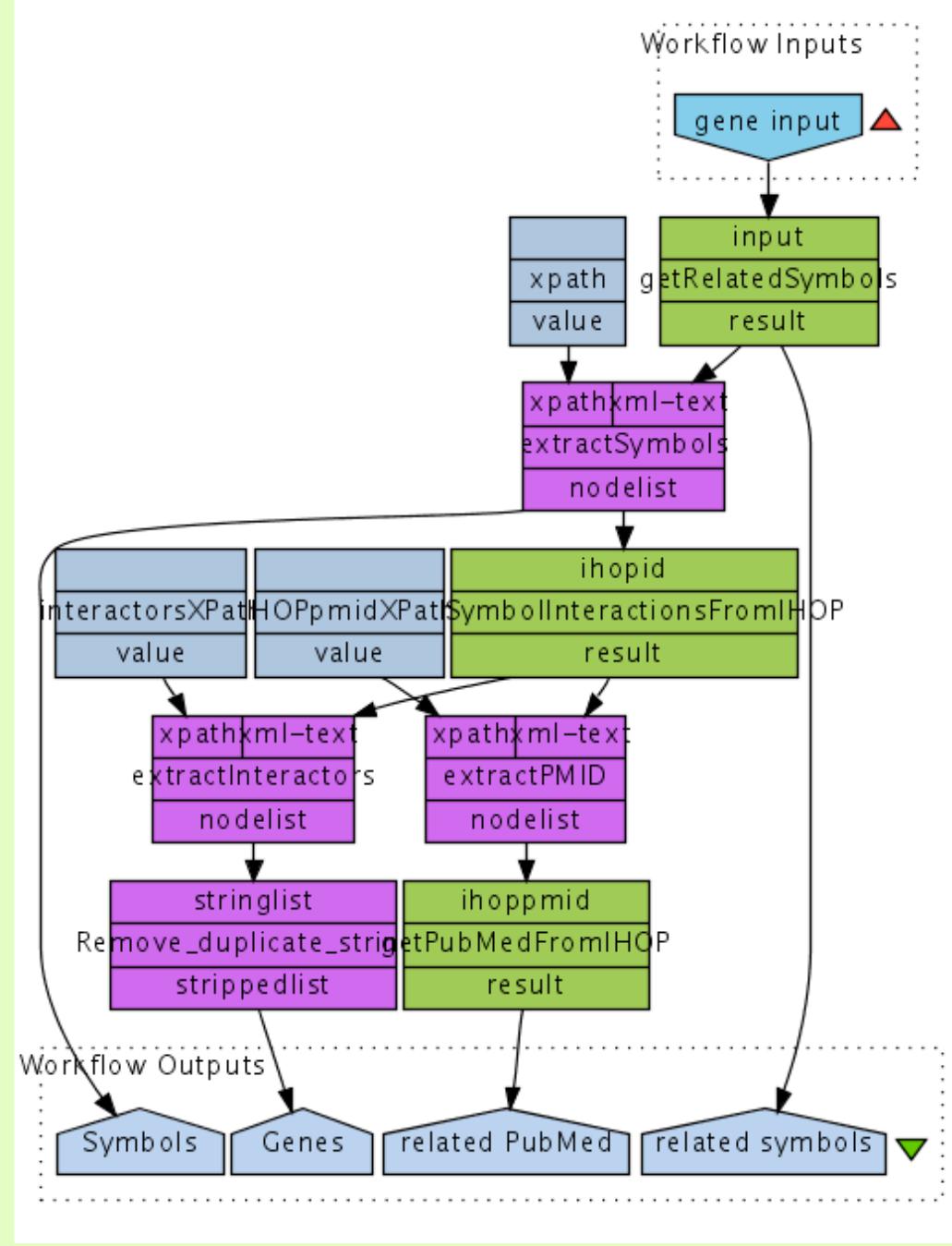
The **BRCA1** ★ 135G allele was detected in **BRCA1** ★ carriers.

However, many other mutations have been found in **BRCA1** ★ carriers.

