

Databases, data integration and applications

MISB, ISB101 Genomics & Databases course

University of Luxembourg

08 January 2015

Venkata P. Satagopam

(venkata.satagopam@uni.lu)

Research Scientist, Bioinformatics core

Luxembourg Centre for Systems Biomedicine (LCSB)

University of Luxembourg

Agenda

Part 1

Data, databases and data management lecture

Exercises

Part 2

Ensembl, BioMart lecture

Exercises

Part 3

bioCompendium & functionality lecture

Exercises

Part 4

Reflect demo

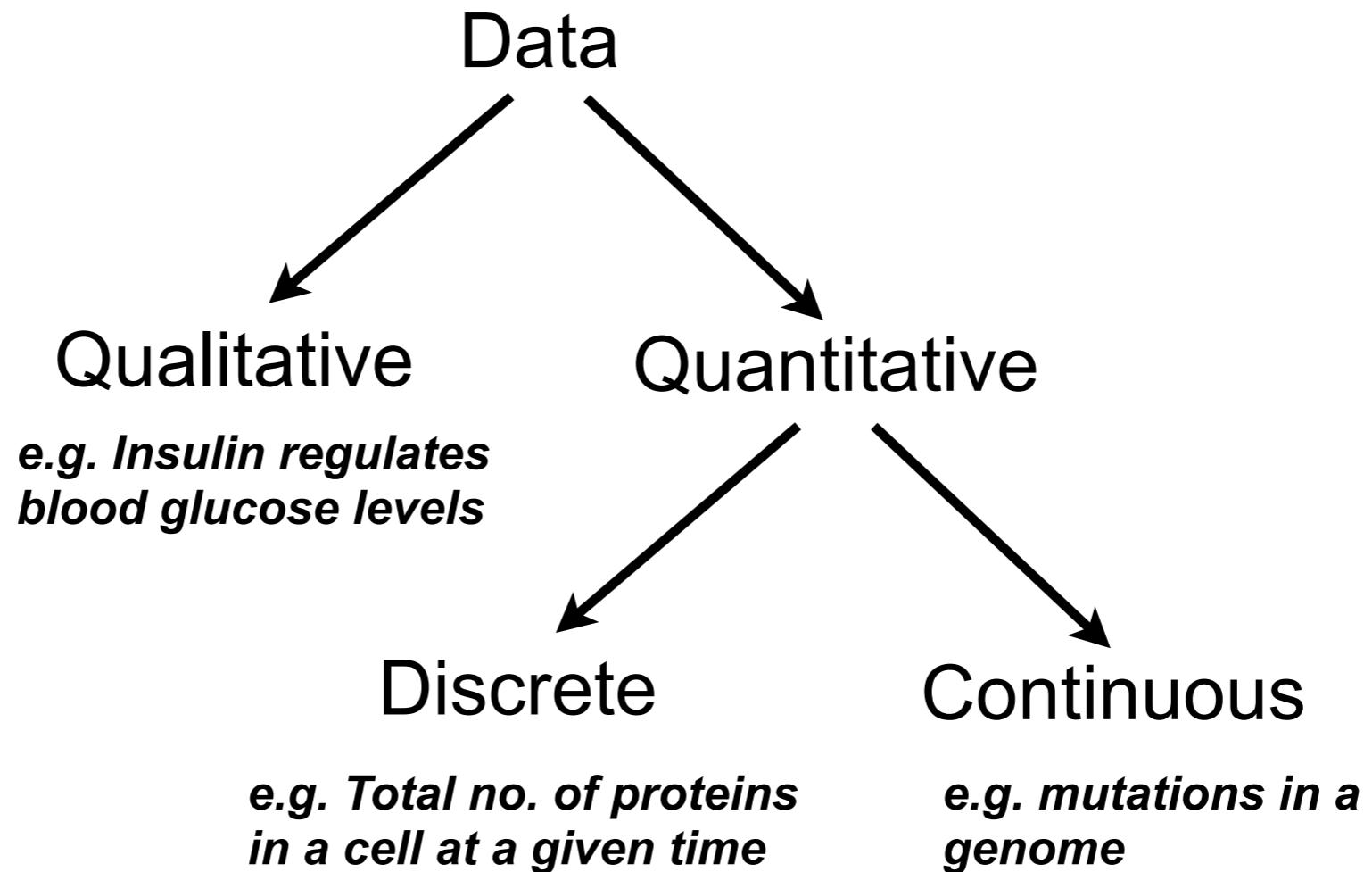
Part 5

Parse EMBL records

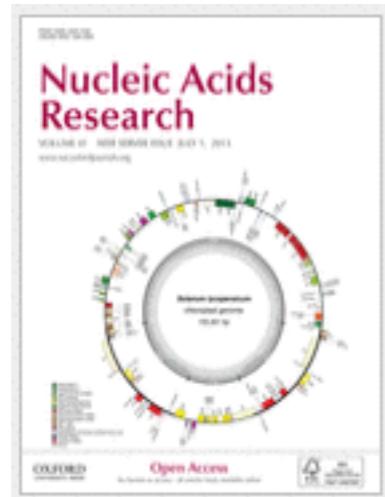
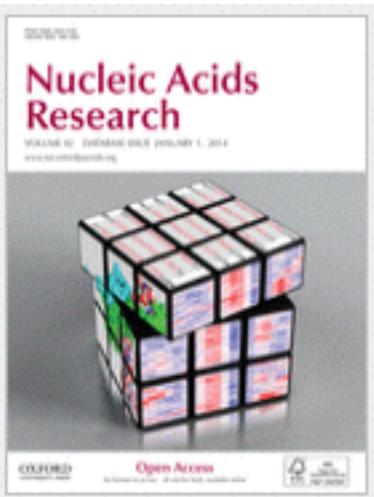
Data & Databases

Data

- Data is a collections of facts, such as values or measurements
- Data can be qualitative or quantitative
- Qualitative data is descriptive
- Quantitative data is numerical



Databases



 **bioinformatics.ca**
links directory

Pathguide

<http://www.oxfordjournals.org/nar/database/c/>
<http://www.oxfordjournals.org/nar/database/cap/>
http://bioinformatics.ca/links_directory
<http://www.pathguide.org>

● 2014 NAR database issue reports on 1552 databases and 2013 web server issue reports on 95 web servers

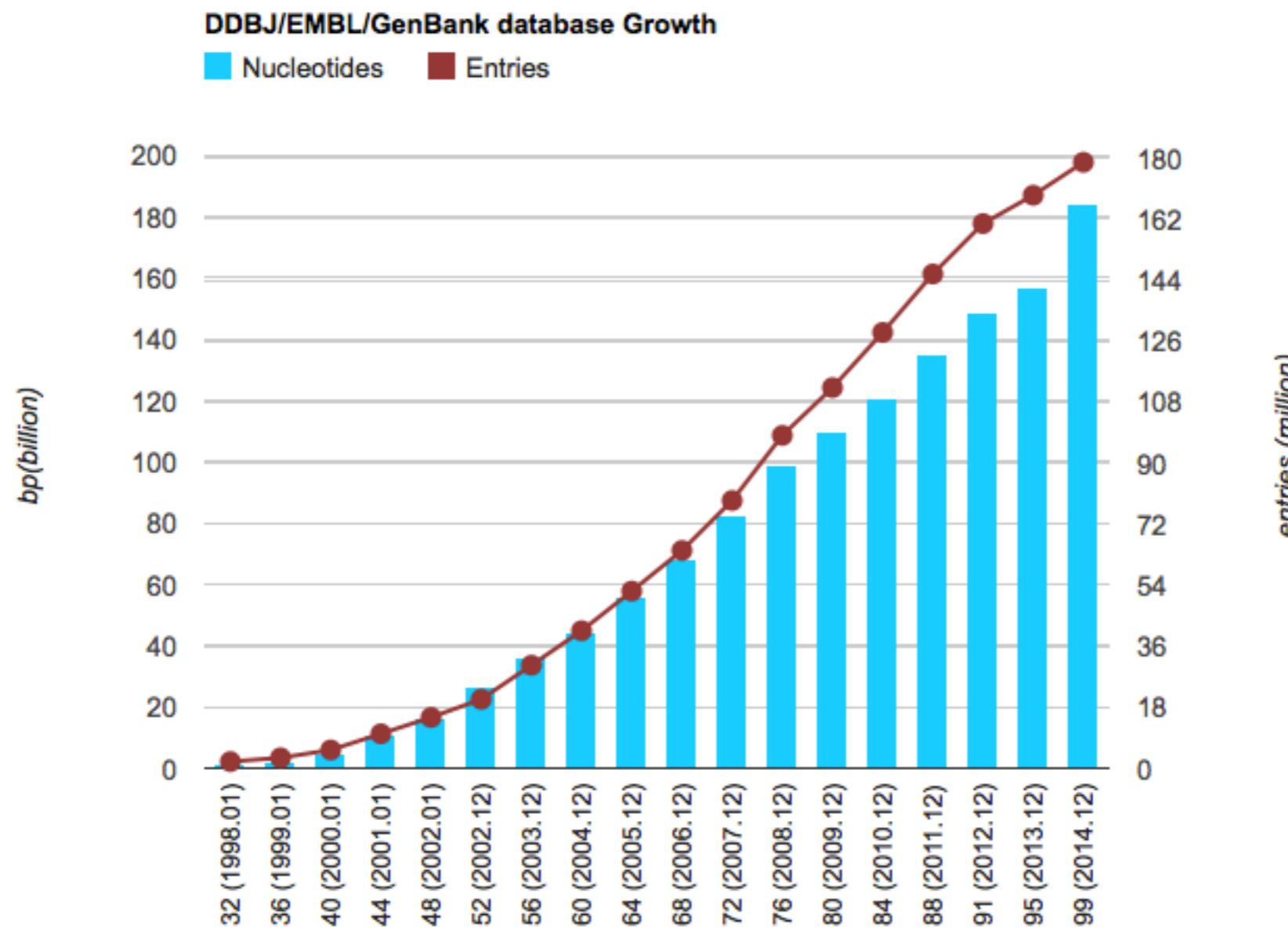
● Currently mentions 174 resources, 623 databases, 1549 tools

● ~550 biological pathway and interaction related resources

Databases



Exponential growth of experimental data



* Note : CON and TPA divisions are not counted in the following Release statistic.

http://www.ddbj.nig.ac.jp/breakdown_stats/dbgrowth-e.html#dbgrowth-graph

SCIENCE IN 2015

The global research enterprise is already vast, whether measured in people, publications, patents or refreshments. And it is getting bigger fast. Assuming current trends continue, here is what the New Year has in store.

BY MARK ZASTROW

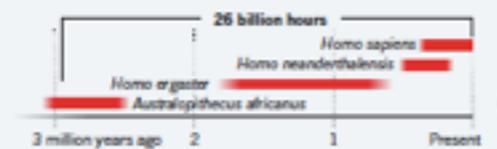
R&D SPENDING

Global investments in research and development have been growing by some 7% annually for more than a decade.



This year, the total should roughly equal the gross domestic product of India.

~26 BILLION RESEARCHER HOURS



With some 10 million researchers averaging 50-hour working weeks, the global human effort going into R&D is equivalent to 2.9 million years — a span of time that would extend back to the heyday of *Australopithecus africanus*.

~234,000

COOKIES CONSUMED AT McMURDO STATION

These easy-to-grab treats are a favourite fuel for those long hours in the laboratory. In Antarctica's McMurdo Station alone, cookies are devoured at the rate of 4,500 per week.

~1 BILLION CUPS OF COFFEE

Scientists consume this essential stimulant in vast quantities. The order-of-magnitude estimate of global consumption given here may well be on the low side, but it is enough to give every resident of the European Union a cup and a refill.

~2.6 MILLION PATENTS FILED

~1.2 MILLION PATENTS GRANTED

These key measures of innovation will have more than doubled since 2000.

2 DWARF PLANETS VISITED

NASA's Dawn spacecraft will arrive at the dwarf planet Ceres, located in the asteroid belt between Mars and Jupiter, in March. The agency's New Horizons spacecraft will fly by the dwarf planet Pluto in July.

~920,000 PUBLICATIONS

~470 RETRACTIONS

The world's output of scientific articles is growing at about 2.8% per year. Only a tiny fraction are ever retracted — but that number also seems to be growing.

~260,000 NEW PhD HOLDERS WORLDWIDE

This is a crowd that could fill Beijing's 80,000-seat National Stadium more than 3 times over. If these new degree holders were magically reassembled into a single 'Gradzilla' with the same volume, the resulting creature would be over 100 metres tall.



© Nature

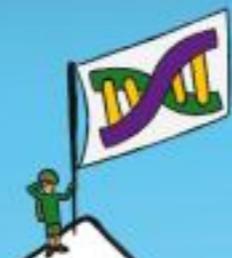
ARTWORK BY CHRIS RYAN/NATURE



SCIENCE IN 2015

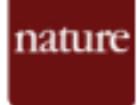
go.nature.com/rdnns2

nature



**~1.5 TRILLION
BASES FILED IN GENBANK**

The online sequence-data repository has been doubling in size every two years or so, and the pace is picking up. By December it could reach the size of 500 human genomes — and will still represent only a fraction of the sequence data stored worldwide.



Nature News&Comment @NatureNews · Jan 5

2015 #science in numbers - from patents to publications, cash to caffeine consumption ow.ly/GNxMk
pic.twitter.com/u4y1Rts0MA



114

58

...

Databases

- Exponential growth of experimental data
- Very heterogeneous in content and format point of view
- Content: genes, proteins, chemicals, diseases, literature, interactions, pathways etc

Gene Databases

EMBL Nucleotide Sequence Database

EMBL-EBI 

Enter Text Here Terms of Use | Privacy | Cookies

Databases Tools Research Training Industry About Us Help Site Index  

- ENA Home
- EMBL-Bank Home
- Access
- Documentation
- News
- Submission
- Publications
- People
- Contact

EMBL Fetch 
Fetch an EMBL record by id

News 
5th January 2010: INSDC and Genome Reference Consortium discussed in Bioinform... [more](#)

Collaborations 

- [INSDC](#) - International Nucleotide Sequence Database Collaboration
- [NCBI](#) - The Nucleotide Sequence Database is produced in collaboration with GenBank (USA)
- [DDBJ](#) - The Nucleotide Sequence Database is also produced in collaboration with the DNA Database of Japan (DDBJ)

EBI > Databases > EMBL-Bank

EMBL Nucleotide Sequence Database

The EMBL Nucleotide Sequence Database (also known as EMBL-Bank) constitutes Europe's primary nucleotide sequence resource. Main sources for DNA and RNA sequences are [direct submissions](#) from individual researchers, genome sequencing projects and patent applications.

The database is produced in an international [collaboration](#) with GenBank (USA) and the DNA Database of Japan (DDBJ). Each of the three groups collects a portion of the total sequence data reported worldwide, and all new and updated database entries are exchanged between the groups on a daily basis. The [current database release](#) (Release 111, March 2012), with according [Release notes](#) and [user manual](#) are available from the EBI servers. A sample database entry is shown [here](#).

A publication in [Nucleic Acids Research 2009 37: D19-D25](#), provides further information and details.

The EMBL nucleotide sequence database forms part of the [European Nucleotide Archive](#), an EBI project led by [Guy Cochrane](#) as part of the [The Protein and Nucleotide Database Group \(PANDA\)](#) under [Ewan Birney](#).

Link	Explanation
Access	Database queries , Completed genomes webserver , FTP archives (EMBL release, alignments etc), EMBL sequence version archive (SVA), Browse by geography .
Submission	Primary sequence submissions, third party annotation, updates.
Documentation	Release notes user manual , Information for Submitters , FAQ , Release information , Forthcoming Changes , EMBL database statistics , Feature table , XML documentation , Sample entry , Examples of annotation , EMBL Features & Qualifiers , DE line standards , Database Policies
Publications	Group publications
People	Group members
Contact	How to contact the EMBL Nucleotide Sequence Database
News	List of recent changes on this site

Contact

For information, comments and/or suggestions, please use the EBI Support Form page
<http://www.ebi.ac.uk/support/>

11

ENA European Nucleotide Archive

[Services](#)[Research](#)[Training](#)[About us](#)

European Nucleotide Archive

[ENA Home](#) | [Search & Browse](#) | [Submit & Update](#) | [About ENA](#) | [Contact](#) | [FAQ](#)

ENA >

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation ... [more](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

Text search

[Advanced Search](#)

Sequence Search

[Advanced Search](#)

- ENA Home
- Search & Browse
- Submit & Update
- About ENA
- Contact
- FAQ

NEWS AND ANNOUNCEMENTS

[Change to date format for advanced search](#)

23 May 2014

From 16th June 2014, the date format used in the advanced search will be changed to ISO format (YYYY-MM-DD).

[Update to the ENA SAMPLE checklist](#)

20 May 2014

From 10th of June 2014 the ENA SAMPLE checklist XML will be updated and the older version will be deprecated.

