Alcune risorse importanti

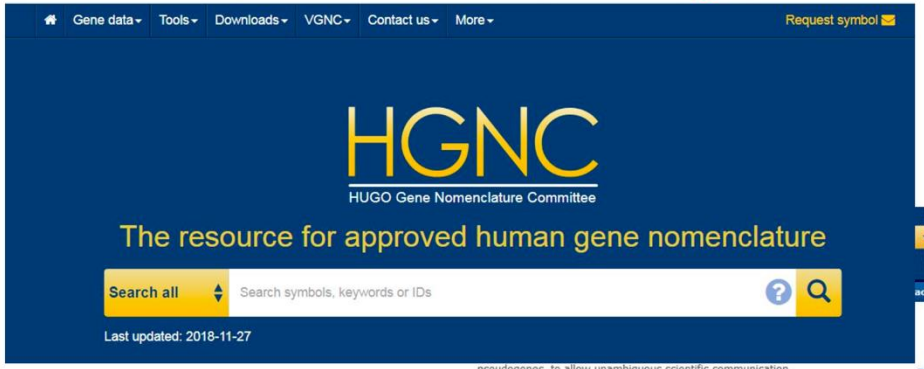
HGNC & Genome Browser

Nomenclatura Genica

Purtroppo **non esiste un modo univoco per indicare uno specifico gene**. Questo crea confusione e non facilita la ricerca informatica.

Ad esempio il gene tcap a seconda del *record* è riportato come titin-cap protein o telethonin.

Human Genome Organisation





https://www.genenames.org

pseudogenes, to allow unambiguous scientific communication.

genenames.org is a curated online repository of HGNC-approved gene nomenclature, gene families and associated resources including links to genomic, proteomic and phenotypic information.

Search our catalogue of more than 39,000 symbol reports using our improved search engine (see Search help), search lists of symbols using our Multi-symbol checker and identify possible orthologs using our HCOP tool.

Download our ready-made data files from our Statistics and Downloads page, create your own datasets using either our Custom Downloads tool or BioMart service, or write a script/program utilising our REST service.

Submit your gene symbol and name proposals to us to be accredited with HGNC approved nomenclature for use in publications, databases and presentations.



HGNC

The Human Genome Organisation ha istituito un comitato allo scopo di dare un unico nome significativo a tutti i geni umani. Con questo intento è stato costruito il database **HGNC.**

HGNC

HUGO Gene Nomenclature Committee

Il progetto HUGO Gene Nomenclature Committee (HGNC) è stato fondato nel **1979** come parte dell'iniziativa della **Human Genome Organisation (HUGO)** un'organizzazione internazionale creata per promuovere la cooperazione globale nella ricerca sul genoma umano

Motivo:

- **sviluppare una nomenclatura standard** per i geni umani, in vista della futura mappatura completa del genoma umano
- risolvere i **problemi di confusione e ambiguità** nella denominazione dei geni, che ostacolavano la comunicazione e la condivisione dei risultati scientifici

L'HGNC **continua tutt'ora** ad aggiornare e migliorare la nomenclatura genetica umana, lavorando con altri enti internazionali per assicurare coerenza con le nomenclature di altri organismi modello.

HGNC
HUGO Gene Nomenclature Committee

Obiettivi:

1. Uniformità e chiarezza: La mancanza di una nomenclatura standard rendeva complesso associare univocamente un nome o simbolo a uno specifico gene.

2. Facilitare la ricerca : Avere una nomenclatura standard facilita l'aggregazione di dati, il confronto tra studi e di integrare risultati in modo più efficace.

1. Collaborazione interdisciplinare e internazionale: Una nomenclatura condivisa permette ai ricercatori di tutto il mondo di parlare la stessa "lingua" in ambito genetico, facilitando la collaborazione e la condivisione dei risultati su scala globale.

Symbol report for PTEN Stable symbol

Report | HCOP homology predictions

HGNC data for PTEN

Approved symbol	PTEN
Approved name	phosphatase and tensin homolog
Locus type	gene with protein product
HGNC ID	HGNC:9588
Symbol status	Approved
Previous symbols	BZS; MHAM
Alias symbols	MMAC1; TEP1; PTEN1
Alias names	" mutated in multiple advanced cancers 1 " " Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN "
Chromosomal location	10q23.31
Gene groups	PTEN protein phosphatases C2 tensin-type domain containing Phosphoinositide phosphatases

Gene resources for PTEN

Ensembl	ENSG00000171862 Curated Ensembl region in detail, Ensembl gene sequence	**NCBI Gene**	5728 Curated
UCSC	uc001kfb.4	**Alliance of Genome Resources**	HGNC:9588

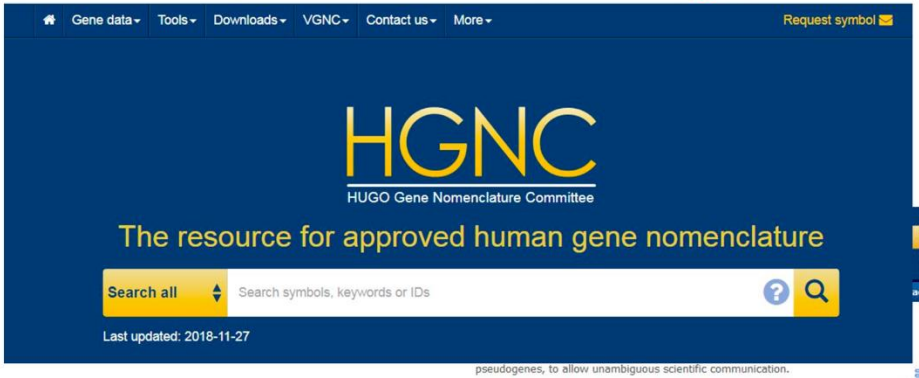
Nucleotide resources for PTEN

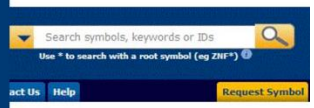
MANE Select	NM_000314.8 ENST00000371953.8	**INSDC**	U92436 Curated ENA, GenBank, DDBJ
RefSeq	NM_000314 Curated NCBI sequence viewer	**CCDS**	CCDS31238 Curated

Protein resources for PTEN

UniProt/Swiss-Prot	P60484	**AlphaFold**	AF-P60484-F1

Human Genome Organisation





https://www.genenames.org

pseudogenes, to allow unambiguous scientific communication.

genenames.org is a curated online repository of HGNC-approved gene nomenclature, gene families and associated resources including links to genomic, proteomic and phenotypic information.