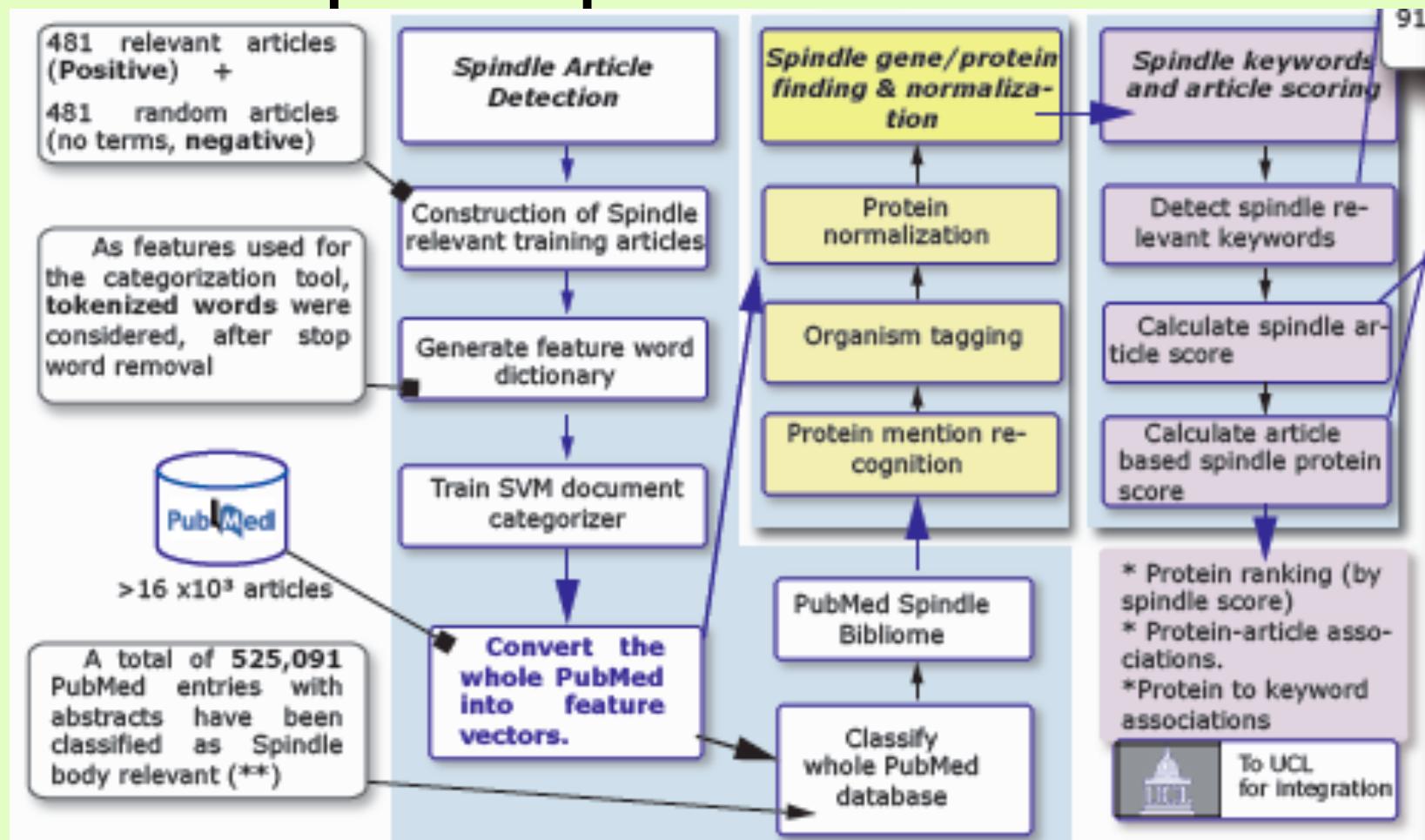


## iHOP web services

<http://www.ihop-net.org/UniPub/iHOP/webservices/>



# Spindle protein classification



## Improvements on progress:

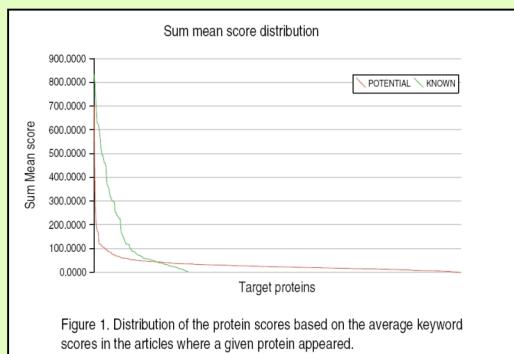
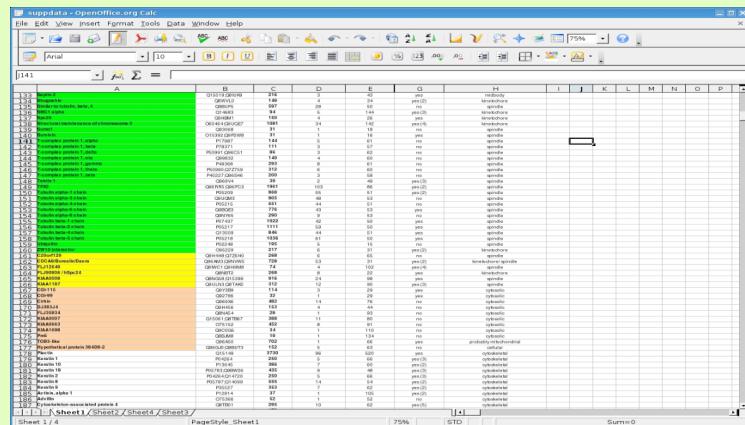
- High-recall of PubMed<sup>\*\*</sup> (filtering due to small training dataset).
- Improvement in gene mention and normalization (using external databases as GeneBank).
- Improve spindle body keyword term detection.
- Improve term scoring.
- Regenerate the bibliome of spindle body articles with SVM (using better features selection strategy).
- Comparison of the Positive with potential set scores (for defining cut off).
- Fold cross validation.

Some  
-Small  
-Missing  
-The S  
(Precision)



Sauer et al. Proteome analyses of the human mitotic spindle. *Mol Cell Proteomics.* (2005)

## Original data



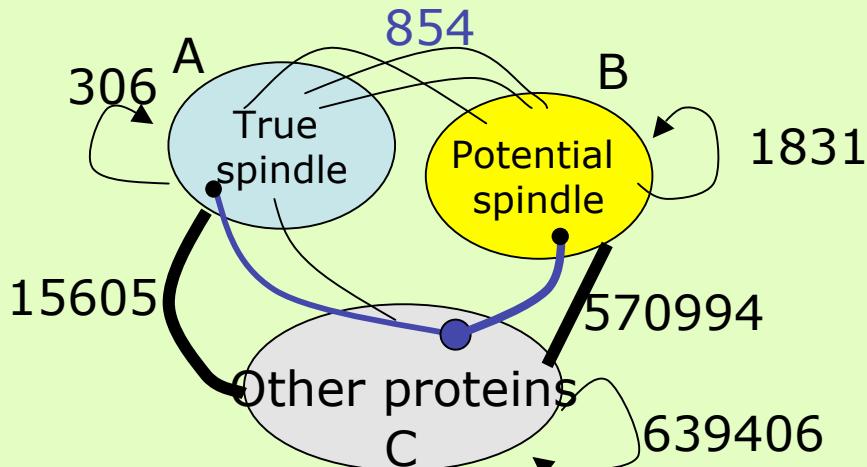
## SVM scoring based on Keyword scoring

By Martin Krallinger

Table 1. Ranked list of the top scoring potential spindle proteins based on the sum mean scores. The last column corresponds to the manual classification of the top ranking proteins according to a revision of the corresponding SwissProt records. Y: records provides positive evidence in favor of having an association to spindle proteins, U: neither positive nor negative evidence found, N: negative evidence found of not being a spindle related protein.