[View all news](#)

KEGG Genes



KEGG GENES Database

Molecular building blocks of life in the genomic space

[KEGG2](#) [PATHWAY](#) [BRITE](#) [MODULE](#) [KO](#) [GENOME](#) [GENES](#) [SSDB](#) [Organisms](#)

Enter org:gene (Example) syn:ssr3451

Gene Catalogs

KEGG GENES is a collection of gene catalogs for all complete genomes (see [release history](#)) generated from publicly available resources, mostly NCBI RefSeq. They are subject to SSDB computation and KO assignment (gene annotation) by KOALA tool. KEGG DGENES for draft genomes of some eukaryotes and KEGG EGENES for EST datasets of mostly plants are supplementary gene catalogs, which are given automatic KO assignment by KAAS with GENES used as a reference data set. There is now a fourth type of gene catalogs, MGENES for metagenomes (see also [KEGG GENOME](#)) with automatic annotation. The viral gene catalog, VGENES, is not yet fully integrated in the KEGG system.

Gene catalog	Category	Remark
GENES	Complete genomes	High-quality genomes with KOALA and manual annotations
DGENES		Draft genomes with automatic (KAAS) annotation
EGENES	EST datasets	EST contigs with automatic (KAAS) annotation
MGENES	Metagenomes	Metagenomes with automatic (KAAS) annotation
VGENES	Viruses	No annotation; available only in DBGET

Search for

bfind mode bget mode

Search for

bfind mode bget mode



NCBI Gene / EntrezGene

TP53 tumor protein p53 [Homo sapiens] – Gene – NCBI

http://www.ncbi.nlm.nih.gov/gene/7157

Apple Yahoo! GMail m-w YouTube Wikipedia News (524) Popular toi TL HIV-Portal Synonym DB2 bioCompendium Syr

NCBI Resources How To

Gene Gene Limits Advanced

Display Settings: Full Report

Send to:

TP53 tumor protein p53 [*Homo sapiens*]

Gene ID: 7157, updated on 3-Jun-2012

Summary

Official Symbol TP53 provided by [HGNC](#)

Official Full Name tumor protein p53 provided by [HGNC](#)

Primary source [HGNC:11998](#)

See related [Ensembl:ENSG00000141510](#); [HPRD:01859](#); [MIM:191170](#); [Vega:OTTHUMG00000162125](#)

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as P53; LFS1; TRP53; FLJ92943

Summary This gene encodes tumor protein p53, which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. p53 protein is expressed at low level in normal cells and at a high level in a variety of transformed cell lines, where it's believed to contribute to transformation and malignancy. p53 is a DNA-binding protein containing transcription activation, DNA-binding, and oligomerization domains. It is postulated to bind to a p53-binding site and activate expression of downstream genes that inhibit growth and/or invasion, and thus function as a tumor suppressor. Mutants of p53 that frequently occur in a number of different human cancers fail to bind the consensus DNA binding site, and hence cause the loss of tumor suppressor activity. Alterations of this gene occur not only as somatic mutations in human malignancies, but also as germline mutations in some cancer-prone families with Li-Fraumeni syndrome. Multiple p53 variants due



GeneCards



GeneCards®

The Human Gene Compendium

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE



with

XENOME
a division of
LifeMap
SCIENCES

Free for academic non-profit institutions. ALL other users need a [commercial license from LifeMap Sciences, Inc.](#)

[Home](#)[GeneCards Guide](#)[Suite](#)[Terms and Conditions](#)[About Us](#)[User Feedback](#)[Mirror sites](#)

Search the GeneCards human gene database

[f Like](#) 138 [g +1](#) 40

Search by keyword(s)

for

[Search](#)

[Advanced Search](#)

[About V3 Search](#)

GeneCards Suite

Explore GeneCards

About GeneCards®: GeneCards is a searchable, integrated, database of human genes that provides concise genomic related information, on all known and predicted human genes. [more...](#)

Extract information for many genes at once:

[GeneALaCart](#)

[GeneDecks](#)

[Hot genes](#)

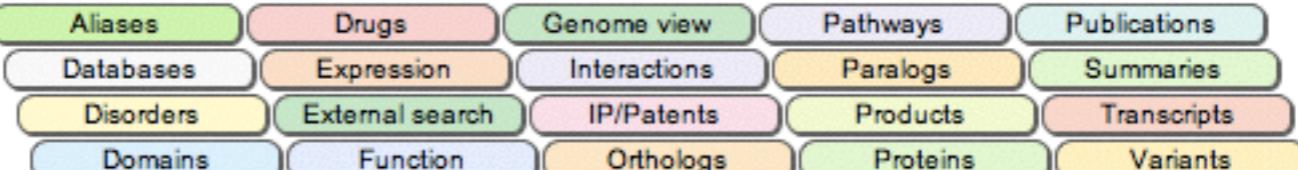
[Disease genes](#)

View Sample Gene

[GeneCards Sections](#)

ERBB2

v-erb-b2 erythroblastic
leukemia viral oncogene
homolog 2



View Random Gene

[Category](#)

RNA genes

[Go](#)

[GIFTs Group](#)

Medium

SNORA78

(GIFTs: 23)

small nucleolar RNA, H/ACA box 78

News and Views

Version 3.08

20 May 2012

[New Features](#)

[more pathways](#)
[more RNA genes](#)
[mouse/rat products](#)
[more...](#)

In our pipeline:

MalaCards

Collaborations:

[SysKid](#)
[SPIRE/MOPED](#)
[RNACentral](#)

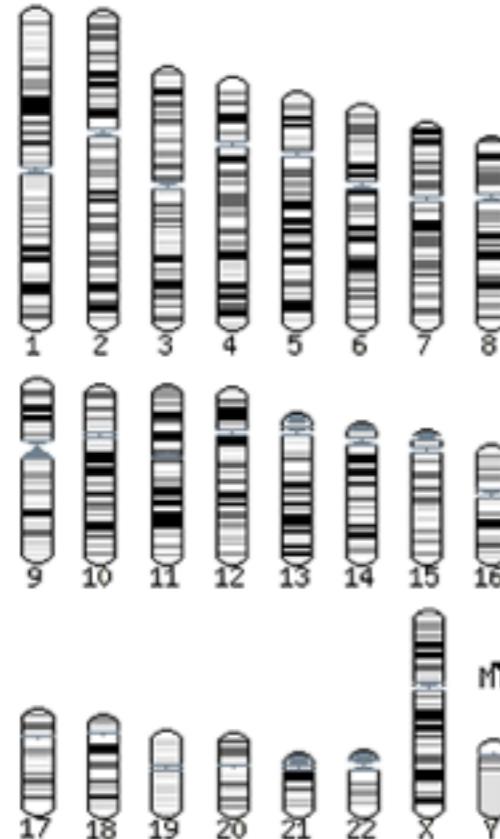
[User comments](#)

[LifeMap PR](#)

[Site Map](#)

3.08.060 7 June '12

[Previous site](#)



Browse approved symbols by chromosome

The HUGO Gene Nomenclature Committee (HGNC) has assigned unique gene symbols and names to over 33,000 human loci, of which around 19,000 are protein coding. [genenames.org](#) is a curated online repository of HGNC-approved gene nomenclature and associated resources including links to genomic, proteomic and phenotypic information, as well as dedicated gene family pages.

Quick Gene Search

Search symbols, keywords or IDs for:

Results that equal begin contain

Display hits

[Search Genes](#)

FAQ

[What is the HGNC?](#)

[What is HGNC-approved nomenclature and why do we need it?](#)

[Where can I find information about existing human gene symbols?](#)

[What is a stem symbol?](#)

Latest News



[KATNAL1 has a role in spermiogenesis](#)

A study in mice has shown that the KATNAL1 ortholog is needed for the final stages of sperm development. The...



MGI has a [job opening](#) for a biologist.



Mouse Genome Informatics

[Search](#) ▾ [Download](#) ▾ [More Resources](#) ▾ [Submit Data](#) [Find Mice \(IMSR\)](#) [Analysis Tools](#) [Contact Us](#)



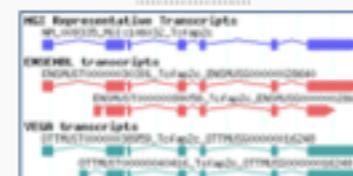
Keywords, Symbols, or IDs

[Quick Search](#)

Explore MGI

[All Search Tools](#)

Genes



Recombinases (cre)



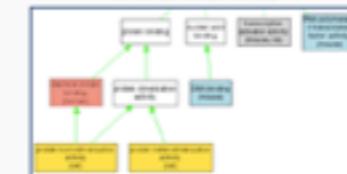
Strains / SNPs

Variation Type	B6CJ	PRBLU	BL6CBLU	Allele Summary (all strains)
SNP	G G	A	A/G	
SNP	C C	T	C/T	

Phenotypes & Disease Models



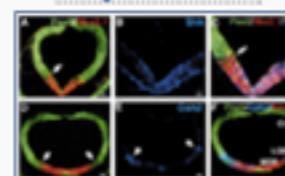
Function



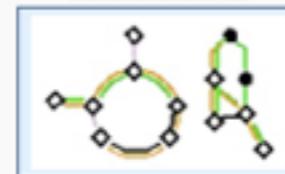
Orthology



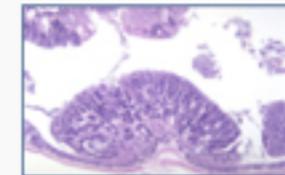
Expression



Pathways



Tumors



FAQs

How do I...

- .. search for genes by genomic interval? [FAQ](#)
- .. find mutations for phenotypes or diseases? [FAQ](#)
- .. find expression data? [FAQ](#)
- .. view a structural genomic map? [FAQ](#)

News

April 11, 2012

- The search forms for querying all MGI references and the GXD gene expression literature contain new features. [Read more...](#)
- The International Mouse Strain Resource (IMSR) features new search and performance enhancements. [Read more...](#)
- Marker detail pages now feature human orthology data, including associated human diseases. [Read more...](#)





New & Noteworthy

[Cancer's Chromosomal Chaos Explained \(Partly\)](#)

06/01/2012

One reason cancer is so tricky to treat has to do with its adaptability. It can quickly try out new genetic combinations until it hits upon one that can survive whatever treatment a doctor is currently throwing at it. The result is return of the cancer after remission. One way cancer is able to change its genetics so rapidly has to do with chromosome instability. The number of chromosomes in a cancer cell is much less...[read more >](#)

[Yeast, Place your Bets](#)

05/25/2012

Life is a balancing act. An organism needs to grow and divide as fast as possible in its current environment. But it also needs to be able to survive when the environment changes. One way nature has come up with to deal with this balancing act is called bet hedging. Basically some members in a population grow well in one set of circumstances and another set grows well in another. Now this makes obvious sense when looking...[read more >](#)

[Alternatives to Whale Puke](#)

05/11/2012

Imagine scouring the beaches of the world for balls of whale vomit. People are willing to do this because finding one is like finding a huge gold nugget. Perfume companies will pay around \$10,000 per pound for this ambergris (which is the more scientific name for the stuff). Of course perfume companies would rather have a more reliable and less expensive source for their ambergris. And it wouldn't hurt to find one



About SGD

The *Saccharomyces* Genome Database (SGD) provides comprehensive integrated biological information for the budding yeast *Saccharomyces cerevisiae* along with search and analysis tools to explore these data, enabling the discovery of functional relationships between sequence and gene products in fungi and higher organisms.

Upcoming Meetings

[General Meeting of the American Society for Microbiology](#)

June 16, 2012 - San Francisco, CA

Abstract deadline: January 17, 2012

[EMBO Conference 2012 on Gene Transcription in Yeast "From Mechanisms to Gene Regulatory Networks"](#)

June 16, 2012 - St. Feliu de Guíxols, Girona, Spain

Abstract deadline: March 16, 2012

[Model Organisms to Human Biology: Cancer Genetics](#)


[Home](#) [Help](#) [Contact](#) [About Us](#) [Login/Register](#)
[Search](#) [Browse](#) [Tools](#) [Portals](#) [Download](#) [Submit](#) [News](#) [ABRC Stocks](#)

The Arabidopsis Information Resource

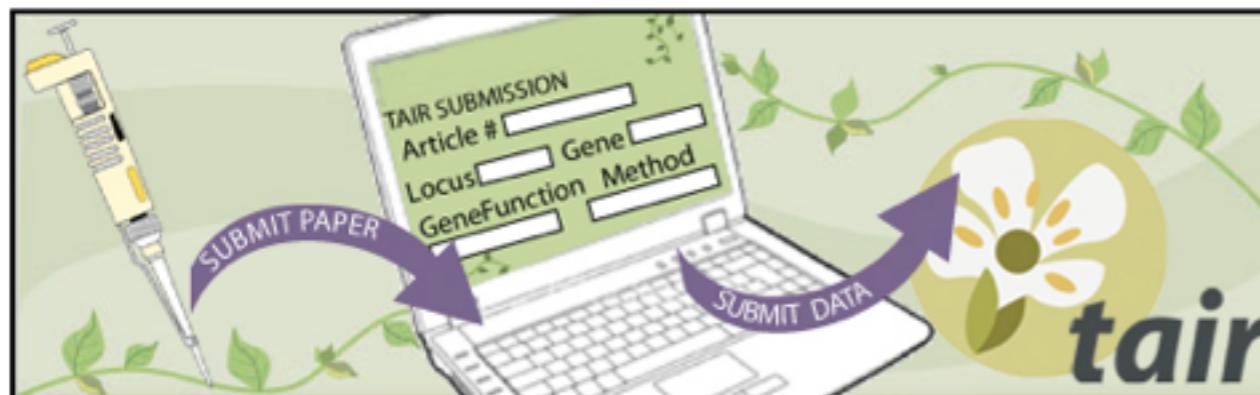
The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every two weeks from the latest published research literature and community data submissions. Gene structures are updated 1-2 times per year using computational and manual methods as well as community submissions of new and updated genes. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

The Arabidopsis Biological Resource Center at The Ohio State University collects, reproduces, preserves and distributes seed and DNA resources of *Arabidopsis thaliana* and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.

CARNEGIE TAIR is located at the Carnegie Institution for Science Department of Plant Biology and funded by the National Science Foundation with additional support from TAIR sponsors.



Updates on TAIR funding are available [here](#).



Click here
to try our new online submission form
and submit the molecular function (e.g. protein kinase), biological

Breaking News

[Subscribe to news feed](#)

[Follow our Twitter feed](#)

[Join our Facebook group](#)

New Protein Chip and Cell Cultures at ABRC

[May 9, 2012]

A new protein chip ([AtProteinChip 2](#)) developed by [M. Snyder](#) and [S.P. Dinesh-Kumar](#), is now available. Cell line PSB-D ([CCL84840](#)), donated by Geert De Jaeger and derived from MM2d developed by M. Menges and J. Murray, is also available. These stocks can be found using the [ABRC catalog](#).

Share Your Education Resources [February 1, 2012]

Do you teach an undergraduate Lab using Arabidopsis? Can you imagine your research project on Arabidopsis illustrating an important scientific concept? Contribute to moving Arabidopsis from the bench to the classroom!

GO Annotations At TAIR

[January 25, 2012]

TAIR integrates [GO annotations](#) for *A. thaliana* from the research community, UniProtKB, the Gene Ontology Consortium, BIOGRID and IntAct. Our community /127



WikiGenes – Collaborative Publishing

http://www.wikigenes.org/ wigi genes

Apple Yahoo! GMail m-w YouTube Wikipedia News (528) Popular toi TL HIV-Portal Synonym DB2 bioCompendium

wikigenes

collaborative publishing

Hoffmann, R. A wiki for the life sciences where authorship matters. *Nature Genetics*.

[Sign In / Create account]

About

- Introduction
- Tutorial
- Contact us
- Language [English]

Share

- Send to a friend
- Share

Personal Info

- Your homepage
- Your articles
- Create article

Links

- NCBI entrez
- Uniprot
- IHOP resource
- Nature
- Science
- Wikipedia

watch the guided tour  WikiGenes Introduction & Tutorial

Search thousands of genes, chemicals, pathologies and much more... advanced search

[WikiGenes Search]

POWERED WITH
AUTHORSHIP TRACKING TECHNOLOGY 

Simply click in the text to find out who wrote what. Fair credit for authors. Always know your sources.

mememoir.org

society in science
The Branco Weiss Fellowship

Switzerland

Queen's

Italy

Purdue

Michigan

Université du Luxembourg

Protein Databases

[Search](#)[Blast](#)[Align](#)[Retrieve](#)[ID Mapping](#)

Search in

Protein Knowledgebase (UniProtKB)

Query

[Search](#)[Advanced Search »](#)[Clear](#)

THE FIRST 10 YEARS OF UNIPROT

Saturday 8 September 2012, Basel, Switzerland

Celebrate UniProt's 10th anniversary with us! Renowned speakers from the fields of interactions & protein modelling, proteomics, protein structure & function, and genome analysis & annotation will highlight how protein databases are underpinning life sciences.

Submit abstracts for talks and posters to abstractsUP12@isb-sib.ch

WELCOME

The mission of [UniProt](#) is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed. Includes complete and reference proteome sets .
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
 data	Literature citations, taxonomy, keywords, subcellular locations, cross-referenced databases and more.

NEWS



UniProt release 2012_05 - May 16, 2012

Sex by deception | Update to Reference proteomes in UniProtKB

- › Statistics for UniProtKB:
[Swiss-Prot](#) · [TrEMBL](#)
- › Forthcoming changes
- › News archives

 [Follow @uniprot](#) 327 followers

SITE TOUR



Learn how to make best use of the tools and data on this site.