Search our catalogue of more than 39,000 symbol reports using our improved search engine (see Search help), search lists of symbols using our Multi-symbol checker and identify possible orthologs using our HCOP tool.

Download our ready-made data files from our Statistics and Downloads page, create your own datasets using either our Custom Downloads tool or BioMart service, or write a script/program utilising our REST service.

Submit your gene symbol and name proposals to us to be accredited with HGNC approved nomenclature for use in publications, databases and presentations.



The Human Genome Organisation ha istituito un comitato allo scopo di dare un unico nome significativo a tutti i geni umani. Con questo intento è stato costruito il database **HGNC.**

HGNC

Andate sul sito (scansionate il QR code)

1. Cercate il gene **ZBTB7A**. Quanti ALIAS ha ? Ce n'è qualcuno strano ?

Genome Browser

L'enorme aumento dei dati riguardanti interi genomi, in particolare quelli derivanti dai progetti di sequenziamento di vertebrati, ha richiesto lo sviluppo di veri e propri ***browser* genomici** per facilitarne la consultazione.

Per questo motivo presso la University of California Santa Cruz (UCSC) venne sviluppato nel 2000 uno dei primi ***genome browser*** in grado di fornire una rapida **visualizzazione grafica** di ogni regione di genoma di qualsivoglia lunghezza assieme ad una grande quantità di informazioni come:

- geni noti, geni predetti, ESTs, mRNA, esperimenti di RNA-seq
- elementi regolativi, geni omologhi di altri organismi, ecc.

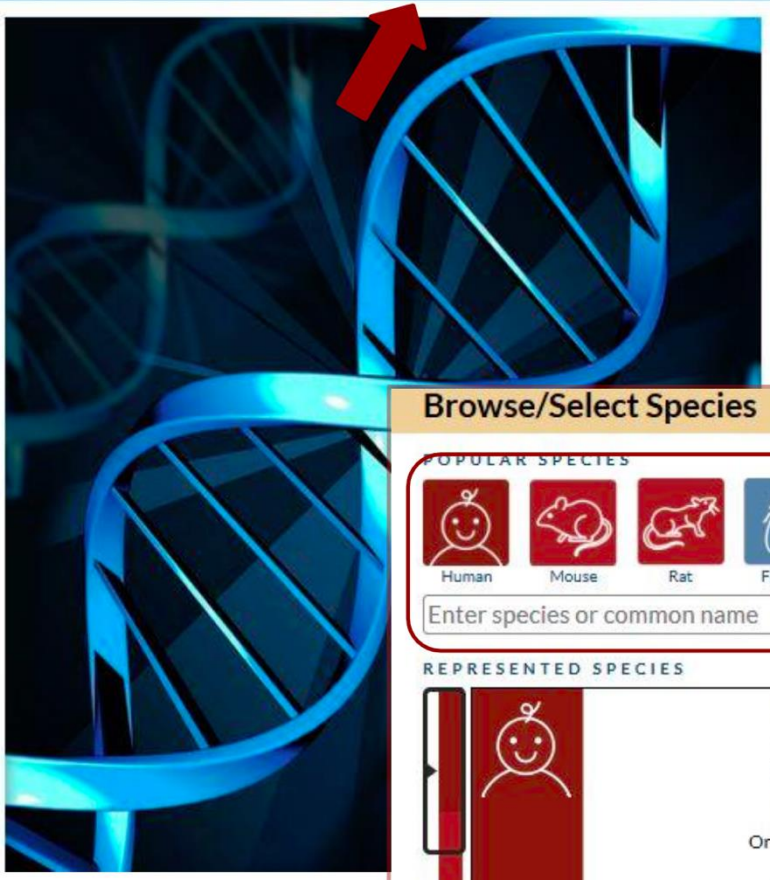
Successivamente anche i principali siti di informazioni scientifiche come **l'NCBI e l'EBI** hanno sviluppato piattaforme per la visualizzazione dati sempre più complesse, in grado di integrare il maggior numero possibile di informazioni sul genoma umano e su numerosi altri organismi.

UNIVERSITY OF CALIFORNIA SANTA CRUZ UCSC **Genome Browser**

http://genome.ucsc.edu/

University of California Santa Cruz

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Help | About Us



Our tools

- **Genome Browser** interactively visualize genomic data
- **BLAT** rapidly align sequences to the genome
- **Table Browser** download data from the Genome Browser database
- **Variant Annotation Integrator** get functional effect predictions for variant calls

Browse/Select Species

POPULAR SPECIES

Human | Mouse | Rat | Fruitfly | Worm | Yeast

Enter species or common name

REPRESENTED SPECIES

Human
Chimp
Bonobo
Gorilla
Orangutan
Gibbon
Green monkey
Crab-eating macaque
Rhesus
Baboon (anubis)
Baboon (hamadryas)
Proboscis monkey
Golden snub-nosed monkey