RANK	SWISSPROT_ID	SCORE	REVISION
1	GCP6_HUMAN	701.74	Y
2	CASC5_HUMAN	312.84	Y
3	DLG7_HUMAN	209.03	Y
4	NOC3L_HUMAN	177.47	Y
5	SMC4_HUMAN	169.77	Y
6	CEBPZ_HUMAN	121.06	Y
7	FRAP_HUMAN	117.92	Y
8	MGN_HUMAN	115.46	Y
9	SMC2_HUMAN	107.45	Y
10	RNPC2_HUMAN	106.27	Y
11	IPO7_HUMAN	103.91	Y
12	CTR9_HUMAN	100.32	Y
13	LAP2A_HUMAN	94.15	Y
14	LAP2B_HUMAN	94.15	Y
15	MPP10_HUMAN	87.21	Y
16	PWP1_HUMAN	85.55	Y
17	PRP19_HUMAN	85.23	Y
18	NOG1_HUMAN	83.58	U
19	NUP88_HUMAN	76.51	Y
20	SAM68_HUMAN	74.83	Y

**iHOP interactions: interactions extracted From literature**



By Edu Leon  
David de Juan  
Jose M Fernandez

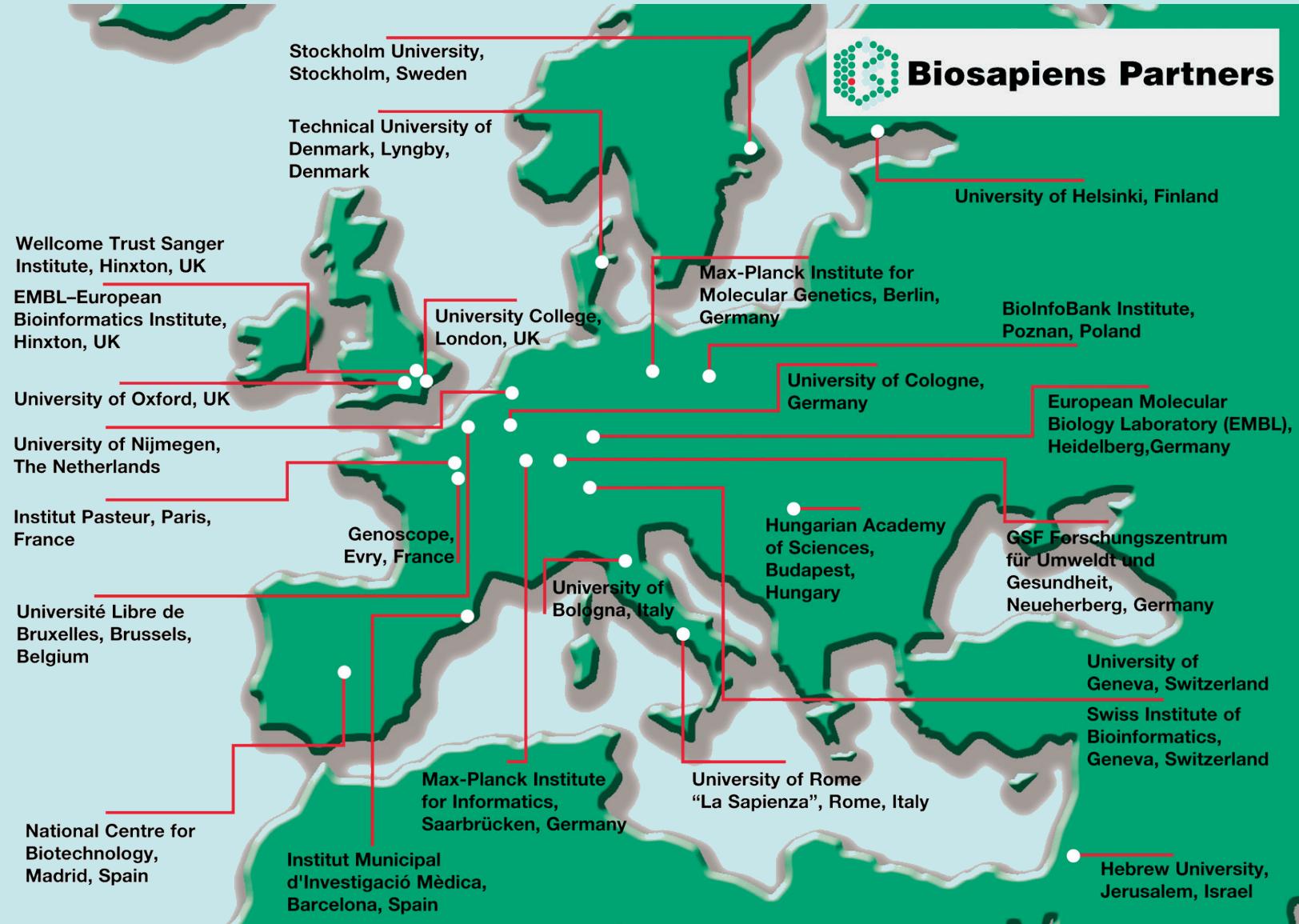
Prot(B): Q99567 Prot(A): P49792 Fmid: 14993277  
<Sentence> Nup88 and Nup214 showed an interdependence at the NPC and were not affected by the absence of Nup358. </Sentence>  
Prot(B): Q99567 Prot(A): P49792 Fmid: 14993277  
<Sentence> RNA interference of either Nup88 or Nup214 in human cells caused a strong reduction of Nup358 at the NE. </Sentence>  
Prot(B): Q99567 Prot(A): P49792 Fmid: 14993277  
<Sentence> These data indicate that Nup88 and Nup214 mediate the attachment of Nup358 to the NPC. </Sentence>  
Prot(B): Q99567 Prot(A): P52948 Fmid: 12589057  
<Sentence> Additionally, we find that the pore-targeting domain of Nup98 interacts directly with the cytoplasmic nucleoporin Nup88, a component of the Nup214, Nup88, Nup62 subcomplex. </Sentence>  
Prot(B): Q99567 Prot(A): P05239 Fmid: 15140359  
<Sentence> DNMT3B co-purifies and interacts, both *in vivo* and *in vitro*, with several components of the condensin complex (hCAP-C, hCAP-E and hCAP-G) and KIF4A. </Sentence>



# BioSapiens



## Biosapiens Partners





[European Bioinformatics Institute – Janet Thornton](#)

# BioSapiens

University of Rome “La Sapienza” - [Anna Tramontano](#)

Consejo Superior de Investigaciones Científicas - [Alfonso Valencia](#)

Technical University of Denmark - [Søren Brunak](#)

European Molecular Biology Laboratory – Peer Bork

German National Centre for Health and Environment – Dmitrij Frishman

Université Libre de Bruxelles - Shoshona Wodak

Institut Municipal d'Investigació Mèdica - Roderic Guigó

[The Wellcome Trust Sanger Institute - Tim Hubbard](#)