PROTEIN SPOTLIGHT

on sex, drugs and satisfaction May 2012

Pleasure is not a human invention. Experiences that arouse a feeling of contentment are as old as life. They have, in fact, kept life going. It is yet another of Mother Nature's tricks...



Chemical Databases

PubChem

[BioAssay](#) [Compound](#) [Substance](#)

[GO](#) [Advanced search](#)

[Chemical structure search](#) | [BioActivity analysis](#)

New The PubChem BioAssay summary pages now display the [Gene Ontology \(GO\)](#) classification of the gene/protein target(s) that were tested by the bioassay.

In addition, the bioassay data summary for a compound now displays a graphical summary of the bioassays that have tested the compound, categorizing the bioassays by bioactivity outcomes, top targets, bioactivity types, and bioassay types. [See more..](#)

New A new RESTful web interface to PubChem data and services (PUG REST) is released in beta form. [See more..](#)

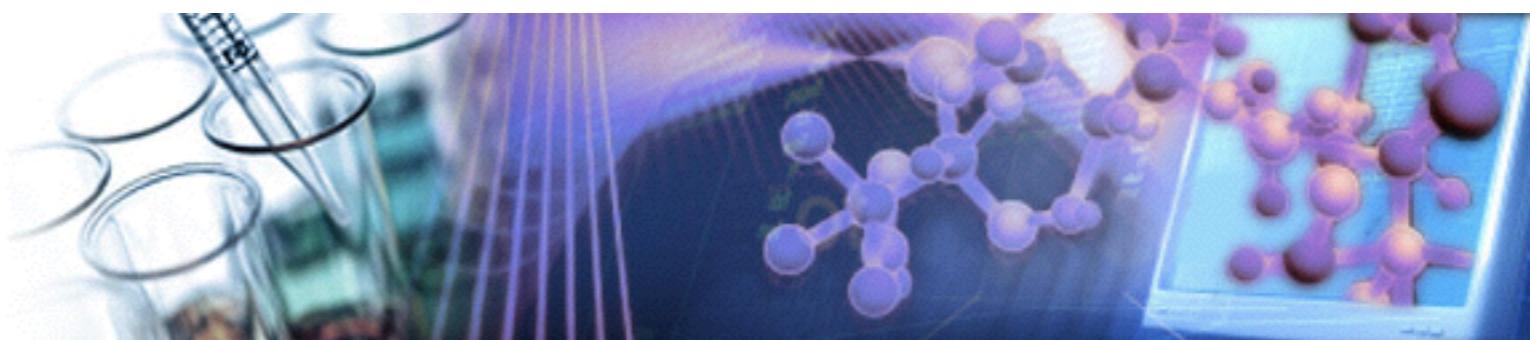
[more ...](#) 

Bioactivity summary	
Bioactivity datatable	
Bioactivity structure-activity	
Chemical structure search	
3D conformer viewer	
Chemical structure clustering	
Deposition gateway	
Structure download	
Bioassay download	
PubChem FTP	

[Write to Helpdesk](#) | [Disclaimer](#) | [Privacy Statement](#) | [Accessibility](#) | [Data Citation Guidelines](#)
National Center for Biotechnology Information
[NLM](#) | [NIH](#) | [HHS](#)

DRUGBANK

Open Data Drug & Drug Target Database



Home Browse Search Downloads News & Updates About Help Contact Us

Search: [Help / Advanced](#)

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 6711 drug entries including 1447 FDA-approved small molecule drugs, 131 FDA-approved biotech (protein/peptide) drugs, 85 nutraceuticals and 5080 experimental drugs. Additionally, 4227 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 150 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.



DrugBank is supported by [David Wishart](#), Departments of [Computing Science](#) & [Biological Sciences](#), [University of Alberta](#).

DrugBank is also supported by [The Metabolomics Innovation Centre](#), a Genome Canada-funded core facility serving the scientific community and industry with world-class expertise and cutting-edge technologies in metabolomics.

[More about DrugBank](#)



ChEMBL

ChEMBLdb

ChEMBL-NTD

Kinase SARfari

GPCR SARfari

DrugEBility

ChEMBL Group

Downloads

Web Services

FAQ

ChEMBLdb Statistics

- DB: ChEMBL_13
- Targets: 8,845
- Compound records: 1,296,266
- Distinct compounds: 1,143,682
- Activities: 6,933,068
- Publications: 44,682

ChEMBL Blog



26 | June - June

EBI > Databases > Small Molecules > ChEMBL Database > Home



OPEN KNOWLEDGE

OPEN SERVICE

Search ChEMBLdb...

Compounds

Targets

Assays

[Activity Source Filter](#)

ChEMBLdb

Compound Search

Protein Target Search

Browse Targets

Browse Drugs

Drug Approvals

ChEMBL is a database of bioactive drug-like small molecules, it contains 2-D structures, calculated properties (e.g. logP, Molecular Weight, Lipinski Parameters, etc.) and abstracted bioactivities (e.g. binding constants, pharmacology and ADMET data).
[read more](#)

Getting Started

- Search [target data](#) via keyword, protein sequence search (BLAST), or by navigating the target classification hierarchy.
- Search [compound data](#) with lists of keywords, SMILES strings, compound identifiers, or by drawing the chemical structure.
- Search [assay data](#) via keyword search using the main search bar.

Support and Feedback

We positively encourage [feedback on the interface](#) and search capabilities, since this will shape our future development.
[read more](#).

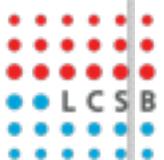
Staying in Touch

To keep up to date with ChEMBL news and data releases subscribe to the [ChEMBL-announce mailing list](#).
[read more](#).

Training

The group run a series of [webinars](#) detailing the interface and schema.
[read more](#).

Data Licensing



LCSB

Human Metabolome Database Version 2.5

Search:

[\[Advanced\]](#)



The Human Metabolome Database (HMDB) is a freely available electronic database containing detailed information about small molecule metabolites found in the human body. It is intended to be used for applications in metabolomics, clinical chemistry, biomarker discovery and general education. The database is designed to contain or link three kinds of data: 1) chemical data, 2) clinical data, and 3) molecular biology/biochemistry data. The database (version 2.5) contains over 7900 metabolite entries including both water-soluble and lipid soluble metabolites as well as metabolites that would be regarded as either abundant ($> 1 \mu\text{M}$) or relatively rare ($< 1 \text{nM}$). Additionally, approximately 7200 protein (and DNA) sequences are linked to these metabolite entries. Each MetaboCard entry contains more than 110 data fields with 2/3 of the information being devoted to chemical/clinical data and the other 1/3 devoted to enzymatic or biochemical data. Many data fields are hyperlinked to other databases (KEGG, PubChem, MetaCyc, ChEBI, PDB, Swiss-Prot, and GenBank) and a variety of structure and pathway viewing applets. The HMDB database supports extensive text, sequence, chemical structure and relational query searches. Four additional databases, [DrugBank](#), [T3DB](#), [SMPDB](#) and [FooDB](#) are also part of the HMDB suite of databases. [DrugBank](#) contains equivalent information on ~1500 drugs, [T3DB](#) contains information on 2900 common toxins and environmental pollutants, [SMPDB](#) contains pathway diagrams for 350 human metabolic and disease pathways, while [FooDB](#) contains equivalent information on ~2000 food components and food additives.

The HMDB is supported by [David Wishart](#), Departments of [Computing Science & Biological Sciences](#), [University of Alberta](#).

The HMDB is also supported by [The Metabolomics Innovation Centre](#), a Genome Canada-funded core facility serving the scientific community and industry with world-class expertise and cutting-edge technologies in metabolomics.

[More about the HMDB](#)

MATADOR: Manually Annotated Targets and Drugs Online Resource

Please enter your search below, or browse the [list of drugs](#) or [proteins](#).

About MATADOR

MATADOR is a resource for protein-chemical interactions. It differs from other resources such as [DrugBank](#) in its inclusion of as many direct and indirect interactions as we could find. In contrast, DrugBank usually contains only the main mode of interaction. The manually annotated list of direct (binding) and indirect interactions between proteins and chemicals was assembled by automated text-mining followed by manual curation. Each interaction contains links to PubMed abstracts or OMIM entries that were used to deduce the interaction. (These articles are not necessarily useful review articles.)

Indirect interactions are caused by many different mechanisms. For example, binding a metabolite of a drug as well as changes in gene expression fall under that category. In order to capture as many interactions as possible, all the different mechanisms are grouped together. You as the user can decide if you rather trust only the direct interactions (with a known mechanism) or also indirect interactions.

How to cite

Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, Ahmed J, Urdiales EG, Gewiess A, Jensen LJ, Schneider R, Skoblo R, Russell RB, Bourne PE, Bork P, Preissner R.

[SuperTarget and Matador: resources for exploring drug-target relationships.](#)

Nucleic Acids Res. 2008 Jan;36(Database issue):D919-22. Epub 2007 Oct 16.



SIDER 2 Side Effect Resource

SIDER contains information on marketed medicines and their recorded adverse drug reactions. The information is extracted from public documents and package inserts. The available information include side effect frequency, drug and side effect classifications as well as links to further information, for example drug–target relations.

Please enter your search for drugs or side effects below and click 'Search'.

 Search

Database statistics

Number of drugs and side effects

# of SE	# of drugs	# of drug-SE pairs	Pairs with frequency information
4199	996	100049	39.6%

Number of drug–side effect pairs in different frequency ranges

	frequent (with exact data)	infrequent (with exact data)	rare (with exact data)	postmarketing	total
drug	11283 (10138)	9347 (3185)	6564 (2039)	20977	39603
placebo	4257 (4257)	2019 (2019)	1397 (1397)	0	6262

Version Information

The current version has been released on March 16, 2012. This release uses the MedDRA dictionary and provides access to preferred terms and lower-level terms. The number of drugs has increased from 888 to 996. SIDER 1 is still available via [FTP](#).

Diseases



All Databases

OMIM

Online Mendelian Inheritance in ManJohns
Hopkins
UniversityMy NCBI
[\[Sign In\]](#) [\[Register\]](#)

PubMed

Nucleotide

Protein

Genome

Structure

PMC

OMIM

Search OMIM

for

Go

Clear

Limits

Preview/Index

History

Clipboard

Details

- Enter one or more search terms.
- Use **Limits** to restrict your search by search field, chromosome, and other criteria.
- Use **Index** to browse terms found in OMIM records.
- Use **History** to retrieve records from previous searches, or to combine searches.

NCBI is implementing changes to help you find current content in OMIM based on resources at NCBI, and then directing you to omim.org. Please be aware that you will leave NCBI to view OMIM records. Access to full records from NCBI (e.g. web, ftp, eutils) will no longer be supported.

OMIM® - Online Mendelian Inheritance in Man®

Welcome to OMIM®, Online Mendelian Inheritance in Man®. OMIM is a comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 12,000 genes. OMIM focuses on the relationship between phenotype and genotype. It is updated daily, and the entries contain copious links to other genetics resources.

This database was initiated in the early 1960s by Dr. Victor A. McKusick as a catalog of mendelian traits and disorders, entitled Mendelian Inheritance in Man (MIM). Twelve book editions of MIM were published between 1966 and 1998. The online version, OMIM, was created in 1985 by a collaboration between the National Library of Medicine and the William H. Welch Medical Library at Johns Hopkins. It was made generally available on the internet starting in 1987. In 1995, OMIM was developed for the World Wide Web by NCBI, the National Center for Biotechnology Information.

OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh.

Disease databases

HuGENet – Gene-disease association databases

http://www.hugenet.org.uk/resources/databases/table1.html

Apple Yahoo! GMail m-w YouTube Wikipedia News (528) Popular toi TL HIV-Portal Synonym DB2 bioCompendium

Human Genome Epidemiology Network UK Coordinating Centre

HuGENet

ABOUT US REVIEWS TRAINING RESOURCES METHODS FIELD SYNOPSES LINKS CONTACT US

Introduction Handbook Protocols Informatics Databases Networks Home

Table 1: Databases containing information on gene-disease associations in more than one gene

Information on these or any other internet-based resources of gene-disease association data should be emailed to databases@hugenet.org.uk.

Name of Website	Website URL	Brief Description	Summary
Alzgene	http://www.alzforum.org/res/com/gen/alzgene	Regularly updated collection of published genetic association studies performed on Alzheimer Disease phenotypes, from database searches and journals' contents lists. Case and control data presented. Performs crude meta-analysis of odds ratios on request.	Alzgene Summary
Cytokine Gene Polymorphism in Human Disease	http://www.nanea.dk/cytokinesnps/	Regularly updated database with Medline-based records from a systematic review of cytokine gene polymorphisms associated with human disease. Data extracted from two publications about the study.	Cytokine Gene Polymorphism in Human Disease Summary
HuGE Navigator	http://hugenavigator.net	HuGE Navigator provides access to a continuously updated knowledge base in human genome epidemiology, including information on population prevalence of genetic variants, gene-disease associations, gene-gene and gene-environment interactions, and evaluation of genetic tests.	HuGE Navigator Summary
GenAtlas	http://www.genatlas.org	Regularly updated database of genes, phenotypes and references. Among numerous databases are brief sections on disorders associated with genes, with lists of citations. May be biased towards statistically significant results.	GenAtlas Summary
	http://genecanvas.idf.inserm.fr	Database of cardiovascular candidate genes available through a search interface.	GeneCanvas

32

UNIVERSITÉ DU LUXEMBOURG

LCSB

Literature

Home – PubMed – NCBI

<http://www.ncbi.nlm.nih.gov/pubmed>

RSS

Apple Yahoo! GMail m-w YouTube Wikipedia News (529) Popular toi TL HIV-Portal Synonym DB2 bioCompendium

NCBI Resources How To My NCBI Sign In

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed Advanced Search Help

PubMed

PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

[PubMed Quick Start Guide](#)
[Full Text Articles](#)
[PubMed FAQs](#)
[PubMed Tutorials](#)
[New and Noteworthy](#)

PubMed Tools

[PubMed Mobile](#)
[Single Citation Matcher](#)
[Batch Citation Matcher](#)
[Clinical Queries](#)
[Topic-Specific Queries](#)

More Resources

[MeSH Database](#)
[Journals in NCBI Databases](#)
[Clinical Trials](#)
[E-Utilities](#)
[LinkOut](#)

You are here: NCBI > Literature > PubMed

[Write to the Help Desk](#)

GETTING STARTED



tion
anual
ook

Training & Tutorials

RESOURCES

Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures

POPULAR

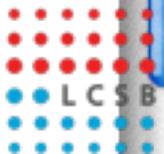
PubMed
Nucleotide
BLAST
PubMed Central

FEATURED

Genetic Testing Registry
PubMed Health
GenBank
Reference Sequences

NCBI INFORMATION

About NCBI
Research at NCBI
NCBI Newsletter
NCBI FTP Site



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

Google

Apple Yahoo! GMail m-w YouTube Wikipedia News (529) Popular toi TL HIV-Portal Synonym DB2 bioCompendium

iHOP

information hyperlinked
over proteins

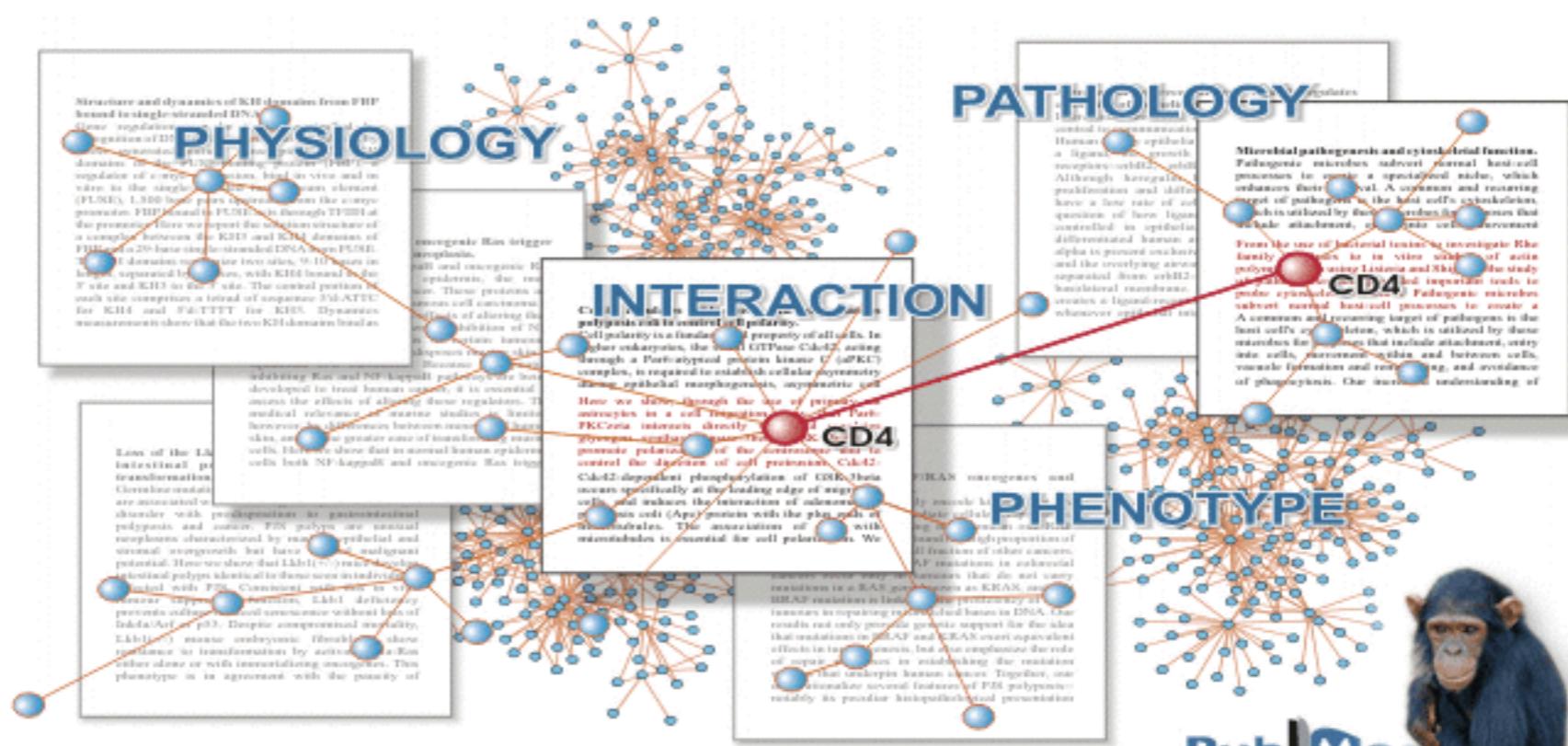
Search Gene

Gene Model
Developer's Zone new
How to cite iHOP

Contact
Links
Help



cocoon



Hoffmann, R., Valencia, A. A Gene Network for Navigating the Literature. *Nature Genetics* 36, more than 2,700 organisms, 110,000 genes, 28.4 million sentences.
...always up to date – every day.