Find Position

Human Assembly

Dec. 2013 (GRCh38/hg38)

GO

Position/Search Term

Enter position, gene symbol or search terms

Current position: chr1:11,102,837-11,267,747

Human Genome Browser - hg38 assembly

view sequences

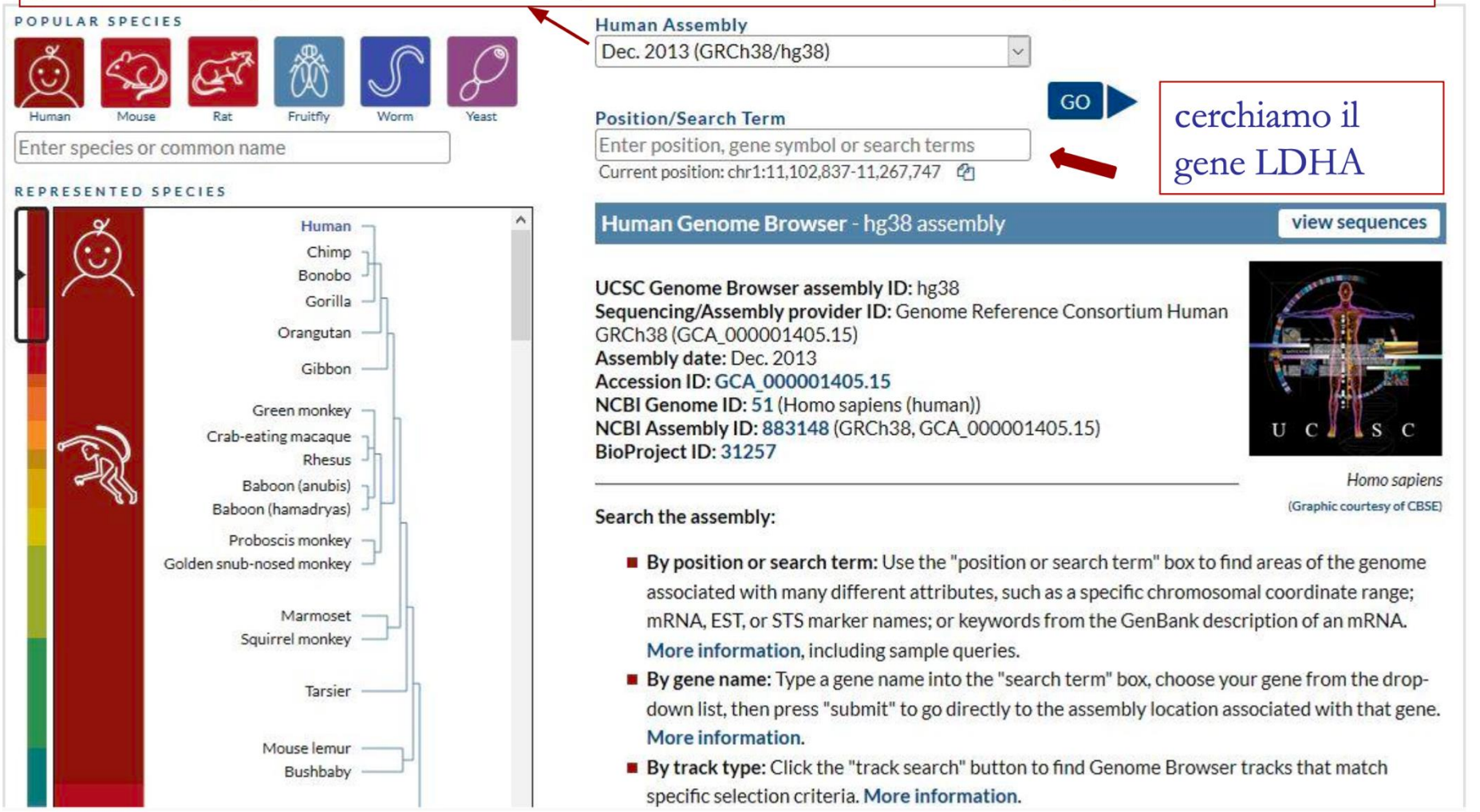
UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38 (GCA_000001405.15)
Assembly date: Dec. 2013
Accession ID: GCA_000001405.15
NCBI Genome ID: 51 (Homo sapiens (human))
NCBI Assembly ID: 883148 (GRCh38, GCA_000001405.15)



Homo sapiens

(Graphic courtesy of CBSE)

Nota: durante il sequenziamento di un genoma, spesso sono rilasciate versioni (*release*) successive che possono essere più o meno definitive. Qui si fa riferimento alla versione del genoma umano del 2013