Max-Planck Institute for Informatics - Thomas Lengauer

The Hebrew University of Jerusalem - Michal Linial

Stockholms Universitet - Gunnar von Heijne

University of Oxford - Richard Mott

University College London - Christine Orengo

Stichting Katholieke Universiteit - Gert Vriend

Swiss Institute of Bioinformatics - Anne-Lise Veuthey

University of Helsinki - Esko Ukkonen

University of Geneva - Stylianos Antonarakis

Institute of Enzymology (Hungarian Academy of Sciences) - László Patthy

Universität zu Köln - Dietmar Schomburg

Institut Pasteur - Antoine Danchin

BioInfoBank Institute - Leszek Rychlewski

Max-Planck Institute for Molecular Genetics – Martin Vingron

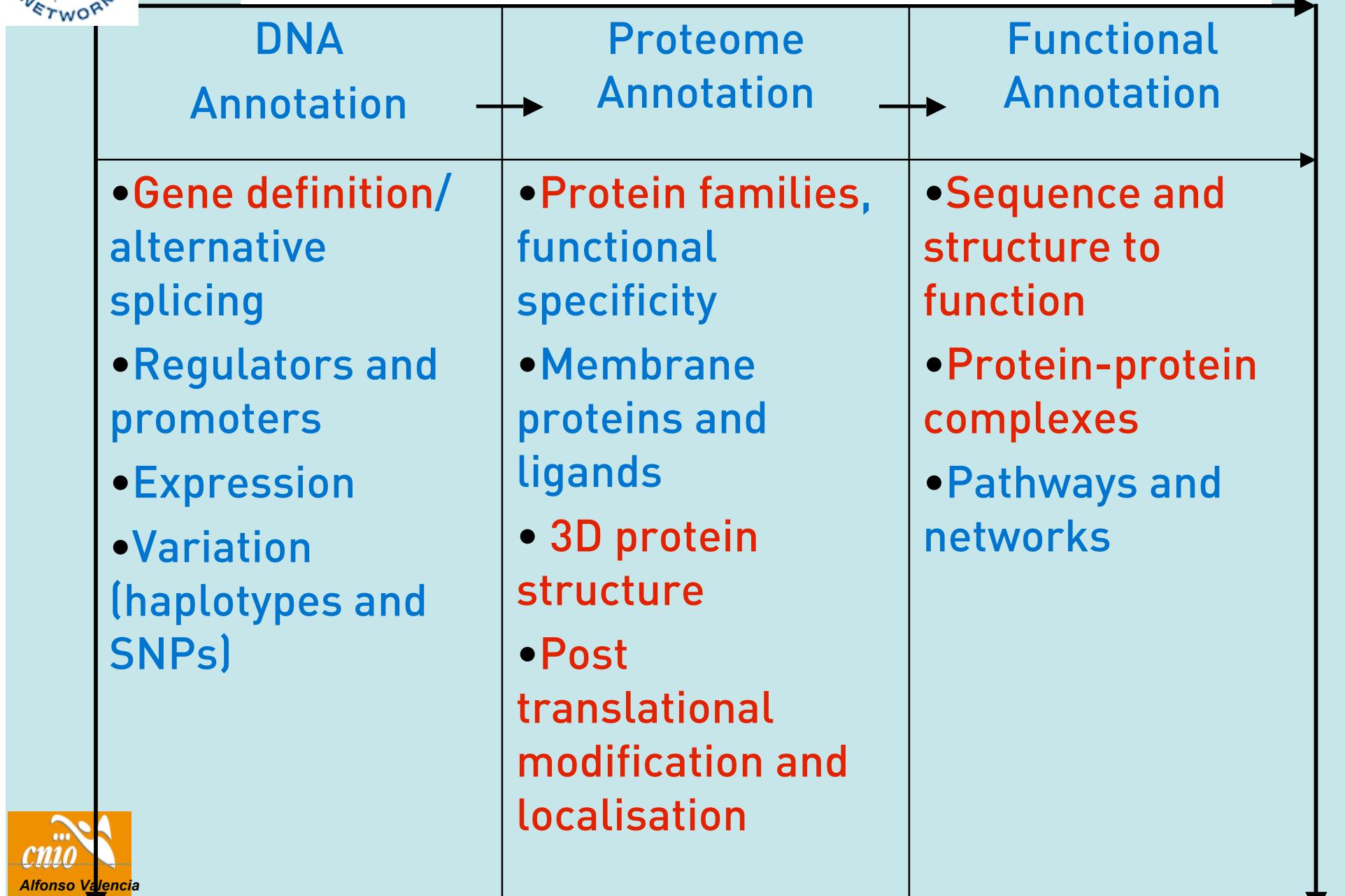
Genoscope – Vincent Shachter

University of Bologna – Rita Casadio

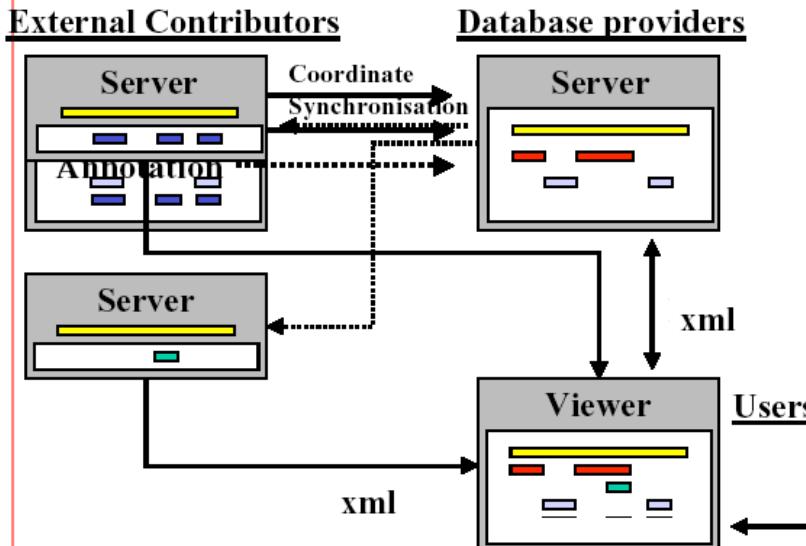




# A European Virtual Institute for Genome Annotation



## Distributed Annotation



## DAS extensions

- **Protein DAS**
  - Provide features in protein coordinates
  - Can handle Uniprot/RefSeq coordinates rather than Ensembl coordinates
- **GeneDAS**
  - Provide text on geneview pages