PubMed

Search for a gene synonym or accession number... (Click here for an example: SNF1)

all fields

in

All organisms

[SEARCH]

Interactions

Home · Download · Help/Info



STRING - Known and Predicted Protein-Protein Interactions

[search by name](#) [search by protein sequence](#) [multiple names](#) [multiple sequences](#)

protein name: (examples: #1 #2 #3)

(STRING understands a variety of protein names and accessions; you can also try a [random entry](#))

organism:

auto-detect

interactors wanted:

COGs

Proteins

Reset

GO !

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



PubMed mips ...

STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

More Info

[Funding / Support](#)

[Acknowledgements](#)

[Use Scenarios](#)

STRING (*Search Tool for the Retrieval of Interacting Genes/Proteins*) is being developed at [CPR](#), [EMBL](#), [SIB](#), [KU](#), [TUD](#) and [UZH](#).

STRING references: [Szklarczyk et al. 2011](#) / [2009](#) / [2007](#) / [2005](#) / [2003](#) / [Snel et al. 2000](#).

Miscellaneous: [Access Statistics](#), [Robot Access Guide](#), [STRING/STITCH Blog](#), [Supported Browsers](#).

What's New? This is version 9.0 of STRING - now covering more than 1100 organisms (and counting) !

Sister Projects: check out [STITCH](#) and [eggNOG](#) - two sister projects built on STRING data!

Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our [old releases](#).

STITCH: chemical association networks

http://stitch.embl.de/ Google

Apple Yahoo! GMail m-w YouTube Wikipedia News (529) Popular toi TL HIV-Portal Synonym DB2 bioCompendium >

STITCH 3.1

[Input Page](#) | [Downloads](#) | [Help/Info](#)

search by name chemical structure protein sequence multiple names multiple sequences

name: (examples: #1 #2 #3)

(STITCH understands a variety of chemical/protein names, accessions and InChIKeys; you can also try a [random entry](#))

organism: auto-detect
(In case of chemicals, the organism with the most likely interaction is chosen)

Reset **GO !**

please enter your protein or chemical of interest...

What is STITCH?

STITCH is a resource to explore known and predicted interactions of chemicals and proteins. Chemicals are linked to other chemicals and proteins by evidence derived from experiments, databases and the literature.

STITCH contains interactions for between 300,000 small molecules and 2.6 million proteins from 1133 organisms.

Interaction network around ethanol:

CO

Mini-Tutorial

To explore the interactions of the beta blocker propanolol, [enter its name](#) or an identifier like the ATC Code [C07AA05](#) in the search box to the left. [Select 'Homo Sapiens'](#) as organism. When you click [GO!](#), you will be taken to the network with various functionally related proteins: adrenergic receptors (the primary targets), serotonin receptors (which are also targeted), etc. Click on "Actions View" to [explore the actions](#) (activation/inhibition/binding) of propanolol. When you click on one of the adrenergic receptors, you can [add it to the set of input nodes](#). This will result in a network containing both proteins and related chemicals.

38

Info Funding / Support Source Databases

UNIVERSITÉ DU LUXEMBOURG

Search Tool for Interactions of Chemicals) is being developed at CPR, EMBL, SIB, KU, TUD and UZH.

LCSB

GeneMania

GeneMANIA » search for CDC7 in H. sapiens

Find genes in **H. sapiens (human)** related to **cdc7**

Showing 20 related genes

File ▾ View ▾ Query ▾ Networks legend Functions legend

Networks Genes Functions Help

Sort by: name, percent weight
Expand: all, only top level, none
Enable all, none

Physical interactions 35.59 %
Co-expression 34.08 %
Predicted 16.53 %
Pathway 7.07 %
Co-localization 5.32 %
Shared protein domains 1.13 %
Genetic interactions 0.27 %

39

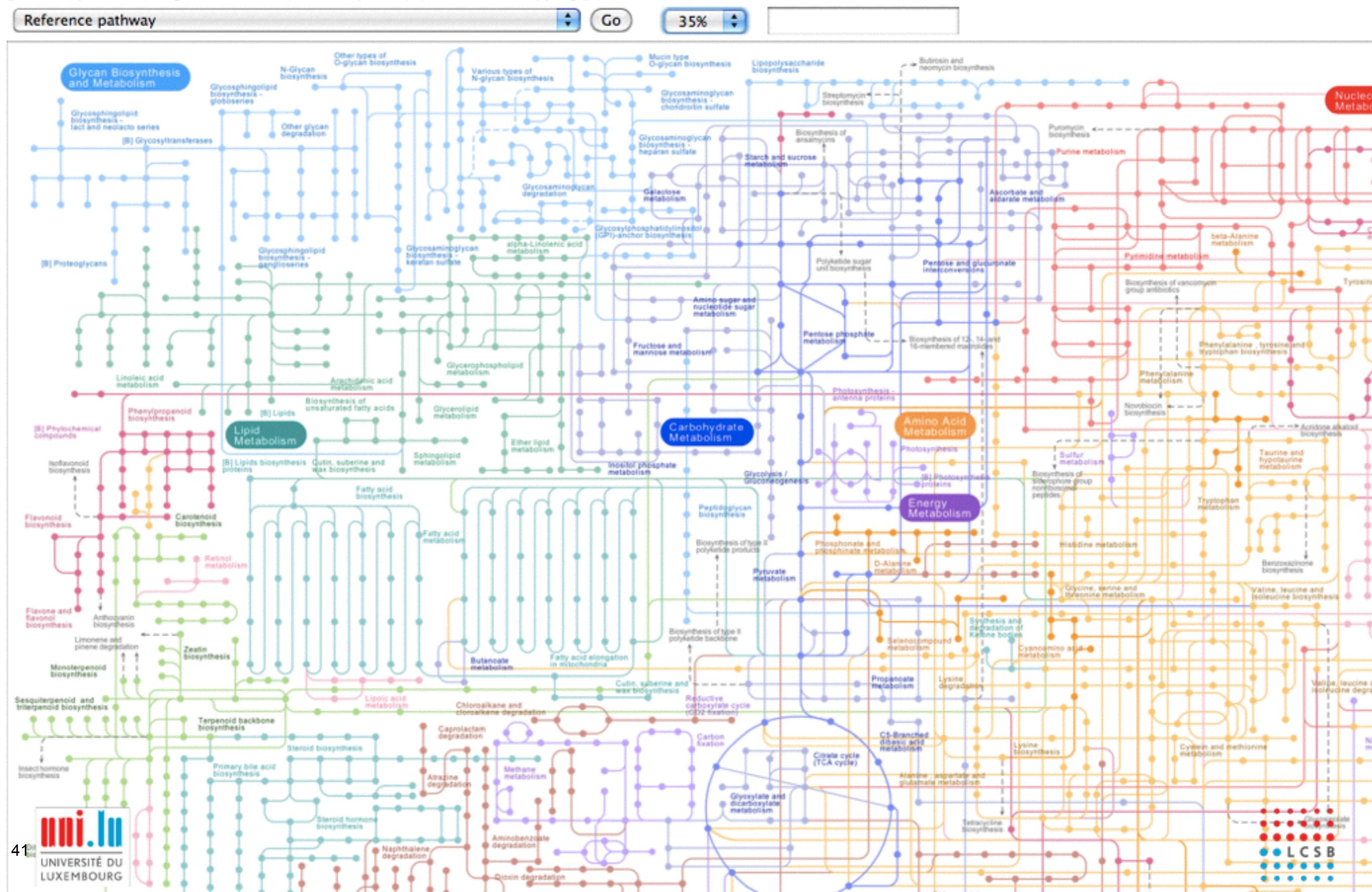
Pathways

[Pathway menu | Organism menu | Pathway entry | User data mapping]

Reference pathway

Go

35%



Panther

PANTHER - Pathway Diagram

PANTHER Classification System

CHOLESTEROL BIOSYNTHESIS ?

Export: Select one

SBML **SBGN**

created with CellDesigner

Click components to make selections. Right-click components for more options.

Search selections for: Genes Go

Pathway Diagram Pathway Description

File Edit Zoom Select Color

Model Molecules

Interactive Standard/Activity Flow SBGN Image View

Standard View Activity Flow

Acetyl CoA Acetoacetyl CoA

H₂O CoA-SH

r26 Hydroxymethyl glutaryl CoA synt

3-hydroxy-3-methyl-glutaryl CoA

2NADPH + H⁺ NADP + CoA-SH

Hydroxymethyl glutaryl CoA redu

Mevalonate Lanosterol Cholesterol

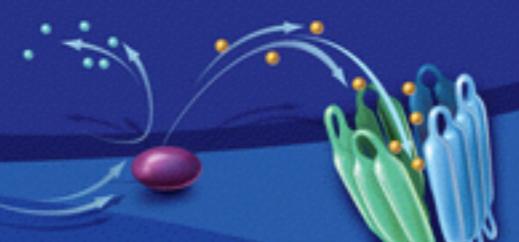
If the applet is not loading correctly, click [here](#) to download Java Plug-in.

Diagram Legend [View complete legend](#)

42

UNIVERSITÉ DU LUXEMBOURG

LCSB



Home About Content Documentation Tools Download Contact Us Outreach

Search examples...

Browse Pathways

Map IDs to Pathways

Compare Species

Analyse Expression Data

If you would prefer to use our old website, click here.

Download

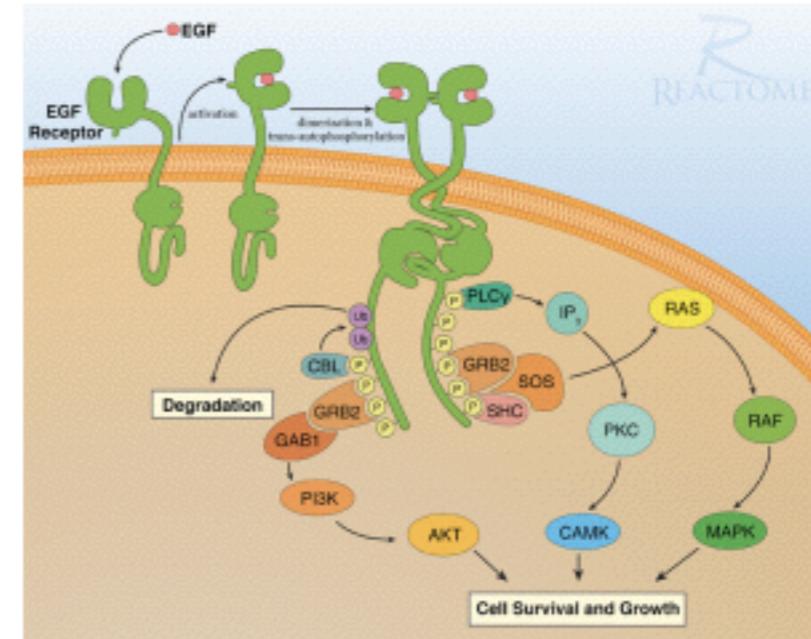
The following links allow you to download Reactome data in various formats:

- BioPax
- SBML
- Textbook
- Other formats

About Reactome

REACTOME is an open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. These include NCBI Entrez Gene, Ensembl and UniProt databases, the UCSC and HapMap Genome Browsers, the KEGG Compound and ChEBI small molecule databases, PubMed, and Gene Ontology. ... [more]

Featured pathway: Signalling by EGFR



Click image to see pathway

Tutorial



News and Notes

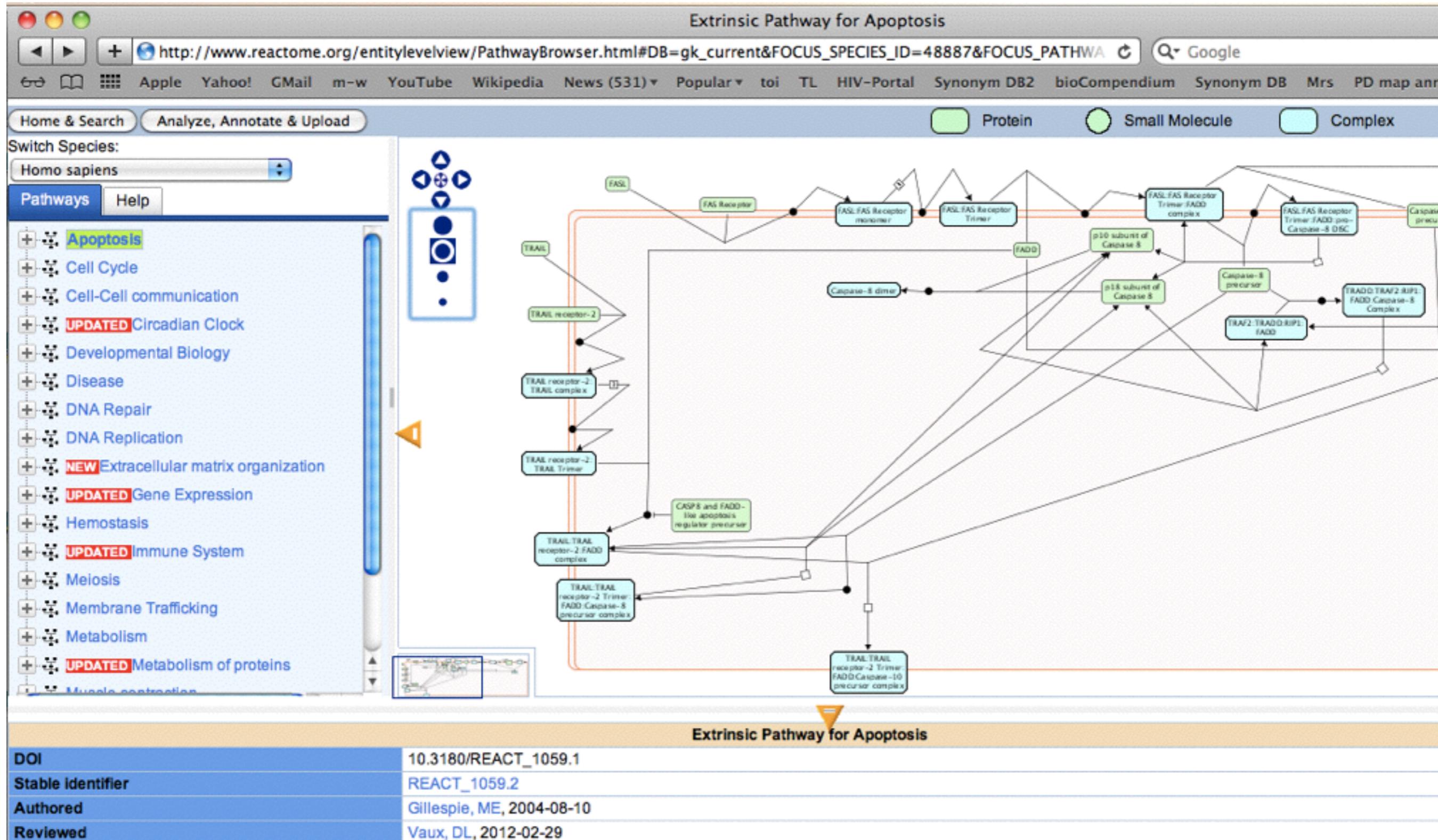
- Mar 22, 2012: Version 40 Released

The pathway Extracellular matrix organization is new and includes Activation ... [more]

- Mar 16, 2012: Reactome will be participating in Google Summer of Code 2012

The Genome Informatics group, organizing the joint efforts of Galaxy, CBioPortal, ... [more]

Reactome



Known as the "death receptor pathway" the extrinsic or caspase 8/10 dependent pathway is activated by ligand binding. The "death receptors" are specialized cell-surface receptors including Fas/CD95 alpha (TNF-alpha) receptor 1, and two receptors, DR4 and DR5, that bind to the TNF-alpha related apoptosis-inducing ligand (TRAIL). The extrinsic and intrinsic pathways unite in the activation of Caspase-3. Pathways communicate through the pro-apoptotic Bcl-2 family member Bid before uniting at the shared activation of Caspase-3.