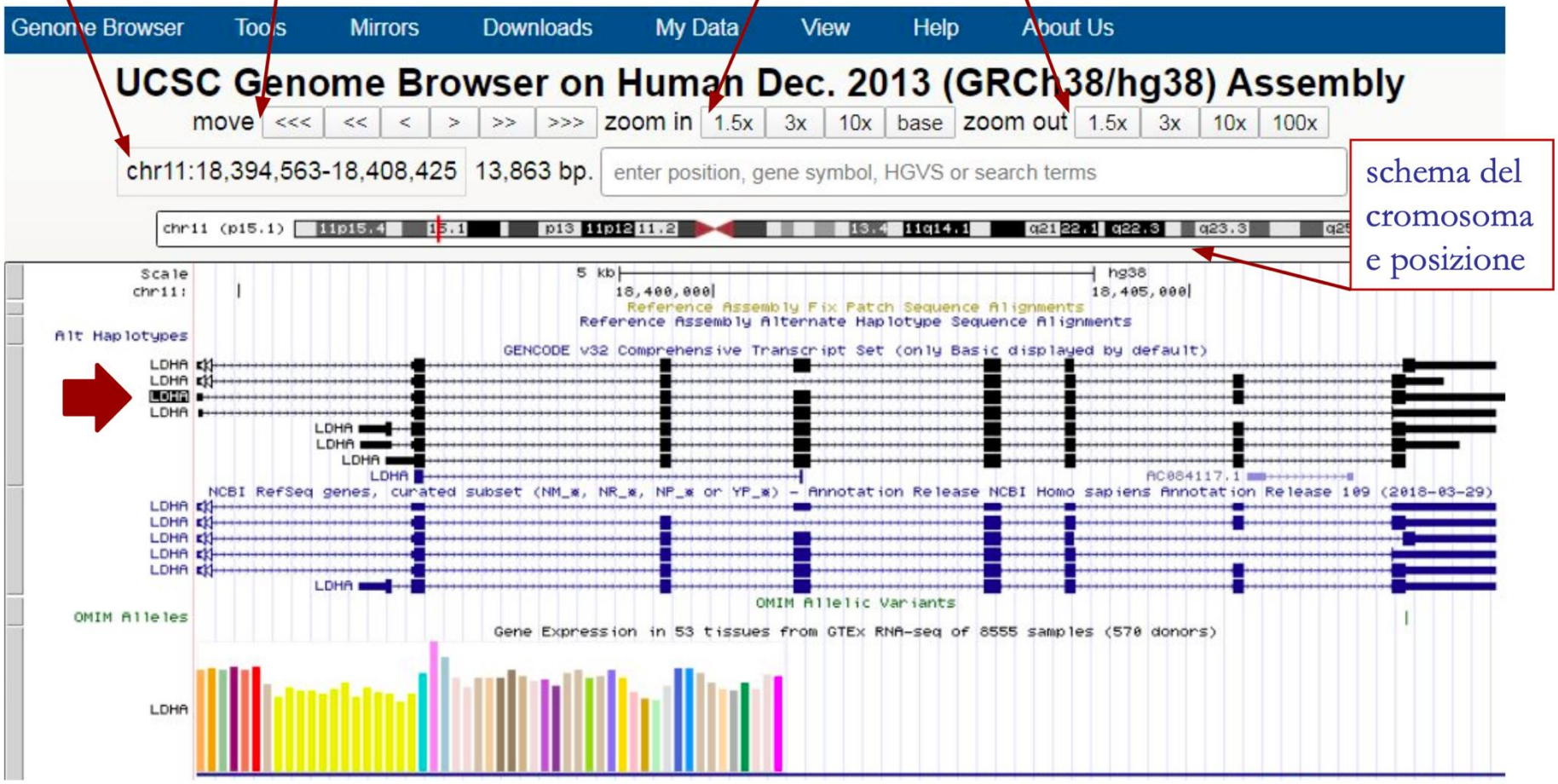


E' possibile scegliere quale **genoma** visualizzare, e cercare al suo interno un particolare gene, o una **specifica regione** di interesse o un particolare tipo di **esperimento**.

Posizione sul genoma

Pulsanti di spostamento sul genoma

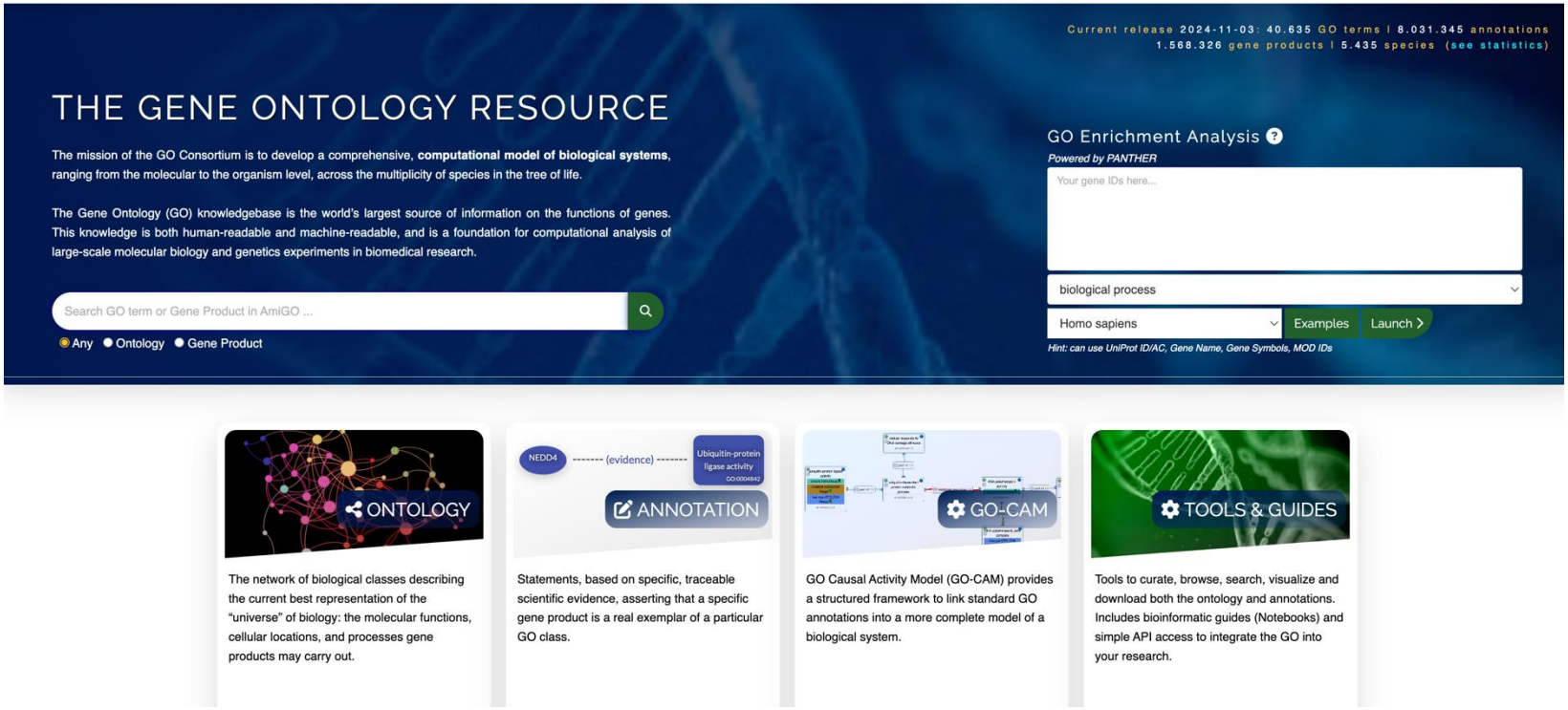
Pulsanti per ingrandire o rimpicciolire l'area