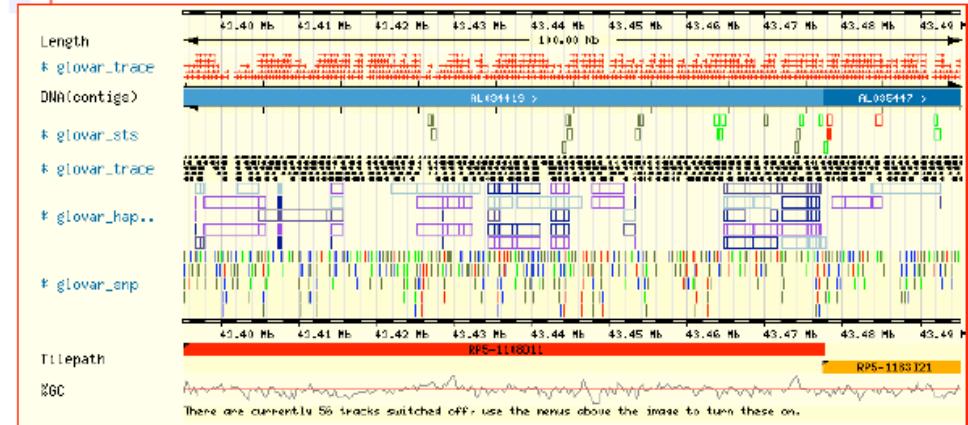


## Why DAS in BioSapiens?

- **Positives:**
  - Established and working now
  - Simple
  - Good fit with many annotation sites (you) and fewer display/viewing sites
  - No need for highly centralised control
- **Negatives:**
  - Not semantically rich
    - Can extend with DAS/2 or other mechanisms

9 of 27

## Genomic DAS



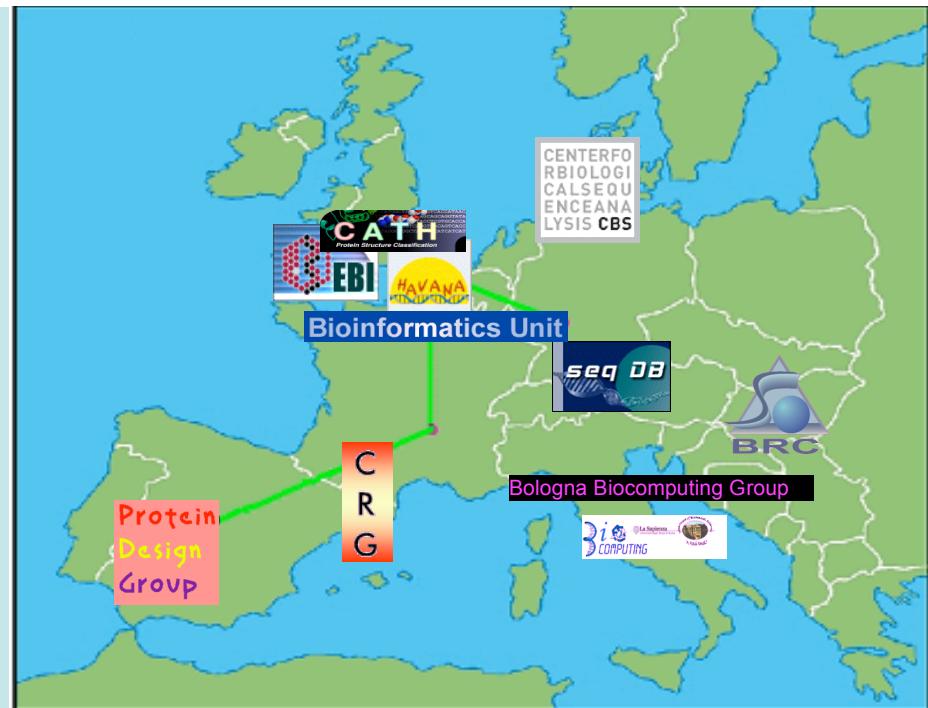
By Ewan Birney

# BioSapiens



The BioSapiens-sponsored project concentrated on the protein coding loci and in particular on the alternatively spliced products.

This work is part of the BioSapiens efforts for the annotation of the human genome ([www.biosapiens.info](http://www.biosapiens.info)).

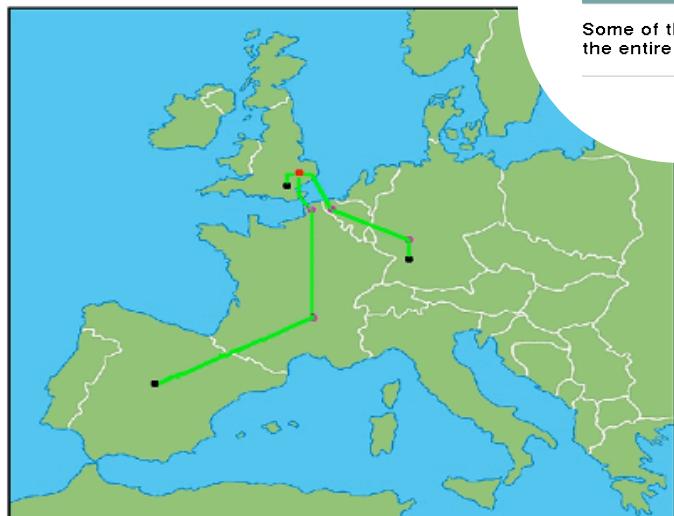
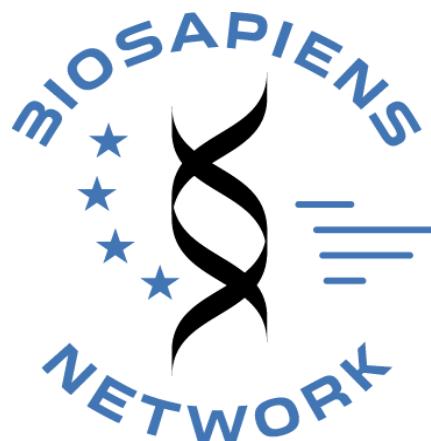


The screenshot shows a complex web interface with several panels:

- ENCODE Data**: Includes a bar chart and links to "Summary of Sequence Data" and "Human homologues, PDB structures, Pfam domains and links to annotated sequences".
- Introductory Page**: Features the BioSapiens logo and navigation links to "Talks at Biosapiens Meeting, Barcelona" (listing PDG, EBI, Institute of Enzymology, and Biocomputing Group), "Data Pages from the PDG" (listing Summary Pages, Best Human, Mouse and Zebra Fish Homologues, GO Functional Terms, PDB Structures, and Domains and Splice Variants), and "CBS Prediction Results" (listing SignalP, TargetP, TMHMM, NetNGlyc, NetOGlyc, ProP, NetAcet, NetNES, NetPhos Predictions, and CBS Data With Links to Annotated Sequences). A "cni" logo is visible in the bottom left corner.
- Design Group**: A sidebar with a red background containing a list of annotations from the PDG.
- Talks at Biosapiens Meeting, Barcelona**: A sidebar listing speakers and their institutions.
- Data Pages from the PDG**: A sidebar listing various data pages and analysis tools.
- CBS Prediction Results**: A sidebar listing prediction results for signal peptides and other bioinformatics tools.