Databases

- Exponential growth of experimental data
- Very heterogeneous in content and format point of view
- Content: genes, proteins, chemicals, diseases, literature, interactions, pathways etc
- Format: flat files, RDBMS (e.g. Oracle, MySQL, PostgreSQL, DB2), XML

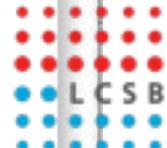
File formats - flat file

http://www.uniprot.org/uniprot/P42858.txt

[+]<http://www.uniprot.org/uniprot/P42858.txt> Google

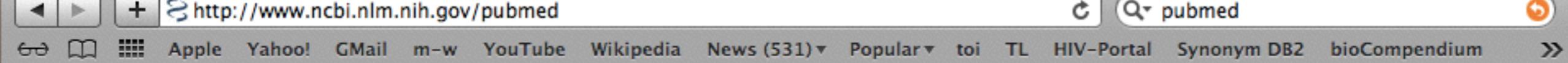
Apple Yahoo! GMail m-w YouTube Wikipedia News (531) Popular toi TL HIV-Portal Synonym DB2 bioCompendium >

ID HD_HUMAN Reviewed; 3142 AA.
AC P42858; Q9UQB7;
DT 01-NOV-1995, integrated into UniProtKB/Swiss-Prot.
DT 18-MAY-2010, sequence version 2.
DT 16-MAY-2012, entry version 125.
DE RecName: Full=Huntingtin;
DE AltName: Full=Huntington disease protein;
DE Short=HD protein;
GN Name=HTT; Synonyms=HD, IT15;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
OC Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [mRNA].
RC TISSUE=Retina;
RX MEDLINE=93208892; PubMed=8458085;
RA Macdonald M., Ambrose C.M., Duyao M.P., Myers R.H., Lin C.S.,
RA Srinidhi J., Barnes G., Taylor S.A., James M., Groot N., McFarlane H.,
RA Jenkins B., Anderson M.A., Wexler N.S., Gusella J.F., Bates G.P.,
RA Baxendale S., Hummerich H., Kirby S., North M., Youngman S., Mott R.,
RA Zehetner G., Sedlacek Z., Poustka A., Frischauf A.-M., Lehrach H.,
RA Buckler A.J., Church D., Doucette-Stamm L., O'Donovan M.C.,
RA Riba-Ramirez L., Shah M., Stanton V.P., Strobel S.A., Draths K.M.,
RA Wales J.L., Dervan P., Housman D.E., Altherr M., Shiang R.,
RA Thompson L., Fielder T., Wasmuth J.J., Tagle D., Valdes J., Elmer L.,
RA Allard M., Castilla L., Swaroop M., Blanchard K., Collins F.S.,
RA Snell R., Holloway T., Gillespie K., Datson N., Shaw S., Harper P.S.;
RT "A novel gene containing a trinucleotide repeat that is expanded and
unstable on Huntington's disease chromosomes.";
RL Cell 72:971-983(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE [mRNA].
RC TISSUE=Brain;
RX MEDLINE=20469406; PubMed=11013077; DOI=10.1006/geno.2000.6317;
RA Matsuyama N., Hadano S., Onoe K., Osuga H., Shouguchi-Miyata J.,
RA Gondo Y., Ikeda J.-E.;
R Identification and characterization of the miniature pig Huntington's
R base gene homolog: evidence for conservation and polymorphism in
R CAG triplet repeat.";
R Biologics 69:72-85(2000).



File formats - XML

http://www.ncbi.nlm.nih.gov/pubmed



```
<PubmedArticle>
  <MedlineCitation Status="Publisher" Owner="NLM">
    <PMID Version="1">22672625</PMID>
    <DateCreated>
      <Year>2012</Year>
      <Month>6</Month>
      <Day>7</Day>
    </DateCreated>
    <Article PubModel="Print-Electronic">
      <Journal>
        <ISSN IssnType="Electronic">1756-0500</ISSN>
        <JournalIssue CitedMedium="Internet">
          <Volume>5</Volume>
          <Issue>1</Issue>
          <PubDate>
            <Year>2012</Year>
            <Month>Jun</Month>
            <Day>6</Day>
          </PubDate>
        </JournalIssue>
        <Title>BMC research notes</Title>
      </Journal>
      <ArticleTitle>Human gene correlation analysis (HGCA): A tool for the identification of transcriptionally
      <Pagination>
        <MedlinePgn>265</MedlinePgn>
      </Pagination>
      <Abstract>
        <AbstractText>ABSTRACT: BACKGROUND: Bioinformatics and high-throughput technologies such as microarra
      </Abstract>
      <AuthorList>
        <Author>
          <LastName>Michalopoulos</LastName>
          <ForeName>Ioannis</ForeName>
          <Initials>I</Initials>
        </Author>
        <Author>
          <LastName>Pavlopoulos</LastName>
          <ForeName>Georgios A</ForeName>
          <Initials>GA</Initials>
        </Author>
        <Author>
          <LastName>Malatras</LastName>
```

File formats - RDBMS

Ensembl Genome Browser

http://www.ensembl.org/index.html

RSS ensembl

Apple Yahoo! GMail m-w YouTube Wikipedia News (531) Popular toi TL HIV-Portal Synonym DB2 bioCompendium

Login · Register

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | More ▾

Search: All species for

e.g. BRCA2 or rat X:100000..200000 or coronary heart disease

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize this list](#))

 **Human**
GRCh37

 **Mouse**
NCBIM37

 **Zebrafish**
Zv9

All genomes

-- Select a species --

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

UNIVERSITÉ DU LUXEMBOURG 

Ensembl is a joint project between  EMBL, EBI and the 

New to Ensembl?

Did you know you can:

-  [Learn how to use Ensembl](#)
with our video tutorials and walk-throughs
-  [Add custom tracks](#)
using our new Control Panel
-  [Upload and analyse your data](#)
and save it to your Ensembl account
-  [Search for a DNA or protein sequence](#)
using BLAST or BLAT
-  [Fetch only the data you want](#)
from our public database, using the Perl API
-  [Download our databases via FTP](#)
in FASTA, MySQL and other formats
-  [Mine Ensembl with BioMart](#)
and export sequences or tables in text, html, or Excel format

Still got questions? Try our [FAQs](#) or [glossary](#)

Did you know...?

Lost?
Try our [tutorials](#) page.

What's New in Release 67 (May 2012)

- [New species: Nile Tilapia \(*Oreochromis niloticus*\)](#)
- [Import of 1000 genomes data](#)
- [Multi-species view renamed](#)
- [New Ensembl virtual machine](#)

Databases

- Exponential growth of experimental data
 - Very heterogeneous in content and format point of view
 - Content: genes, proteins, chemicals, diseases, literature, interactions, pathways etc
 - Format: flat files, RDBMS (e.g. Oracle, MySQL, PostgreSQL, DB2), XML
-
- Most of the Systems biology (SB) data is represented in XML-based markup languages e.g.
 - SBML, CSML, CellML, SBRML, SED-ML, MFAML
 - mzXML, mzData, mzML, TraML, mzIdentML, mzQuantML ..etc (for mass spectrometry data from HUPO-PSI)

Problems in biological data management

- Heterogeneity and complexity of the data
- Huge volume and context dependency of the data
- The nosiness of the data generated by modern high-throughput methods
- Lack of well established data standards and globally unique identifiers, which would allow easy mapping and data integration

Unique identifiers - LSID

- A problem hindering the integration of data is the lack of globally unique identifiers
- Genes and proteins have several different names, it is not possible to merge clearly the data belonging to them.
- The OCLC (Online Community Library Center) proposed to use PURL's (Persistent Uniform Resource Locators)
- Another proposal from life science community to use LSID, the Life Science Identifier , a stable GUID (Globally Unique Identifier)

Data management systems

● Spreadsheet-based approaches

- e.g: MAGE-TAB, ISA-TAB, mzTAB

● Web-based document-sharing tools

- Open-source wikis, semantic wikis, groupware programs like eGroupware, BaseCamp, PBWORKS, DaMaSys (Data Management Systems Biology), BSCW (Be Smart Cooperate Worldwide)

● Laboratory Information Management Systems (LIMS) or Electronic Lab Notebooks (ELN)

● Web-services based workflow systems

- Taverna, Galaxy

Choosing a data management system



Developing a in-house solution

- Using RDBMS as back end and web front-end for managing and processing the data.



Use of open source software, e.g.

- BASE (BioArray Software Environment) from Lund University, Sweden
- SBEAMS (Systems Biology Experiment Analysis Management System) form ISB, Seattle
- AMEN (Annotation, Mapping, Expression and Network suite of tools for molecular systems biology) from ProteHome, France

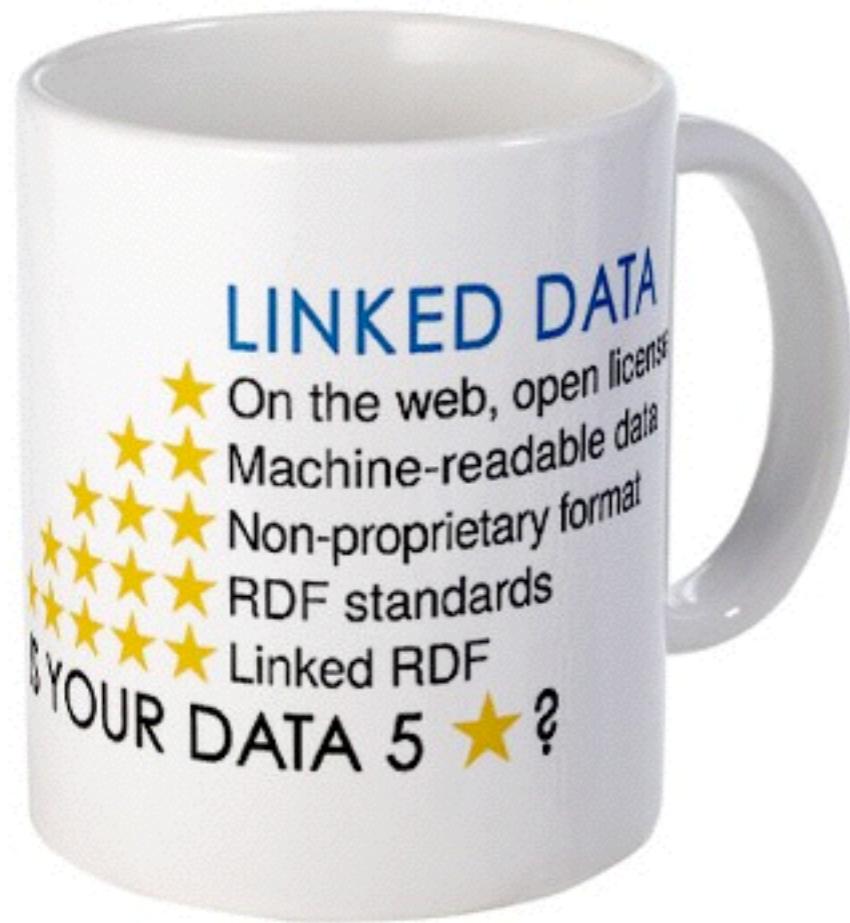


Use of commercial software, e.g.

- Genedata suite of Expressionist, Phylosopher, and Screener
- GeneLogics, GeneBio, Accelrys Pipelien Pilot, Ingenuity Pathway Analysis Suite etc

Data integration

- **Link integration** - based of cross references in the databases, e.g. SRS, Entrez
- **Federated databases** - user queries translated into cross-database query language, e.g. TAMBIS (Transparent Access to Multiple Bioinformatics Information Sources) from University of Manchester, UK
- **Data warehousing** - here source information/databases are regularly scanned and loaded into a central database by an ETL (Extract, Transform, Load) procedure, e.g. **BioMart**, bioCompendium
- **Web services, Services oriented architecture (SOA, WOA)**
 - web services are programmatic interfaces, APIs, allows integration of data by e.g. workflow programs.
 - these are based on protocols like WSDL, SOAP, AJAX, REST, XML-RPC



www.w3.org

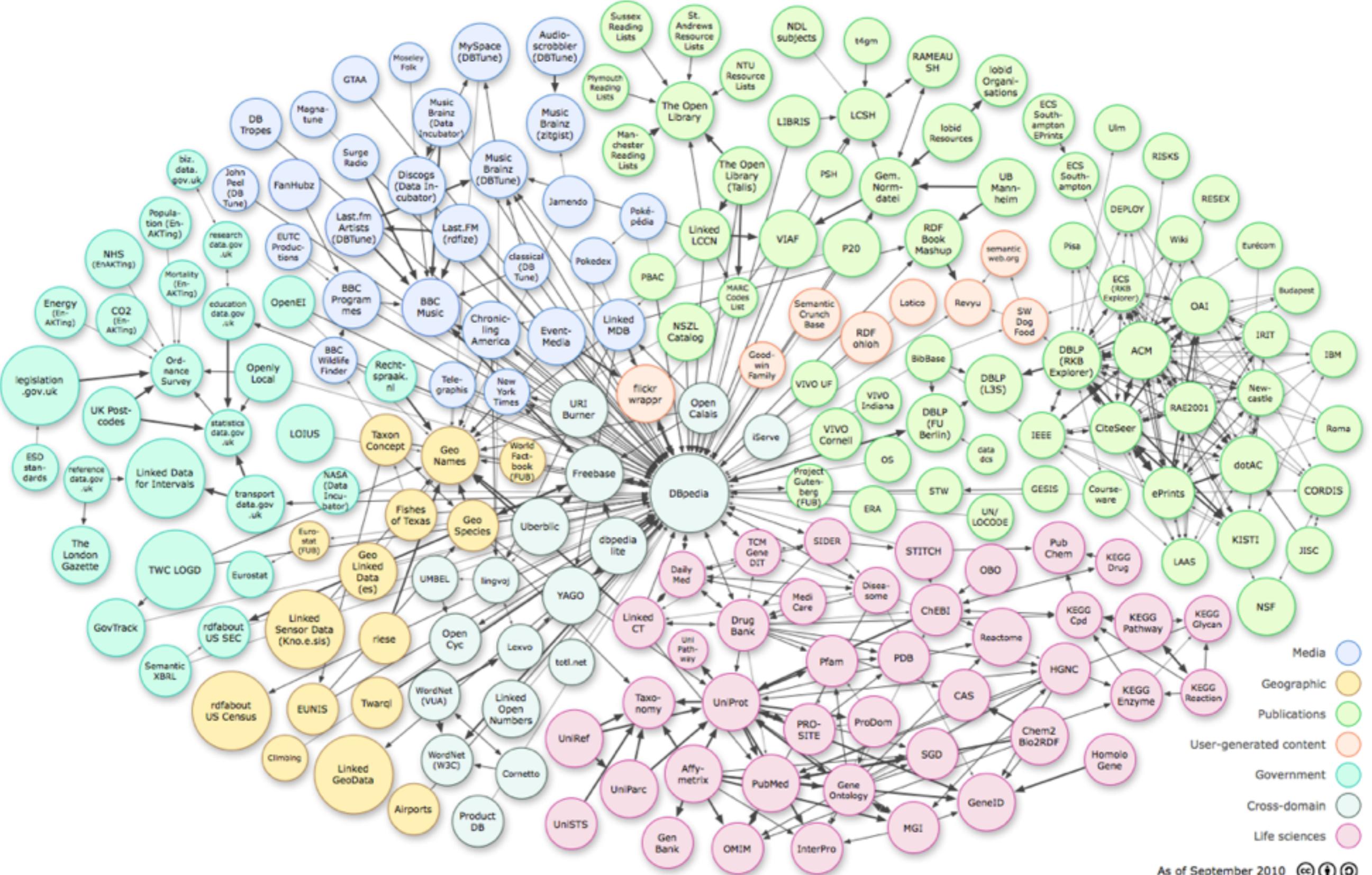
Semantic web and Linked data

- semantically annotating the data with RDF, RDFS, OWL becoming important for reaching the goal of true data integration
- RDF is useful for describing static things or facts, because it uses a simple data model of **subject -- predicate -- object** triples
- This RDF triple knowledge can be treated as a simple graph where subjects and objects are the **nodes** and **edges** are the predicates

```
John | Is a Friend of | James
James | Is a friend of | Jill
Jill | Likes | Snowboarding
Snowboarding | Is a | Sport
```

- New knowledge can be easily added to the existing graph, this is called OWA (Open World Assumption)
- One can query these triple stores with SPARQL (Simple Protocol And RDF Query Language)

Linked data cloud



As of September 2010

Deep web needs semantics

- Lot of valuable biological information on the web
- >24M published articles, lot of knowledge buried in this literature



<http://reflect.ws>

<http://www.ebi.ac.uk/webservices/whatizit/info.jsf>

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

Ensembl

Ensembl Genome Browser

http://www.ensembl.org/index.html

RSS ensembl

Apple Yahoo! GMail m-w YouTube Wikipedia News (531) Popular toi TL HIV-Portal Synonym DB2 bioCompendium

Login · Register

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | More ▾

Search: All species for

e.g. BRCA2 or rat X:100000..200000 or coronary heart disease

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize this list](#))

 **Human**
GRCh37

 **Mouse**
NCBIM37

 **Zebrafish**
Zv9

All genomes

-- Select a species --

[View full list](#)

[Other species](#)

 UNIVERSITÉ DU LUXEMBOURG

New to Ensembl?