E' possibile raggiungere specifiche posizioni digitando direttamente nell'area di ricerca

Gene Ontology And Pathways

Annotazione funzionale



https://geneontology.org/

In informatica, un'**ontologia** è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse.
Si tratta infatti **di una descrizione delle entità e delle loro relazioni per uno specifico dominio di conoscenza**.

Annotazione funzionale

Il **database Gene Ontology** fornisce una definizione precisa del ruolo svolto dalle singole proteine tramite un vocabolario che consente di definire in modo corretto e non arbitrario il o i **processi biologici** cui una proteina partecipa, le sue **funzioni molecolari** e le sue **localizzazioni cellulari**.

- **Molecular function**: attività a livello molecolare del prodotto genico di interesse;

- **Cellular component**: comparto cellulare in cui il prodotto genico di interesse esercita la sua funzione;

- **Biological process**: processi biologici in cui è coinvolto il prodotto genico di interesse.

Quindi in sintesi la Gene Ontology è composta da un numero di termini definiti esplicitamente, che rappresentano nomi di oggetti o eventi biologici.

Annotazione funzionale

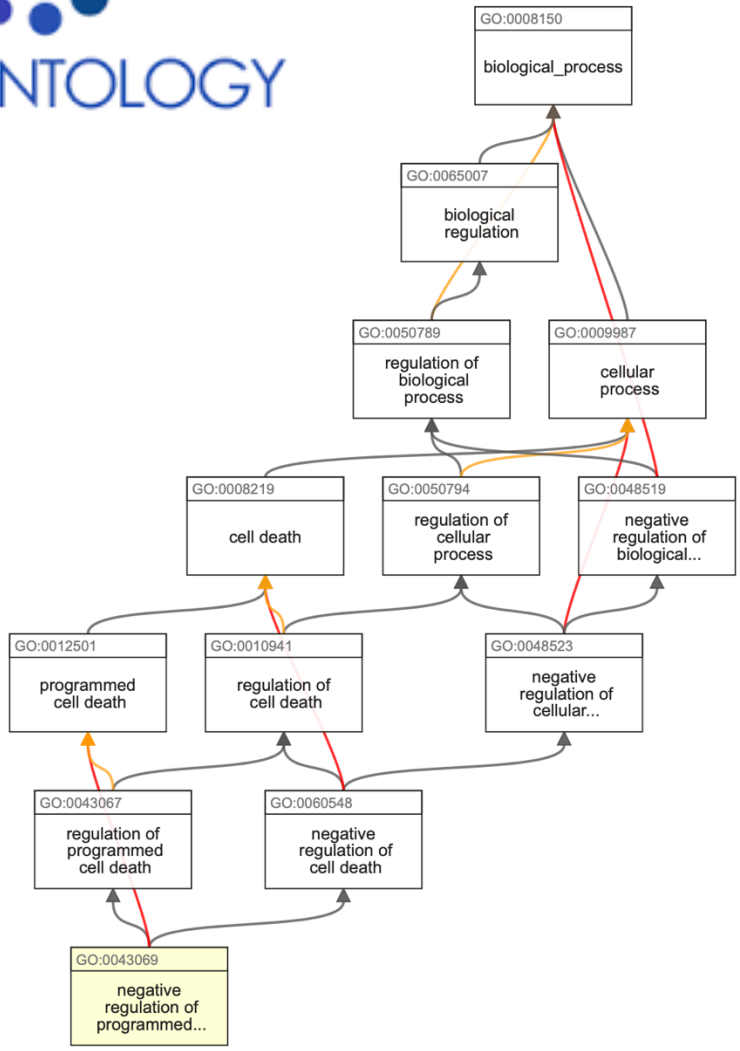
Questi termini sono raffigurati come **nodi** (chiamati anche vertici) in un **grafo** che descrive **le relazioni tra i nodi**.

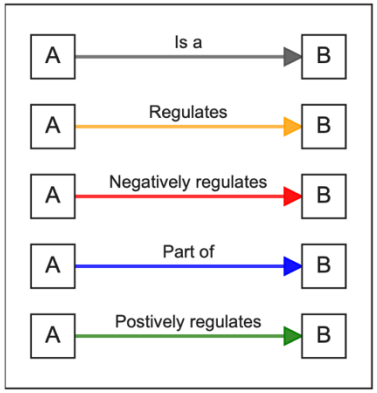
Ad esempio, *"citosol" è un nodo, che è collegato da un arco al suo nodo padre "parte intracellulare". Il tipo di questo arco è "è un" e questa struttura indica semplicemente che il citosol è una parte intracellulare.*

Il grafo formato da questi nodi è un cosiddetto "**grafo aciclico diretto**" o **DAG**.