Splice variants were annotated by the Havana Group as part of the GENCODE consortium.

This complete set of splice variants allows us to pose interesting questions, particularly whether the gene products would fold into functional proteins.



Funded by BIOSAPIENS NETWORK Coordinated by Funded by

A European Virtual Institute for Genome Annotation

### Structure Function Pipeline.

From this page you can submit your own structure or analyse an existing PDB entry.

**Enter Details:**

Email Address:   
 NCBI Taxon ID:  OR Species Name:   
If you want to see the NCBI tax IDs click [here](#)

**Select Option:**

Upload PDB-format file:    
or  
 Get existing PDB file\* PDB code:

Some of the methods take minutes to run; others take hours. You will be notified by e-mail when the entire process is complete, but can check on preliminary results as they become available.

E-mail to:- James Watson [watson@ebi.ac.uk](mailto:watson@ebi.ac.uk)

**•Utilises SOAP interfaces to simultaneously access:**

- ProFunc (EBI, Hinxton, UK)
- CATH (UCL, London, UK)
- FUNCut (CNB, Madrid, Spain)
- STRING (EMBL, Heidelberg, Germany)

### Structure to Function Pipeline - running



# Structure to Function Pipeline - results

**A European Virtual Institute for Genome Annotation**

### Pipeline results for Uploaded structure

**Profunc Results**

**Sequence motifs**

InterPro scan for sequence motifs. Chain A

**8 motifs matched in scan against PROSITE, PRINTS, PFam-A, TIGRFAM, PROFILES and PRODOM motifs**

Type	Motif	Name
1. ???	G3D.3.40.50.300	no description
2. ScanRegExp	PS00039	DEAD_ATP_HELICASE
3. superfamily	SSF52540	P-loop containing nucleoside triphosphate hydrolases
4. HMMSmart	SM00487	no description
5. HHMPfam	PF00270	DEAD
... plus others		

InterPro scan for sequence motifs. Chain B

**10 motifs matched in scan against PROSITE, PRINTS, PFam-A, TIGRFAM, PROFILES and PRODOM motifs**

Type	Motif	Name
1. HMMSmart	SM00487	no description
2. HMMSmart	SM00490	no description
3. ScanRegExp	PS00039	DEAD_ATP_HELICASE
4. ???	G3D.3.40.50.300x2	no description
5. HHMPfam	PF00270	DEAD
... plus others		

**Matches to the Superfamily HMM library**

Search of sequence vs Superfamily HMM library. Chain A

**1 motif matched in scan against the Superfamily HMM library**

Residue range(s)	Motif	Superfamily name
1. 20-221	52540	P-loop containing nucleoside triphosphate hydrolases

**DNA**  
DNA-binding templates.  
**No hits obtained from any of the 1386 DNA-binding templates.**

**REVERSE**  
Reverse template comparison vs structures in PDB.

**20 significant hits out of 1190 auto-generated templates.**

Score	Template	PDB Name
1. 1060.000	TMP01059	1fuu Yeast initiation factor 4a
2. 989.000	TMP01059	1fuk Crystal structure of the aptase domain of translation initiation factor 4a from saccharomyces cerevisiae-the prototype of the dead box protein family
3. 897.000	TMP00259	1fuk Crystal structure of the carboxy terminal domain of yeast eif4a
4. 562.688	TMP00676	1t5i Crystal structure of the C-terminal domain of uap56
5. 648.875	TMP00038	1hv8 Crystal structure of a dead box protein from the hyperthermophile methanococcus jannaschii
... plus others		