Did you know you can:

-  [Learn how to use Ensembl](#)
with our video tutorials and walk-throughs
-  [Add custom tracks](#)
using our new Control Panel
-  [Upload and analyse your data](#)
and save it to your Ensembl account
-  [Search for a DNA or protein sequence](#)
using BLAST or BLAT
-  [Fetch only the data you want](#)
from our public database, using the Perl API
-  [Download our databases via FTP](#)
in FASTA, MySQL and other formats
-  [Mine Ensembl with BioMart](#)
and export sequences or tables in text, html, or Excel format

Still got questions? Try our [FAQs](#) or [glossary](#)

Did you know...?

Lost?
Try our [tutorials](#) page.

<http://www.ensembl.org>



BioMart

http://www.ensembl.org/index.html

RSS ensembl

Apple Yahoo! GMail m-w YouTube Wikipedia News (531) Popular toi TL HIV-Portal Synonym DB2 bioCompendium

Login · Register

e!Ensembl BLAST/BLAT BioMart | Tools | Downloads | More ▾

Search: All species for

e.g. BRCA2 or rat X:100000..200000 or coronary heart disease

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize this list](#))

 **Human**
GRCh37

 **Mouse**
NCBIM37

 **Zebrafish**
Zv9

All genomes

-- Select a species --

[View full list](#)

[Other species](#)

 UNIVERSITÉ DU LUXEMBOURG

New to Ensembl?

Did you know you can:

-  [Learn how to use Ensembl](#)
with our video tutorials and walk-throughs
-  [Add custom tracks](#)
using our new Control Panel
-  [Upload and analyse your data](#)
and save it to your Ensembl account
-  [Search for a DNA or protein sequence](#)
using BLAST or BLAT
-  [Fetch only the data you want](#)
from our public database, using the Perl API
-  [Download our databases via FTP](#)
in FASTA, MySQL and other formats
-  [Mine Ensembl with BioMart](#)
and export sequences or tables in text, html, or Excel format

Still got questions? Try our [FAQs](#) or [glossary](#)

Did you know...?

Lost?
Try our [tutorials](#) page.

<http://www.ensembl.org>



www.ensembl.org/biomart/martview/4d5a02edd3c227c43bf952fa0dfd3b1d

www.ensembl.org/biomart/martview/4d5a02edd3c227c43bf952fa0dfd3b1d Reader

Apple iCloud Facebook Twitter Wikipedia Yahoo! News Popular

Mac App Store... (15) WP 4 | Tr... bioCompendium bioCompendium www.ensembl... Google Transl... Generating SS... + Login/Register

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | More ▾ Search all species...

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh37.p10)
Filters
[None selected]
Attributes
Ensembl Gene ID
Ensembl Transcript ID

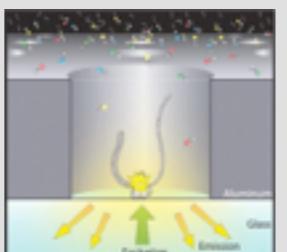
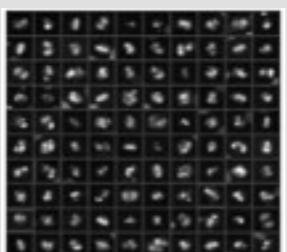
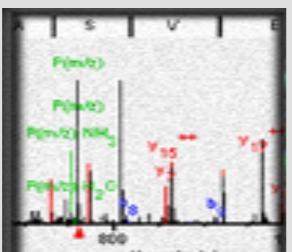
Dataset
[None Selected]

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

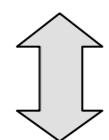
Exercises part 2

bioCompendium

Application



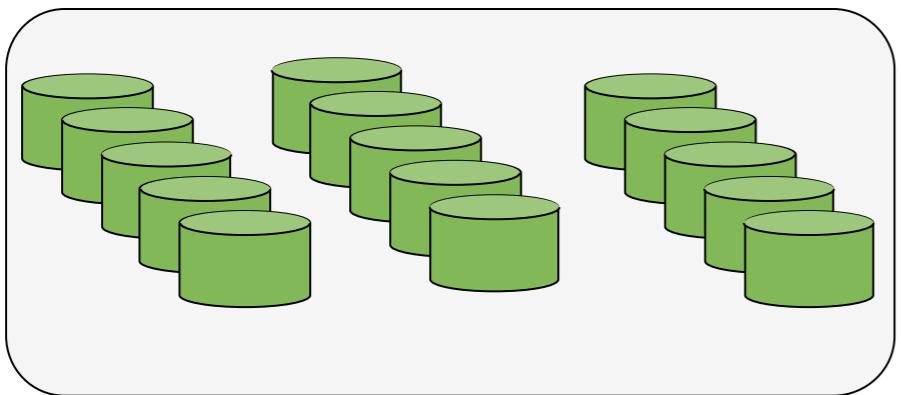
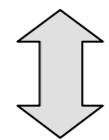
Genome wide high throughput experiments



NR1H2
TP53
LMNA
CDC7
HTT
Rb1

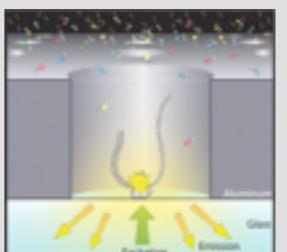
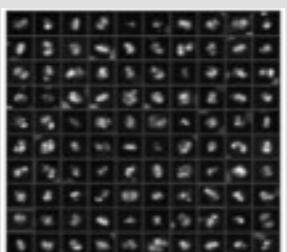
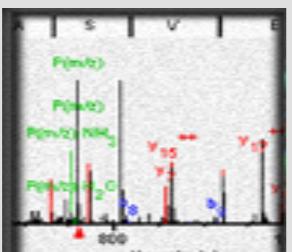


Target prioritization

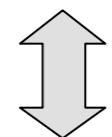


Different biological databases

Application



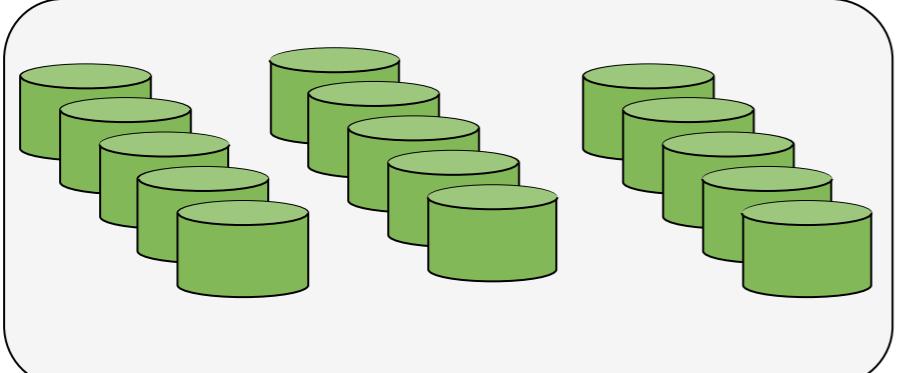
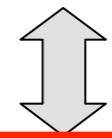
Genome wide high throughput experiments



NR1H2
TP53
LMNA
CDC7
HTT
Rb1



Target prioritization

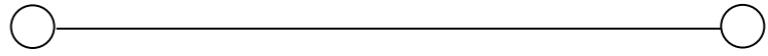


Different biological databases

Experiment set up



Single gene/protein/probe-set list



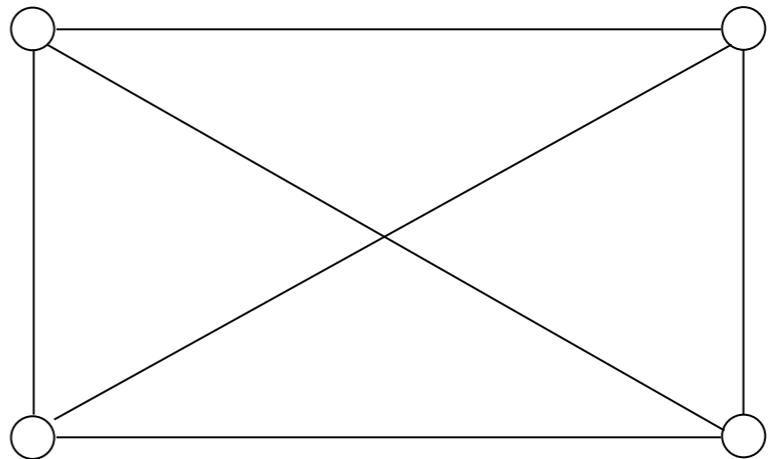
Control

Disease



More result sets

Control Disease



Control +
Drug

Disease+
Drug

Data analysis scenarios

Lamin A/C - Google Search - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Most Visited Mail Yahoo! Mail SRS Support JScriptShell blohma BACE GeneGo Affymetrix m-w ak

Lamin A/C - Google Search

Web Images Videos Maps News Shopping Mail more ▾

Google Lamin A/C Search Advanced Search

Web Show options... Results 1 - 1

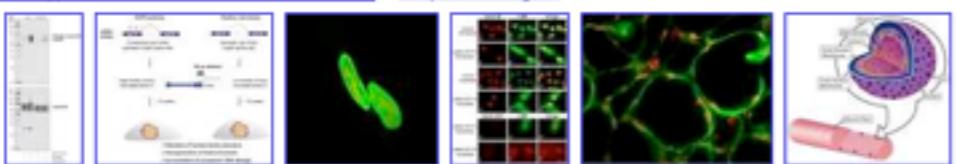
[Lamin A/C \(LMNA\) homepage](#)
5 Jul 2005 ... Lamin C is derived from the LMNA gene using an alternative splice site located in intron 10; Lamin C thus differs C-terminally from the ...
www.dmd.nl/lmna_home.html - Cached - Similar -

[OMIM - LAMIN A/C; LMNA](#)
MIM *150330 · Description · Cloning · Gene Structure · Mapping · Gene Function · Molecular Genetics · Genotype/Phenotype Correlations ...
www.ncbi.nlm.nih.gov/entrez/dispmim.cgi?id=150330 - Cached - Similar -

[4000 - Gene Result](#)
1: LMNA Official Symbol LMNA and Name: **Lamin A/C** [Homo sapiens] Other Aliases: RP11-54H19.1, CDCCD1, CDDC, CMD1A, CMT2B1, EMD2, FPL, FPLD, HGPS, IDC, LDP1, ...
www.ncbi.nlm.nih.gov/gene/4000 - Similar -

Show more results from www.ncbi.nlm.nih.gov

Image results for Lamin A/C - Report images



Data analysis scenarios

The screenshot shows a Mozilla Firefox browser window with the title "Lamin A/C - Google Search - Mozilla Firefox". The address bar displays "Search for Lamin A/C in All the EBI | EBI - Mozilla Firefox" and the URL "http://www.ebi.ac.uk/ebisearch/search.ebi?db=alldb&query=Lamin+A%2FC&formSubmit=3". The main content area is a search results page for "Lamin A/C" across various EBI databases. The results are listed in a grid:

Category	Count
Genomes	51
Nucleotide Sequences	943
Protein Sequences	160
Macromolecular Structures	6
Small molecules	1
Gene Expression	13
Molecular Interactions	9
Reactions & Pathways	0
Protein Families	4
Enzymes	1
Literature	1,335
Ontologies	2
EBI Web Site	0

Below the results, there is a "Refine your search" section with a search bar containing "Search for Lamin A/C in All the EBI" and a "Refine" button. The left sidebar of the browser shows various tabs and links related to Lamin A/C and EBI.

Data analysis scenarios

The screenshot shows a Mozilla Firefox browser window with the title bar "Lamin A/C - Google Search - Mozilla Firefox". The address bar contains the URL "Search for Lamin A/C in All the EBI | EBI - Mozilla Firefox". The main content area displays the NCBI Entrez search results for the query "Lamin A/C". The results are organized into two main sections: "PubMed" and "All Databases".

PubMed:

- 773 PubMed: biomedical literature citations and abstracts
- 1133 PubMed Central: free, full text journal articles
- 4 Site Search: NCBI web and FTP sites

All Databases:

- 46 Books: online books
- 12 OMIM: online Mendelian Inheritance in Man
- 2 OMIA: online Mendelian Inheritance in Animals

Other Database Results:

Database	Count	Description
Nucleotide	94	Core subset of nucleotide sequence records
EST	5	Expressed Sequence Tag records
GSS	none	Genome Survey Sequence records
Protein	224	sequence database
Genome	2	whole genome sequences
Structure	2	three-dimensional macromolecular structures
dbGaP	2	genotype and phenotype
UniGene	16	gene-oriented clusters of transcript sequences
CDD	none	conserved protein domain database
3D Domains	5	domains from Entrez Structure
UniSTS	6	markers and mapping data
PopSet	none	population study data sets

Data analysis scenarios



Data analysis scenarios

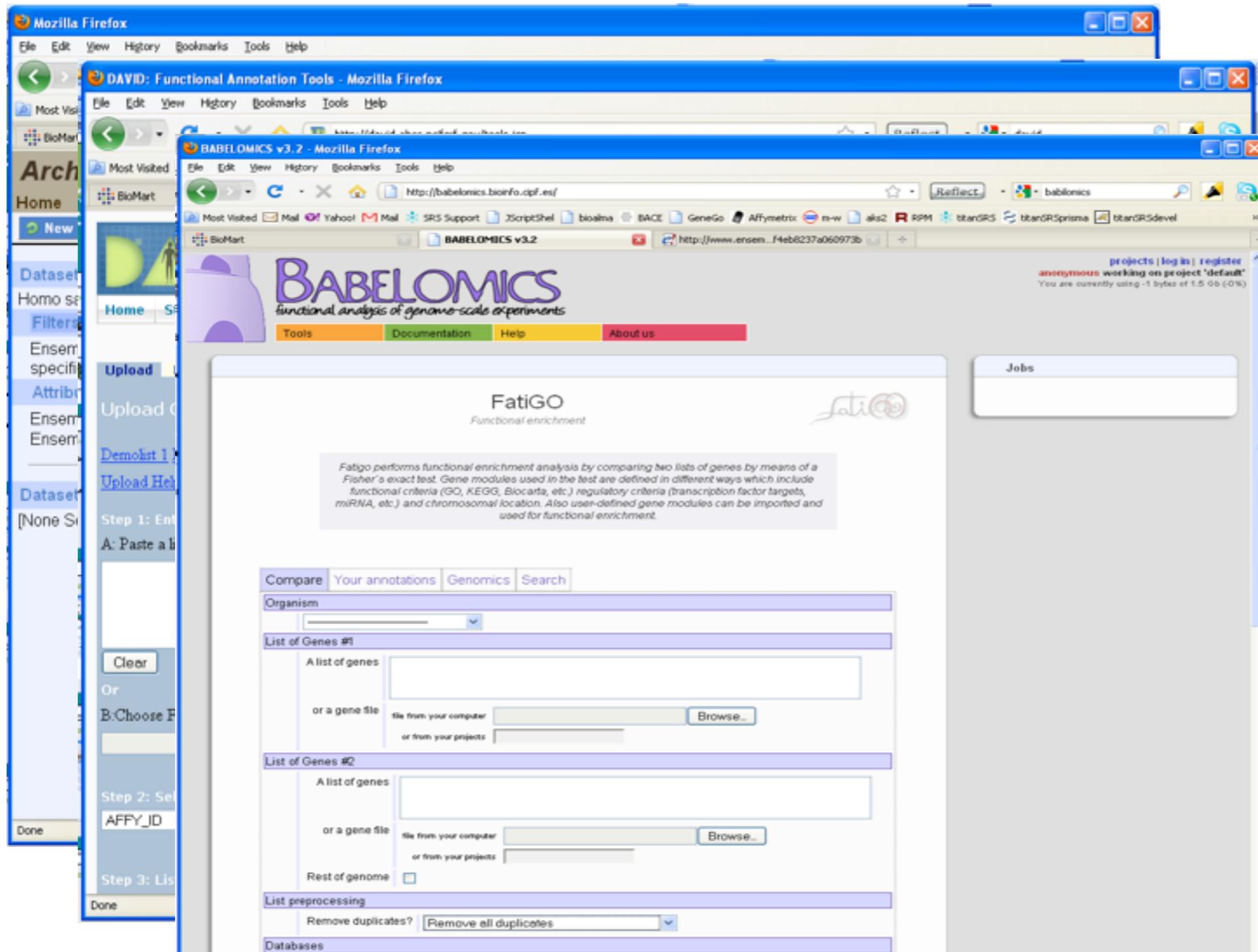
The screenshot shows the Ensembl BioMart interface in Mozilla Firefox. The left sidebar lists datasets: 'Homo sapiens genes (NCBI36)', 'Filters' (Ensembl protein ID(s)), and 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID). The main panel has a header 'Please restrict your query using criteria below'. It includes sections for 'REGION', 'GENE', and 'EXPRESSION'. In the 'GENE' section, the 'ID list limit' checkbox is checked, and a dropdown menu titled 'Ensembl protein ID(s)' is open, showing a list of IDs: ENSP00000005257, ENSP00000005259, ENSP00000005260, ENSP00000005279, and ENSP00000005284. A green oval highlights this dropdown menu. Other dropdowns in the 'GENE' section include 'with Affymetrix Microarray hc g110 ID(s)' (set to 'Only'), 'Gene type' (list: IG_C_gene, IG_D_gene, IG_J_gene, IG_V_gene, IG_pseudogene), 'Source' (ensembl), and 'Status (gene)' (KNOWN). Below these are 'Transcript count >=' and 'Status (transcript)' dropdowns, both set to KNOWN.