- ✓ gli archi sono **direzionati**, cioè c'è una sorgente e una destinazione per ciascun arco.
- ✓ la sorgente è definita come termine padre e la destinazione come termine figlio.
- ✓ **non ha cicli**, il che significa che non è possibile completare un ciclo seguendo gli archi direzionati. Questa restrizione implica che due termini non possono essere sia padre che figlio l'uno dell'altro (altrimenti formerebbero un ciclo tra di loro).

GENEONTOLOGY





Annotazione funzionale

Le evidenze usate per l'annotazione nella **Gene Ontology (GO)** si basano su diversi tipi di dati sperimentali, computazionali o di letteratura, e vengono classificate utilizzando codici standard chiamati **Evidence Codes**.

Evidenze Sperimentali

- ✓ **IDA (Inferred from Direct Assay)**: supportata da dati sperimentali diretti, come test biochimici o esperimenti di localizzazione cellulare.
- ✓ **IPI (Inferred from Physical Interaction)**: Basato sull'interazione fisica tra la proteina annotata e altre molecole
- ✓ **IMP (Inferred from Mutant Phenotype)**: Derivata dall'osservazione di fenotipi in mutanti, come knock-out o mutazioni puntiformi.
- ✓ **IGI (Inferred from Genetic Interaction)**: interazioni genetiche, come epistasi o soppressione genetica.
- ✓ **IEP (Inferred from Expression Pattern)**: Basata su dati di espressione genica, ad esempio profili di RNA o proteomica.

Annotazione funzionale

Le evidenze usate per l'annotazione nella **Gene Ontology (GO)** si basano su diversi tipi di dati sperimentali, computazionali o di letteratura, e vengono classificate utilizzando codici standard chiamati **Evidence Codes**.

Evidenze Computazionali

- ✓ **ISS (Inferred from Sequence or Structural Similarity)**: Derivata da similitudini di sequenza o strutturali con proteine di altre specie, attraverso tecniche di omologia o allineamenti di sequenza.
- ✓ **ISO (Inferred from Sequence Orthology)**: Basato sull'ortologia (ossia corrispondenza di geni derivati da un antenato comune) tra diverse specie.
- ✓ **ISA (Inferred from Sequence Alignment)**: Supportata da allineamenti di sequenza che indicano somiglianza funzionale.

Annotazione funzionale



Le evidenze usate per l'annotazione nella **Gene Ontology (GO)** si basano su diversi tipi di dati sperimentali, computazionali o di letteratura, e vengono classificate utilizzando codici standard chiamati **Evidence Codes**.

Evidenze da Letteratura e Inferenze Manuali

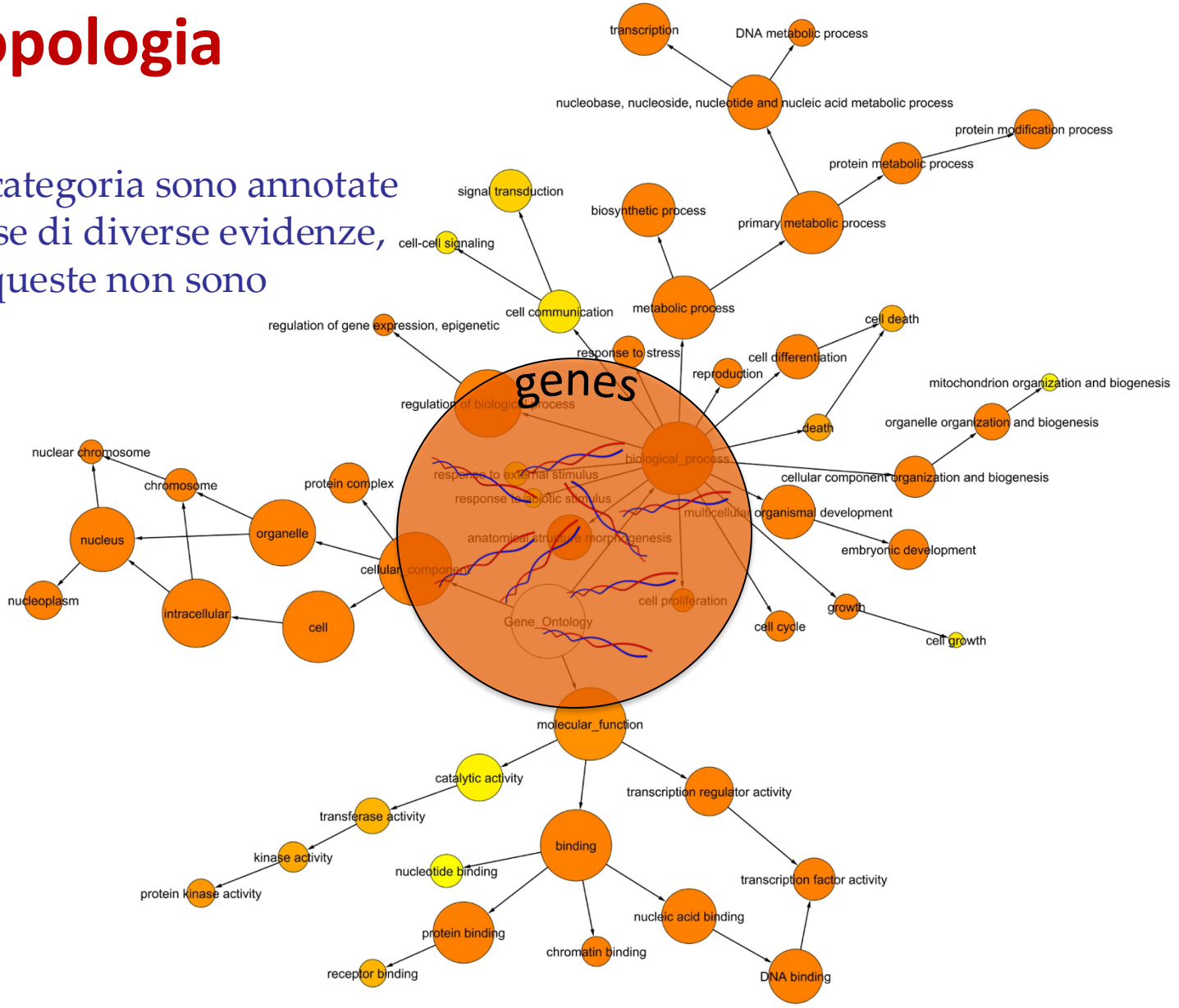
- ✓ **TAS (Traceable Author Statement)**: Annotazioni basate su affermazioni trovate nella letteratura scientifica che sono tracciabili e verificabili.
- ✓ **NAS (Non-traceable Author Statement)**: Derivate da affermazioni presenti in articoli scientifici ma che non forniscono dettagli sufficienti per la verifica diretta.
- ✓ **IC (Inferred by Curator)**: Inferenze fatte direttamente dai curatori basate sulla loro conoscenza e interpretazione delle informazioni disponibili.