Generated on Fri 20 Jan 2006

**CATH database**

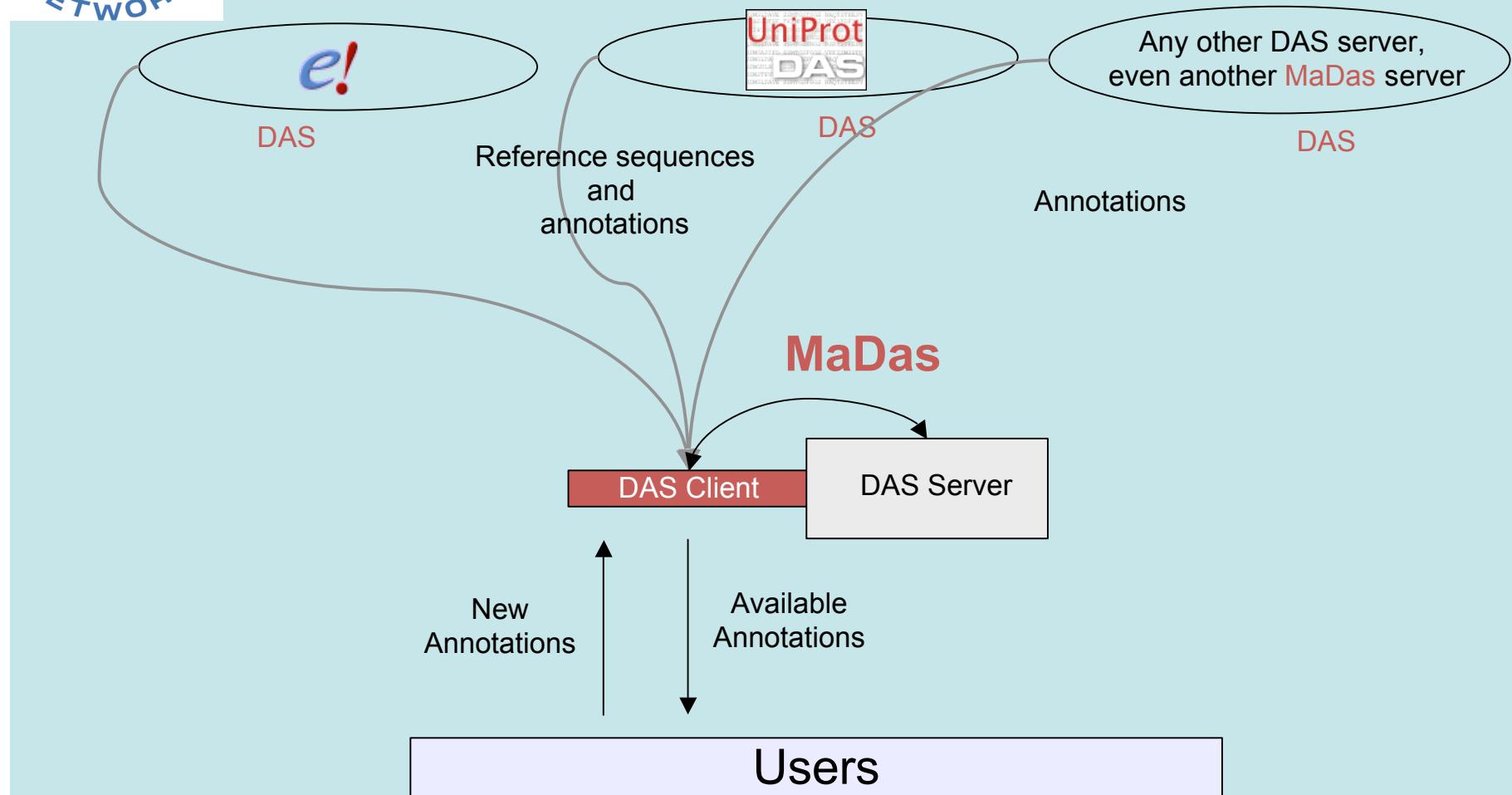
**CATH results for blast scan:**

PDB Code or Match	Description	Sequence Identity	Evalue of the Match
1qde	Gene regulation Translation initiation factor 4a. Chain: a. Fragment: n-terminal domain, residues 9-232. Synonym: eif4a. Engineered: yes	81.517	2e-65 5e-65
1fuk	Translation Eukaryotic initiation factor 4a. Chain: a. Fragment: carboxy terminal domain (residues 230-394). Synonym: yeast initiation factor 4a, eif4a. Engineered: yes	90.323	1e-61
1hv8	RNA binding protein Putative apt-dependent RNA helicase. Chain: a. b. Synonym: dead box helicase. Engineered: yes	35.106	1e-18 5e-19
1g0u	RNA binding protein Btddead. Chain: a. b. Fragment: n-terminal domain. Engineered: yes	34.444	4e-11 9e-11
1b73	Isomerase Glutamate racemase. Chain: a. EC: 5.1.1.3	25.773	0.22
1ayx	Hydrolyase Glucosaminidase. Chain: null. Engineered: yes	26.984	0.67 1.4
1ff9	Oxidoreductase Saccharopine reductase. Chain: a. Engineered: yes	28.571	0.88 1.8
1ng6	Structural genomics, unknown function Hypothetical protein ygey. Chain: a. Engineered: yes	23.529	1.1
1q8y	Transferase Sg protein kinase. Chain: a. b. Fragment: skylpellet(13')delta. Engineered: yes	45.833	2.0 4.1
1ckv	Hydroxylase regulatory protein Protein b. Chain: null. Biological unit: monomer	29.787	3.3 7.0
1d7u	Lysate 2,2-dialkylglycine decarboxylase (pyruvate). Chain: a. Synonym: dgd. Engineered: yes	52.381	5.7
1jp	Oxidoreductase Aspartate dehydrogenase. Chain: a. Engineered: yes	33.333	5.7
1auu	Transcription regulation Sacy. Chain: a. b. Fragment: RNA binding domain, residues 1 - 55. Engineered: yes. Biological unit: dimer. Other details: a terminal region of the sacy protein. The structure is a symmetric dimer in solution	23.404	7.0



# MaDas

## Manual Sequence Annotation System



## MaDas Web Interface

UPLOAD DATA DOWNLOADS HELP PDG

### MaDas

Version 0.5 (For testing purposes only).

Welcome Guest !!!

MaDas is a manual sequence annotation system.

The general purpose is to provide users with a simple tool to store user self-generated sequence annotations in either a local or public repository. The system is based on the DAS (Distributed Annotation System) technology and is composed by two modules: a DAS annotations server and a DAS client. The current version (0.5) allows the users to annotate both genomes and proteins. In addition, the FunCUT generated annotations and the Ensembl and UniProt DAS information are also displayed.

You need to register to use it. For more information check out the [HELP](#) page.

This is a testing version. Therefore comments and suggestions from users are very welcome! So, go ahead and e-mail me [here](#).

Type the protein name or the accession number

#### Overview

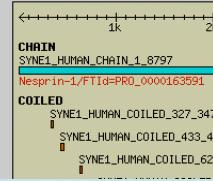


Region from 1 to 8797 bp is showed in the detail window

#### Details

Window size 100 kbp

\* Click on the annotations to perform



Annotate your sequence

Feature name*	<input type="text"/>
Type*	<input type="button" value="Pick one ..."/>
Or enter some new type ...	
Category	<input type="text"/>
Method*	<input type="button" value="MaDas"/>
Start*	<input type="text"/> End* <input type="text" value="8797"/>
Score	<input type="text"/>
Orientation*	<input type="button" value="0"/>
Phase*	<input type="button" value="-"/>
Note	<input type="text"/>
Target name	<input type="text"/>
Target Start	<input type="text"/>
Target End	<input type="text"/>
Group Name	<input type="text"/>
Group Type	<input type="text"/>
<input type="button" value="Submit Annotations"/>	

#### Annotation details

Reference Sequence	Feature Name	Type	Category	Method
Q8NF91	SYNE1_HUMAN_Spectrin_7022_7127	Pfam		Pfam