Data analysis scenarios

The screenshot shows a Mozilla Firefox browser window with multiple tabs open. The main tab displays the DAVID Functional Annotation Tools interface, specifically the Analysis Wizard. The left sidebar contains navigation links for Home, Dataset, Filters, Ensembl specific, Attributes, and Ensembl. The main content area is titled "Analysis Wizard" and "DAVID Bioinformatics Resources 2008, NIAID/NIH". It features a "Upload" section with tabs for "Upload", "List", and "Background". The "Upload" tab is active, showing options to "Upload Gene List" via a text input field or a file browser, and a dropdown menu for "Step 2: Select Identifier" set to "AFFY_ID". Below this is a "Step 3: List Type" dropdown menu. A blue arrow points from the text "Step 1. Submit your gene list through left panel." to the "Upload" tab. To the right, there are links to "Tell us how you like the tool" and "Contact us for questions". A note states: "Note: Affy Exon IDs and Affy Gene Array IDs are now supported in DAVID, as "affy_id" type." An example list of IDs is provided:

1007_s_at
1053_s_at
117_s_at
121_s_at
1255_g_at
1294_s_at
1316_s_at
1320_s_at
1405_j_at
1431_s_at
1438_s_at
1487_s_at
1494_f_at
1598_g_at

Data analysis scenarios



Data analysis scenarios

- Scientists may need to analyze/compare/enrich
- One or more gene list(s)
- Experimental results with already published data
- Experimental results from different model organisms
- Experimental results with some documents (pdf, doc, excel, ppt, text etc)
- Bio-entities present in different documents
- To know wide spectrum of biological information and enrichments from gene list(s).

Problem

- Identifier conversion
- What is the prior knowledge that we have about my data set?
- Comparison to other experimental results
- Analysis and prioritization of results for further experimental studies
 - Genes/proteins/miRNA/siRNAs ... etc

Solution

- Biological data driven approach
- 100s of biological databases / TeraBytes of biological knowledge
- Need a system that contains wide range of biological information and able to
 - Handle list(s) of different identifiers
 - Extract bio-entities from documents
 - Map information between different organisms
 - Analyze & cluster biological information
 - Visualize the results

bioCompendium



bioCompendium
The high-throughput experimental data analysis platform

home examples help search...

Gene list(s) analysis

Select primary organism : **human**

Select background : whole genome other gene list(s)

Upload gene list(s) and/or documents :

Org	Name	File	ID/Document Type
human	gene_list_1	Browse...	
Ensembl Gene ID			

Reset GO!

What is it & what it does

bioCompendium is a publicly accessible, high-throughput experimental data analysis platform. The system is designed to work with large lists of genes or proteins for which it collects a wide spectrum of biological information. It facilitates the analysis, comparison and enrichment of experimental results; either proprietary or publicly available data sets. Typical use cases are the prioritization of potential targets from gene expression analysis studies or from RNAi studies. The current version is designed to work best for human, mouse and yeast but other model organisms will be included in the next releases.

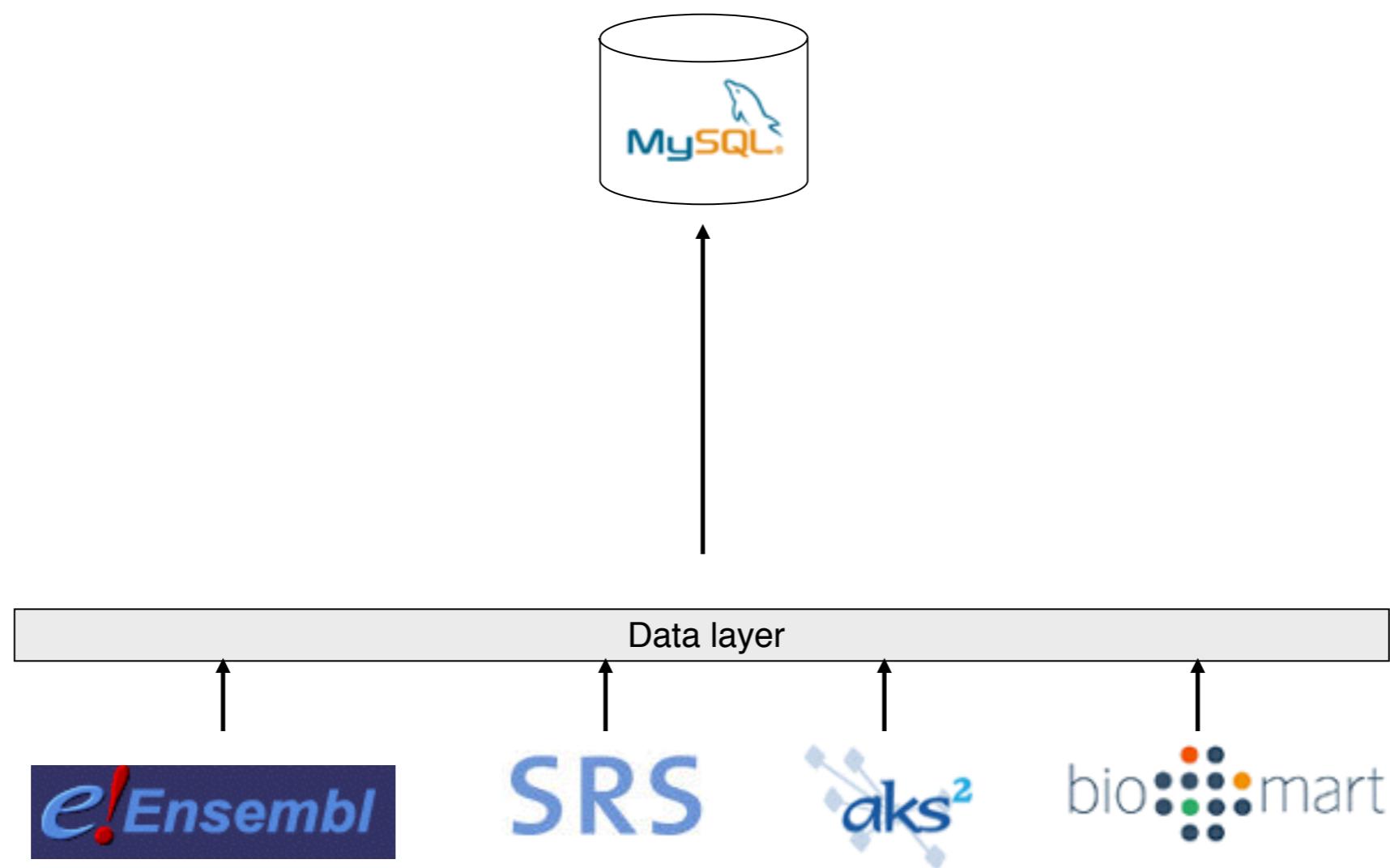
Main features of the system are:

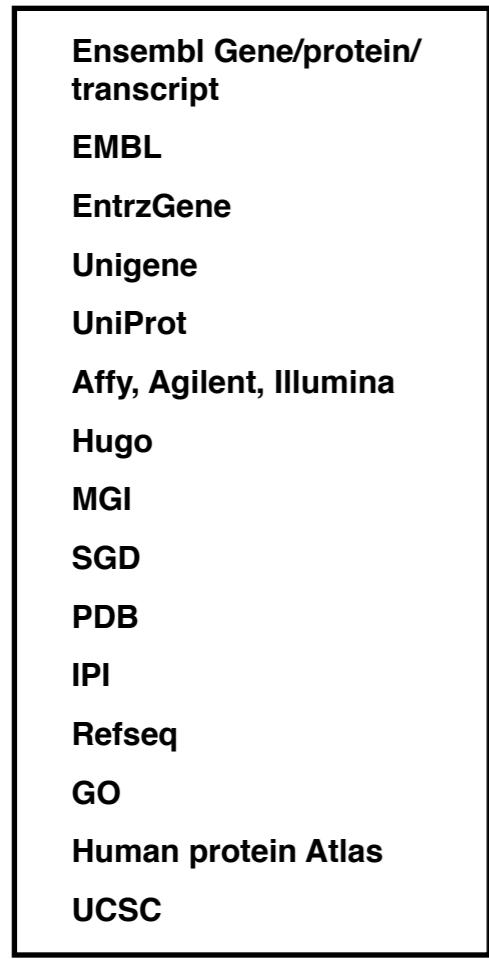
- Input and conversion of a wide range of input ID's like UniProt, GO, Affymetrix and RefSeq
- Extraction of bio-entities from different file formats (MS-Office, PDF and flat text)
- Comprehensive knowledge collection from different biological database for a given list(s) of genes
- Search interface to the knowledge collection to find information like gene annotations, disease associations, sequences domain architectures, interfering chemicals and involved pathways
- Enrichment analysis for GeneOntology terms, diseases, pathways and other biological concepts
- Extraction of the protein-protein, protein-chemistry interactions networks
- Compilation of clusters based on sequence homology & sequence domain architectures in a given list(s) of genes
- Analysis and clustering of transcription factor binding site (TFBS) profiles
- Access to orthology information, clinical trial and patent information
- Comparison of results derived from different experimental conditions, time series or treatments

See [help pages](#) for more details.

Send comments to [Venkata P. Satagopam](#)