Evidenze Automatiche

- ✓ **IEA (Inferred from Electronic Annotation)**: Annotazioni generate automaticamente da algoritmi, senza revisione manuale. Sono tipicamente basate su dati di omologia o altre fonti predittive.

GO: senza topologia

All'interno di ogni categoria sono annotate le proteine, sulla base di diverse evidenze, ma le **relazioni** tra queste non sono riportate.



Pathway biologici: cosa sono

Un **pathway biologico** è una serie di reazioni chimiche o interazioni molecolari che avvengono all'interno di una cellula.

I *pathway* possono riguardare varie funzioni, come la produzione di energia, la comunicazione tra cellule, la risposta a stimoli esterni, o la regolazione del ciclo cellulare. Ne abbiamo di 3 tipi:

1. **Metabolic Pathways (Vie Metaboliche)**
   Includono una serie di reazioni chimiche che trasformano una sostanza in un'altra, spesso per produrre energia o sintetizzare molecole necessarie alla cellula.

2. **Signal Transduction Pathways (Vie di Trasduzione del Segnale)**
   Descrivono il modo in cui una cellula risponde a segnali esterni, come ormoni o molecole di segnalazione. Questi pathway sono fondamentali per la comunicazione tra cellule e per regolare le funzioni cellulari.

3. **Gene Regulatory Pathways (Vie di Regolazione Genica)**
   Coinvolgono l'attivazione o la repressione dell'espressione genica in risposta a vari segnali. Questi pathway determinano quali geni vengono espressi e a quale intensità, permettendo alla cellula di adattarsi a cambiamenti nell'ambiente o di differenziarsi durante lo sviluppo.

Pathway biologici: esistono molti database

A biological pathway is a group of biological entities that are organized in a specified order and perform a specified biological task or function.

(Viswanathan et al., 2008)



http://www.pathguide.org

By Kanehsia Laboratories



By European Bioinformatic Intitute

By BiGCaT Bioinformatics (Maastricht University) and Conklin Lab at the Gladstone Institutes (University of California, San Francisco).





By Computational Biology Research Center of Japan

By National Cancer Institute

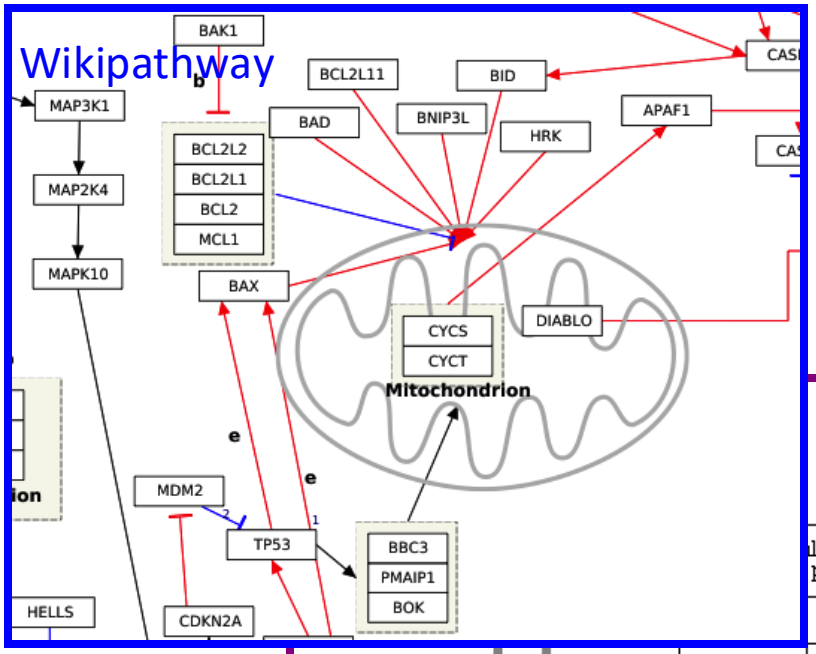




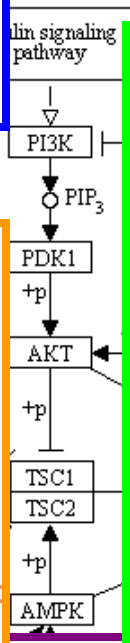
By University of Southern California.