#### Sequence

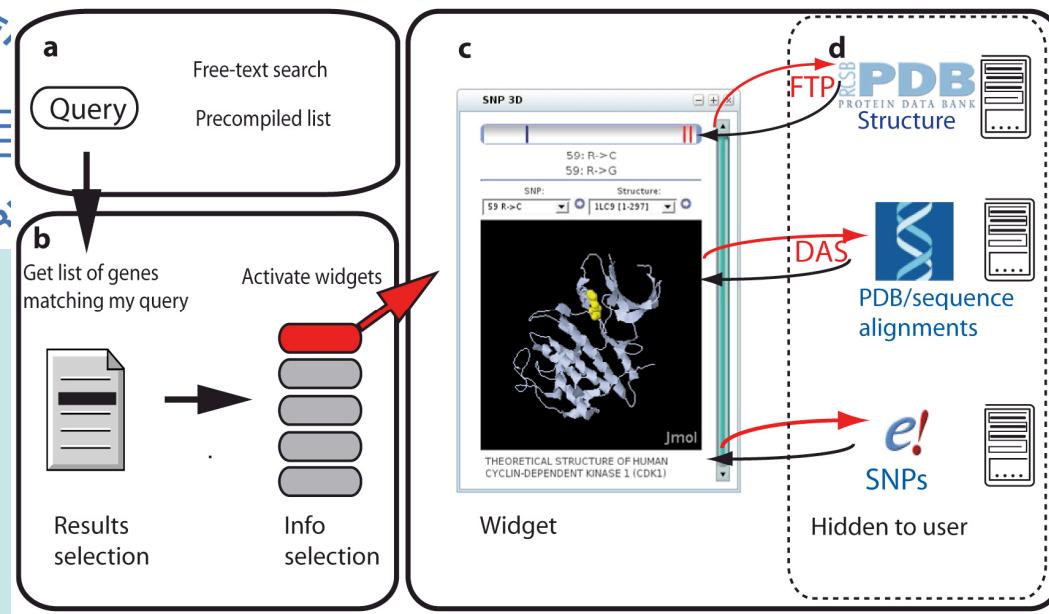
ESWSSEYENNQCLKIWFTEQEKRLKQQHRIGDQASVQNALKDCQDLEDLIKAKEVEKI EQNGLALIQNKEDVSSIVMSTLRELGQTIWANLDHMVGQLKILLKS

☒ MaDas allows users to add, edit, or remove self generated sequence annotations

☒ Allows to upload multiple annotations from a file.

☒ Provides a security system based on projects. The annotations could be public or only available for the project members.

☒ Provides an interface to manage projects, users and collections of annotations.



## CARGO

[bioinfo.cnci.es/cargo/](http://bioinfo.cnci.es/cargo/)

I Cases, AM Rojas, A Carro, JM Fernandez-G, JF Vera, JM Rodriguez, G Gomez, EA Leon, A Valencia, D G Pisano\*

**CARGO: Cancer And Related Genes Online - Mozilla Firefox**

P53  
Cellular tumor antigen p53 Tumor suppressor p53 Phosphoprotein p53 Antigen NY-CO-13  
ENSG00000141510

File Edit View History Bookmarks Tools Help

**CARGO**  
Cancer And Related Genes Online

Literature Mining  
Disease Info  
Transcripts Annotations  
3D Coding SNPs  
Protein Interactions  
Example: P53  
Search  
Example: P53  
Cancer Candidate

How many Structures?  
Coding SNP's?  
Is any allelic variant a coding SNP?  
Where in PDB?  
Any disease associated to this variant?  
What is published?

SNP 3D  
249: R>S  
Structure:  
249 R>S | 1TUP [84-289]

Jmol

OMIM  
+ TEXT  
- ALLELIC VARIANTS  
+ 0001:LI-FRAUMENI SYNDROME 1 [TP53, ARG248TRP]  
+ 0002:LI-FRAUMENI SYNDROME 1 [TP53, GLU258LYS]  
+ 0003:LI-FRAUMENI SYNDROME 1 [TP53, GLY245CYS]  
+ 0004:LI-FRAUMENI SYNDROME 1 [TP53, LEU252PRO]  
+ 0005:TP53 POLYMORPHISM [TP53, PRO72ARG]  
- 0006:HEPATOCELLULAR CARCINOMA [TP53, ARG249SER]

iHOP  
Sentences that define the Gene  
Deletions of Sp21 and TP53 in bladder cancer.  
PubMed 12550767  
Evidence for another tumor suppressor gene at 17p13.3 distal to TP53 in hepatocellular carcinoma.  
MeSH term: Carcinoma, Hepatocellular  
PubMed 1052  
No changes in the mRNA or protein level of TP53 and E6 could be detected.

Interaction with other gene symbols

**cnio**  
Alfonso Valencia



C A R G O

Literature Mining  
Disease Info  
Transcripts Annotations  
3D Coding SNPs  
Protein Interactions  
 Cancer Candidate  
 Search  
Example: P53, Colon

Cellular tumor antigen p53  
Tumor suppressor p53  
Phosphoprotein p53 Antigen  
NY-CO-13  
ENSG00000141510

## CARGO in the BioSapiens analysis of cancer gene list WP

P53  
Cellular tumor antigen p53 Tumor suppressor p53 Phosphoprotein p53 Antigen NY-CO-13  
ENSG00000141510

iHop OMIM FunCut SNP 3D Interactome

**iHop**  
Sentences that define the Gene  
Interaction with other gene symbols

**PubMed 9338106**  
Cell culture studies have clearly demonstrated that TP53 can induce and BCL2 can suppress apoptosis in response to various stimuli.

**PubMed 10963376**  
Tp53 protein may be inactivated by the binding of the MDM2 protein.

**PubMed 12699883**  
hypermethylated in tumors of most of their patients with primary esophageal adenocarcinoma.

Ishii et al. (1999) found mutations in the LZTS1/P

**SNP 3D**  
47: P->S  
SNP: Structure:  
\* 47 P->S 1X0H [369-374]

**FunCut**  
Transcript Annotations from P53 gene

List of transcripts (order-numAminoAcids): 0-341

- Consensus annotated description: Cellular tumor antigen p53 (Tumor suppressor p53) (Phosphoprotein p53) (Antigen NY-CO-13)
- Closer protein in the family: P04637

Interactome  
EMSEMBL ID: ENSG00000141510  
SYNONYM NAMES: P53, TP53

Bioinformatics Unit | Structural & Computational Biology Group | CNIO

Some Widget are DAS mini-clients producing “information digestion”



BioCreAtIvE

COMBIO



ENFIN

Enabling Systems Biology



EMERGENCE COST  
CA on Synthetic Biology

[www.pdg.cnb.uam.es](http://www.pdg.cnb.uam.es)

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



- Osvaldo Graña
- Victor de la Torre
- Ildefonso Cases, Jose M. Fernandez-G, Jaime F. Vera, Jose M. Rodriguez, Gonzalo Gomez, Eduardo A. Leon, Angel Carro, Ana Rojas, David G. Pisano
- Robert Hoffmann, Martin Krallinger, Jose M. Fernandez

- BIOCRA  
TIVE II.  
- CASP7

[biocreative.sourceforge.net](http://biocreative.sourceforge.net)  
[predictioncenter.org](http://predictioncenter.org)