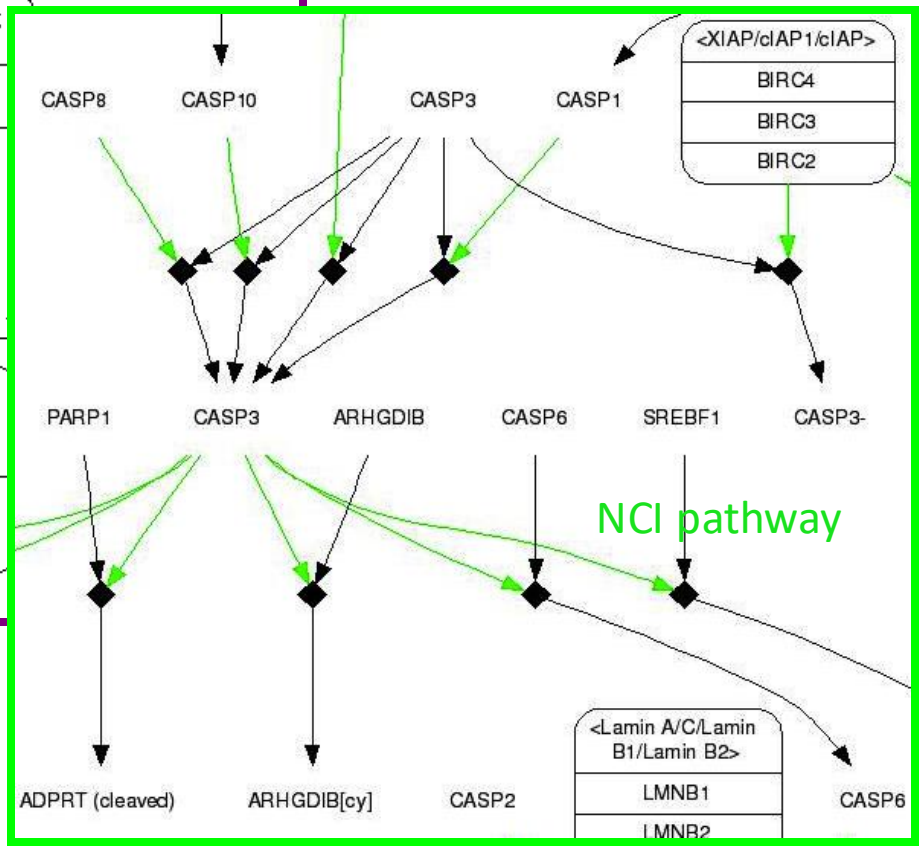
By Invitrogen

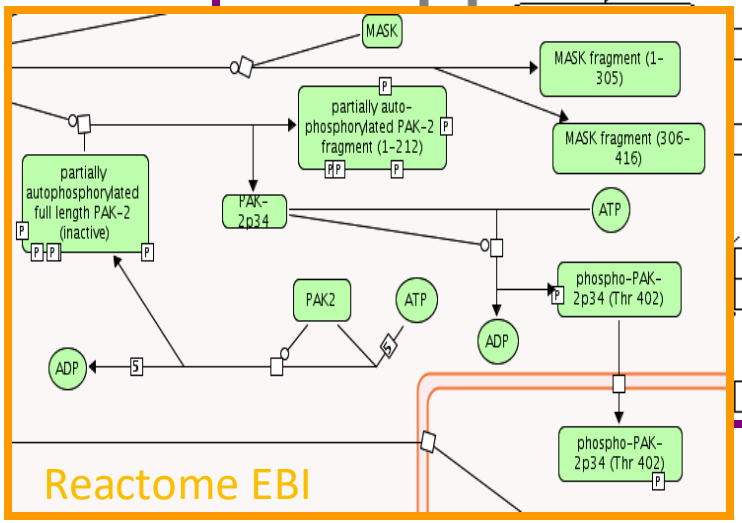






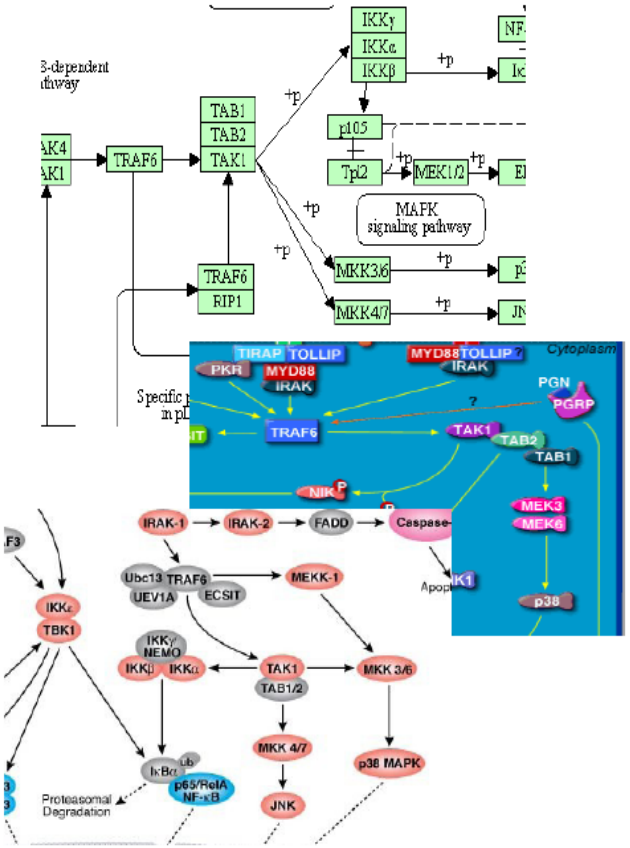






Ognuno con le proprie rappresentazioni grafiche

Graphical representations



SBGN

Machine-readable formats



BioPax SBML ---- KGML(KEGG)
GPML(WikiPathway)

… ma per fortuna ad ogni rappresentazione grafica corrisponde una **rappresentazione codificata in modo standard in linguaggio macchina**