bioCompendium overview





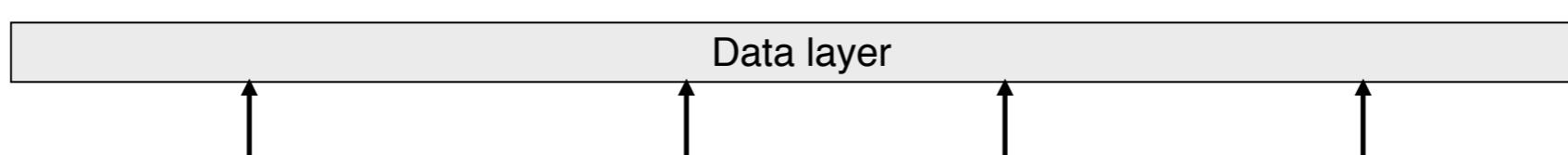
ID Conversion

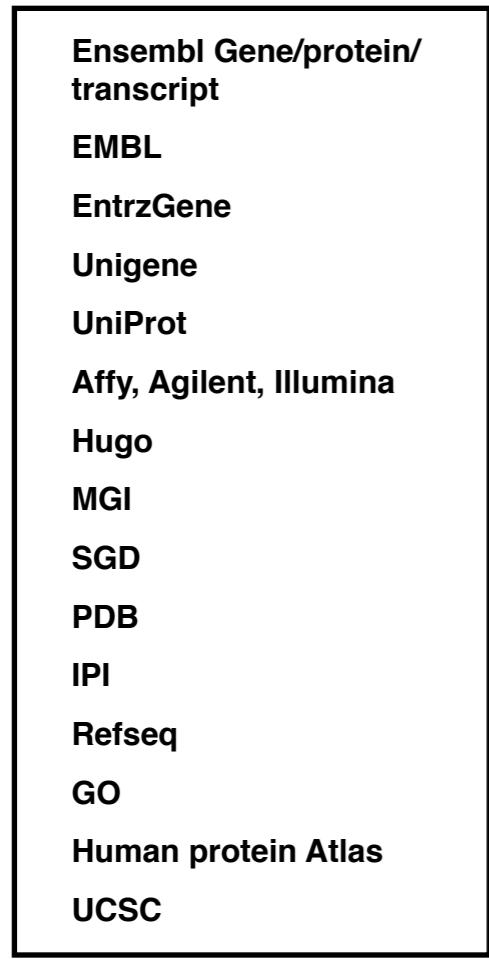


Ensembl genes

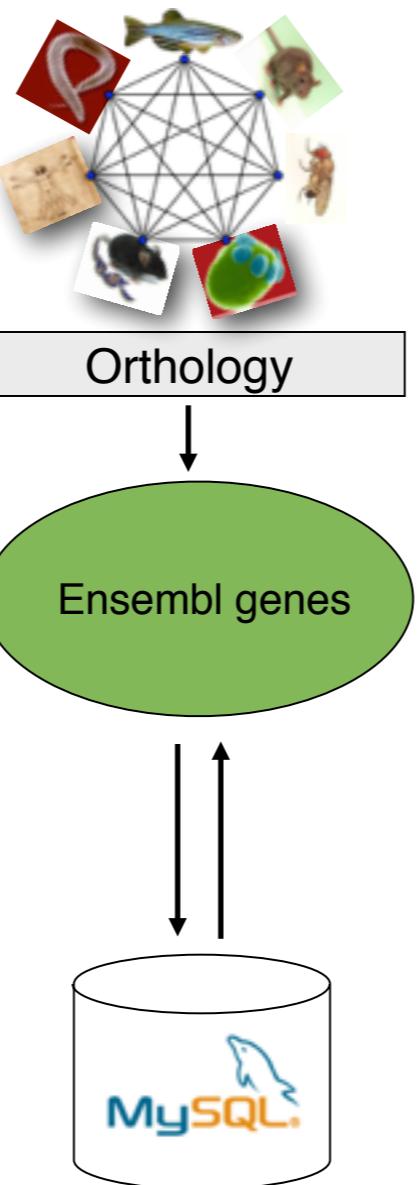
MySQL

Data layer

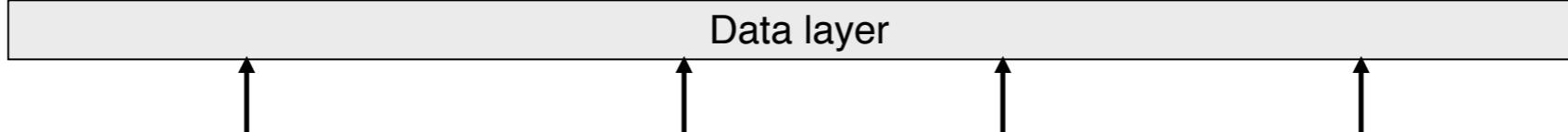


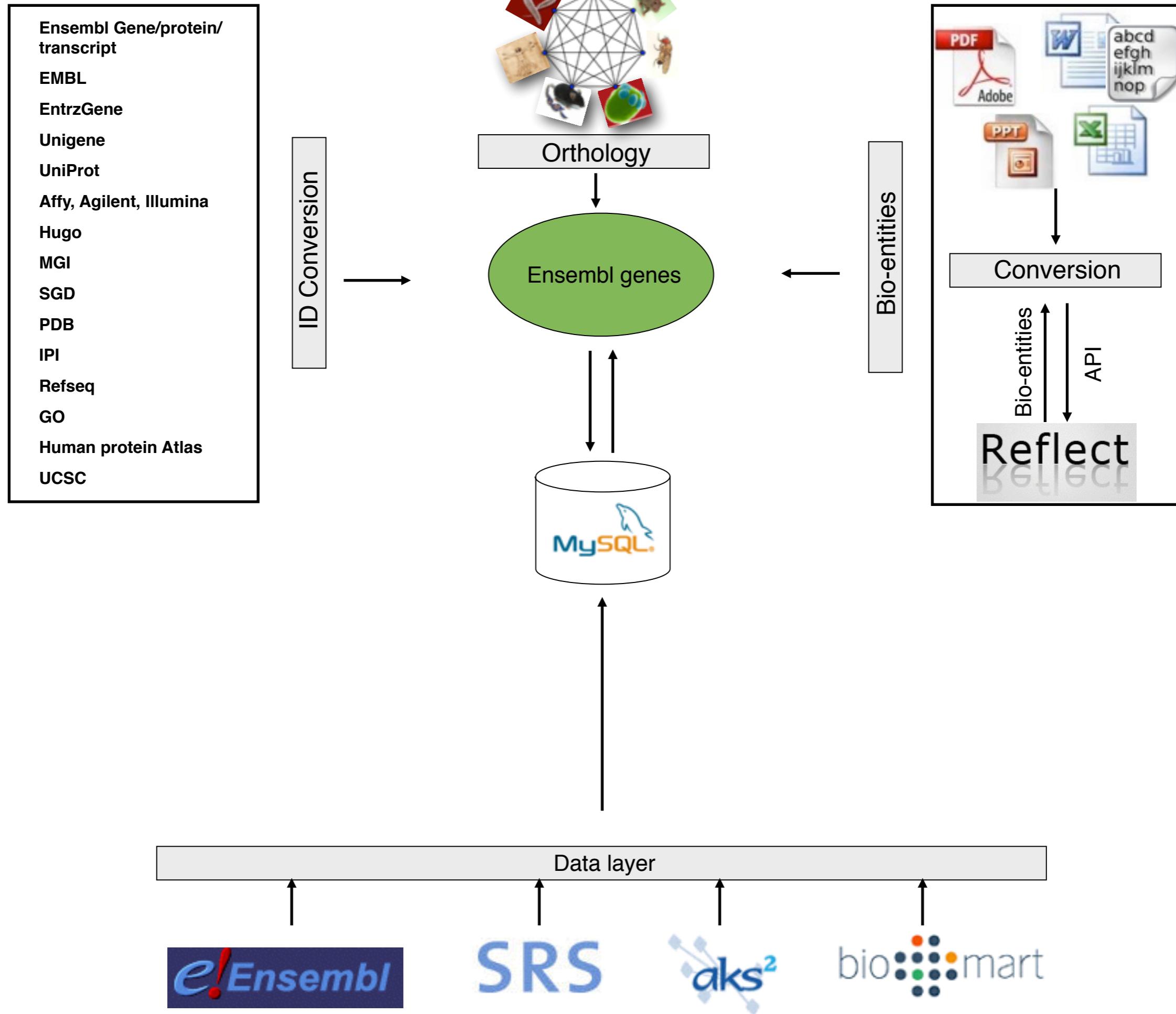


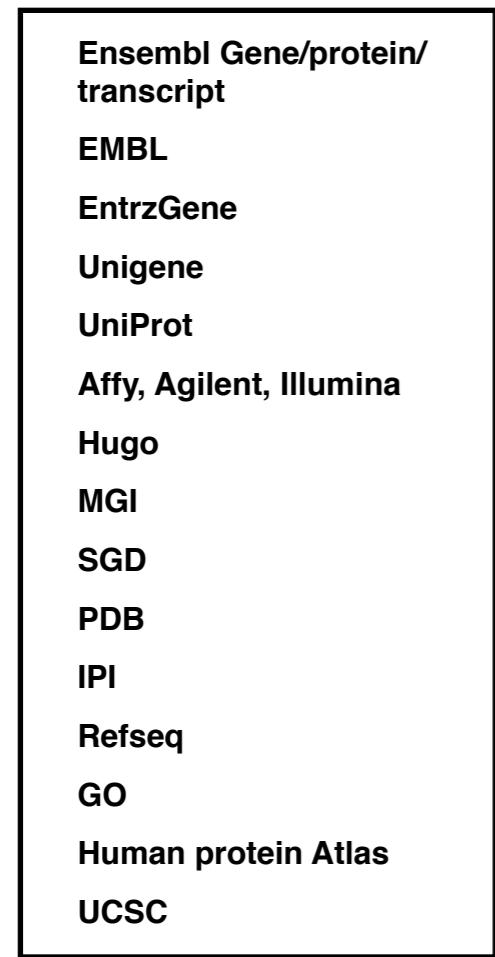
ID Conversion



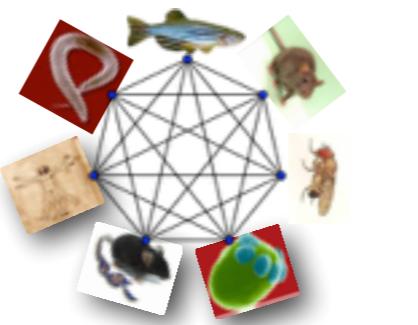
Data layer







ID Conversion



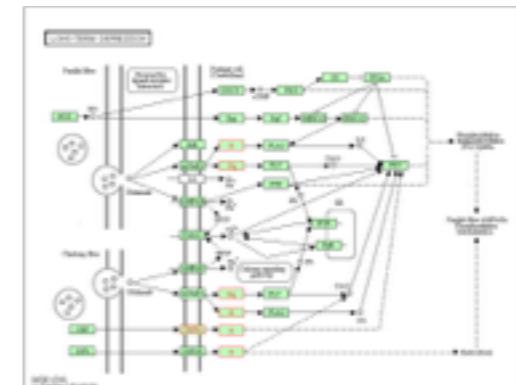
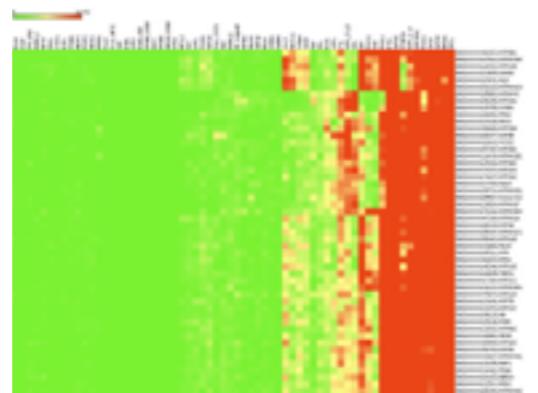
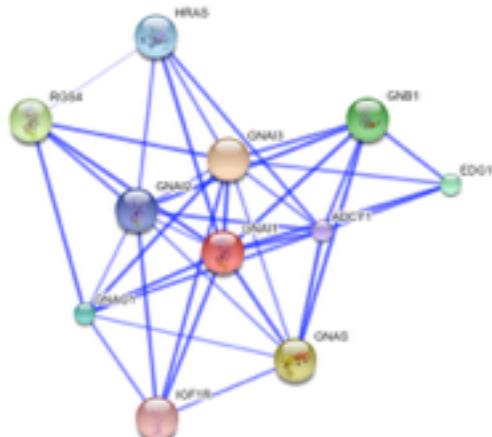
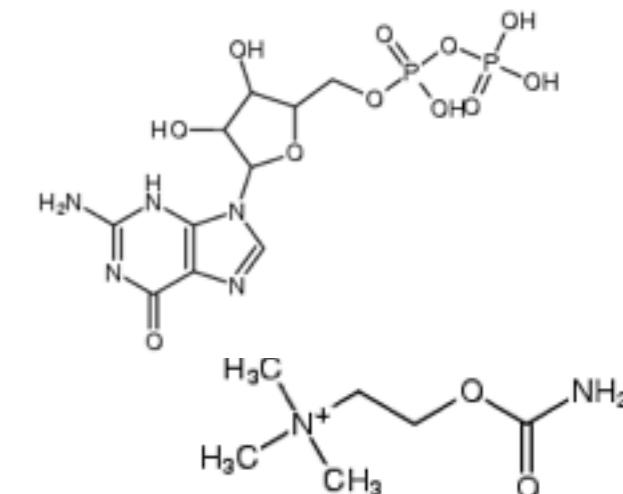
Orthology

Ensembl genes

Bio-entities

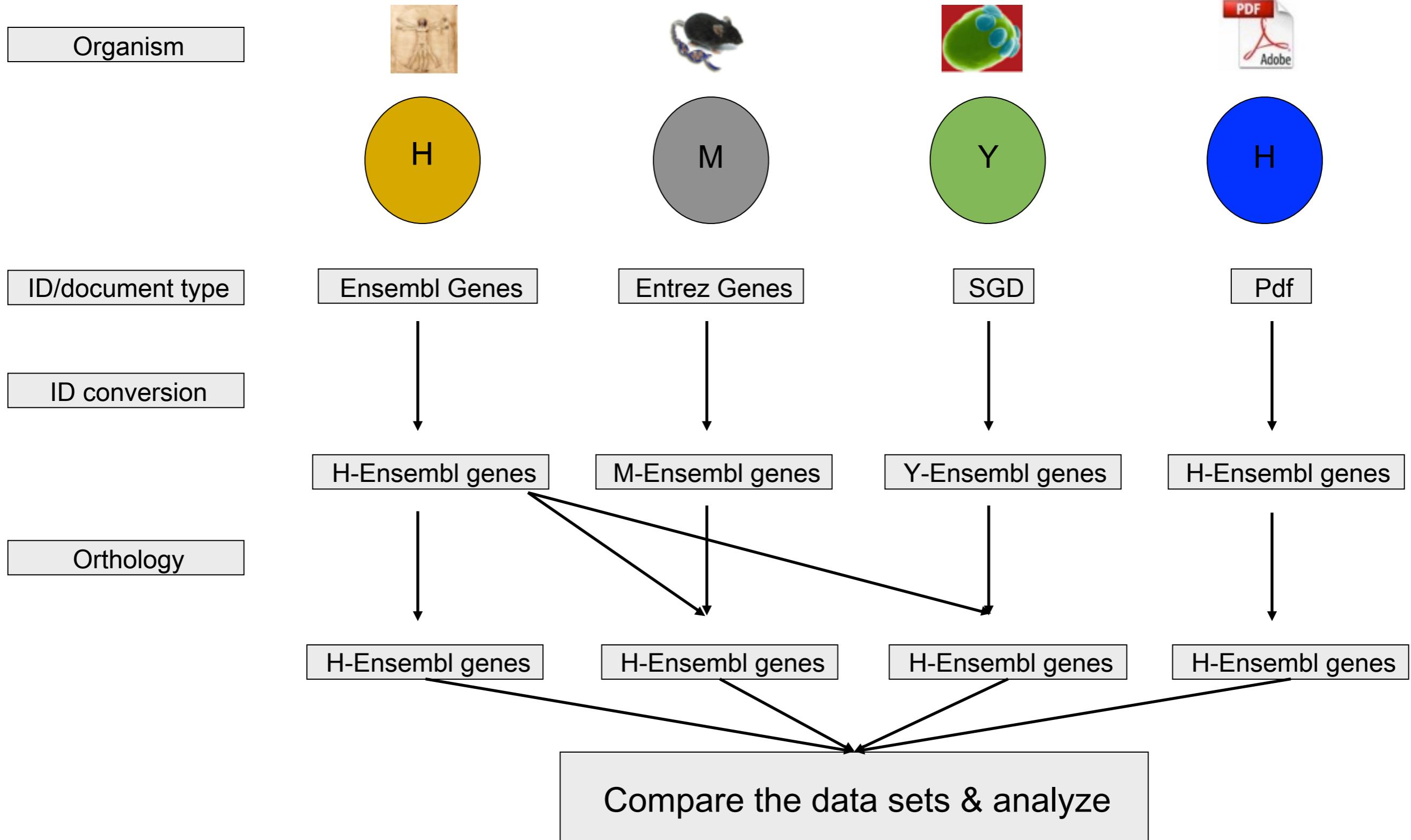
Conversion

Reflect

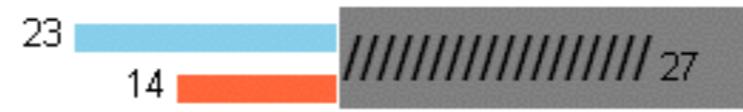
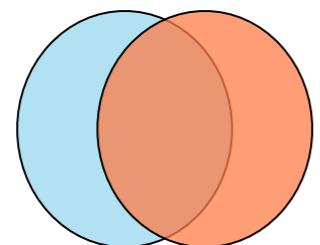
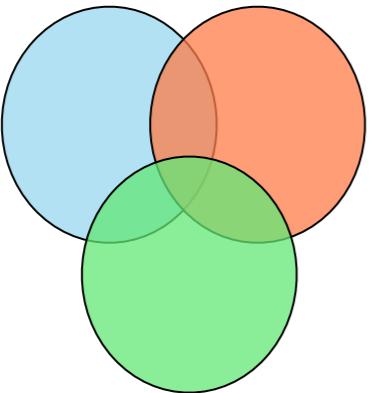
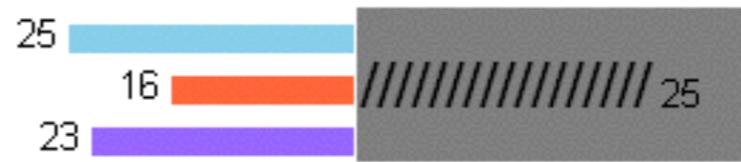
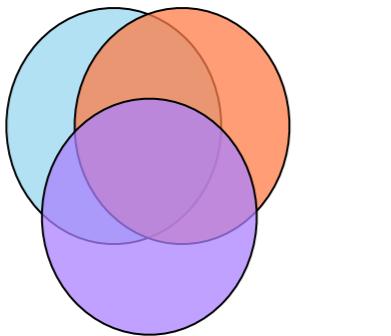
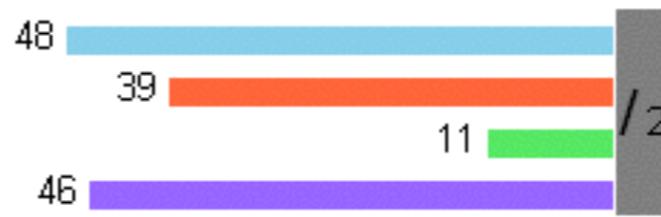
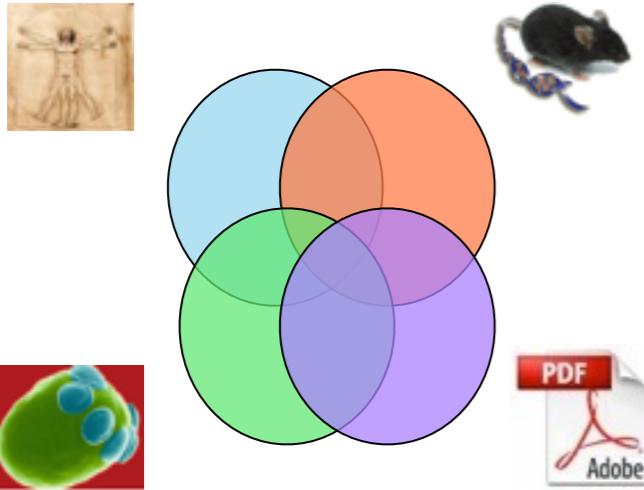


Data layer

Data input



Datasets comparison & analysis





bioCompendium

The high-throughput experimental data analysis platform

home examples help

Gene list(s) analysis

Select primary organism : human ▾

Select background : whole genome other gene list(s)

Upload gene list(s) and/or documents :

Org	Name	File	ID/Document Type
human ▾	gene_list_1	Browse...	Ensembl Gene ID ▾

[Upload another list](#)

[Reset](#)

[GO !](#)



bioCompendium

The high-throughput experimental data analysis platform

home examples help

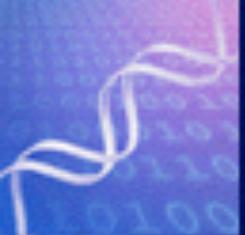
Gene list(s) analysis

Select primary organism : **human**

Select background : whole genome other gene list(s)

Upload gene list(s) and/or documents :

Org	Name	File	ID/Document Type
human	gene_list_1	<input type="button" value="Browse..."/>	Ensembl Gene ID
human	other list		
mouse			
yeast			



bioCompendium

The high-throughput experimental data analysis platform

home examples help

Gene list(s) analysis

Select primary organism : human ▾

Select background : whole genome other gene list(s)

Upload gene list(s) and/or documents :

Org	Name	File	ID/Document Type
human ▾	exp1_human	/Users/Venkata/Desktop/vs	Browse... Ensembl Gene ID ▾

[Upload another list](#)

Reset

GO !



Gene list(s) analysis

Select primary organism : **human**

Select background : whole genome other gene list(s)

Upload gene list(s) and/or documents :

Org	Name	File	ID/Document Type
human	sembl_genes	/Users/Venkata/Desktop/vs/	<input type="button" value="Browse..."/>
Upload another list			

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl protein ID
- Agilent cgh
- Agilent Probe
- EMBL ID
- EntrezGene ID
- GO ID
- HGNC ID
- HGNC symbol
- Illumina v1 ID
- Illumina v2 ID
- IPI ID
- MIM Gene Accession
- MIM Morbid Accession
- miRBase ID
- PDB ID
- Protein ID
- Refseq DNA ID
- Refseq protein ID



Gene list(s) analysis

Select primary organism : **human**

Select background : whole genome other gene list(s)

Upload gene list(s) and/or documents :

Org	Name	File	ID/Document Type
human	sembl_genes	/Users/Venkata/Desktop/vs/	<input type="button" value="Browse..."/>
Upload another list			
<input type="button" value="Reset"/>		<input type="button" value="GO !"/>	

ID/Document Type

- Ensembl Gene ID
- Affy hg u95av2 ID
- Affy hg u95b ID
- Affy hg u95c ID
- Affy hg u95d ID
- Affy hg u95e ID
- Affy hg u133a 2 ID
- Affy hg u133a ID
- Affy hg u133b ID
- Affy hg u133 plus 2 ID
- Affy HuEx
- Affy HuGene
- Affy hugeneID
- Affy u133 x3p ID

----- Document Types -----

- pdf**
- doc
- excel
- ppt
- html
- ascii

Gene list(s) analysis

Select primary organism : **human**

Select background : whole genome other gene list(s)

Upload gene list(s) and/or documents :

Org	Name	File	ID/Document Type
human	sembi_genes	/Users/Venkata/Desktop/v:	Browse... Ensembl Gene ID
mouse	entrezgenes	/Users/Venkata/Desktop/v:	Browse... EntrezGene ID
yeast	east_sgd_ids	/Users/Venkata/Desktop/v:	Browse... Sgd ID
human	exp4_pdf	/Users/Venkata/Desktop/v:	Browse... pdf

[Upload another list](#)

Legend

Click on color bars to explore analysis results

Expand All Collapse All

Selected background

Whole genome

#	Name	Size	Organism
1	exp1_human_ensembl_genes	50	human
2	exp2_mouse_entrezgenes	41	mouse->human
3	exp3_yeast_sgd_ids	13	yeast->human
4	exp4_pdf	43	human

Input after conversion

Combinations

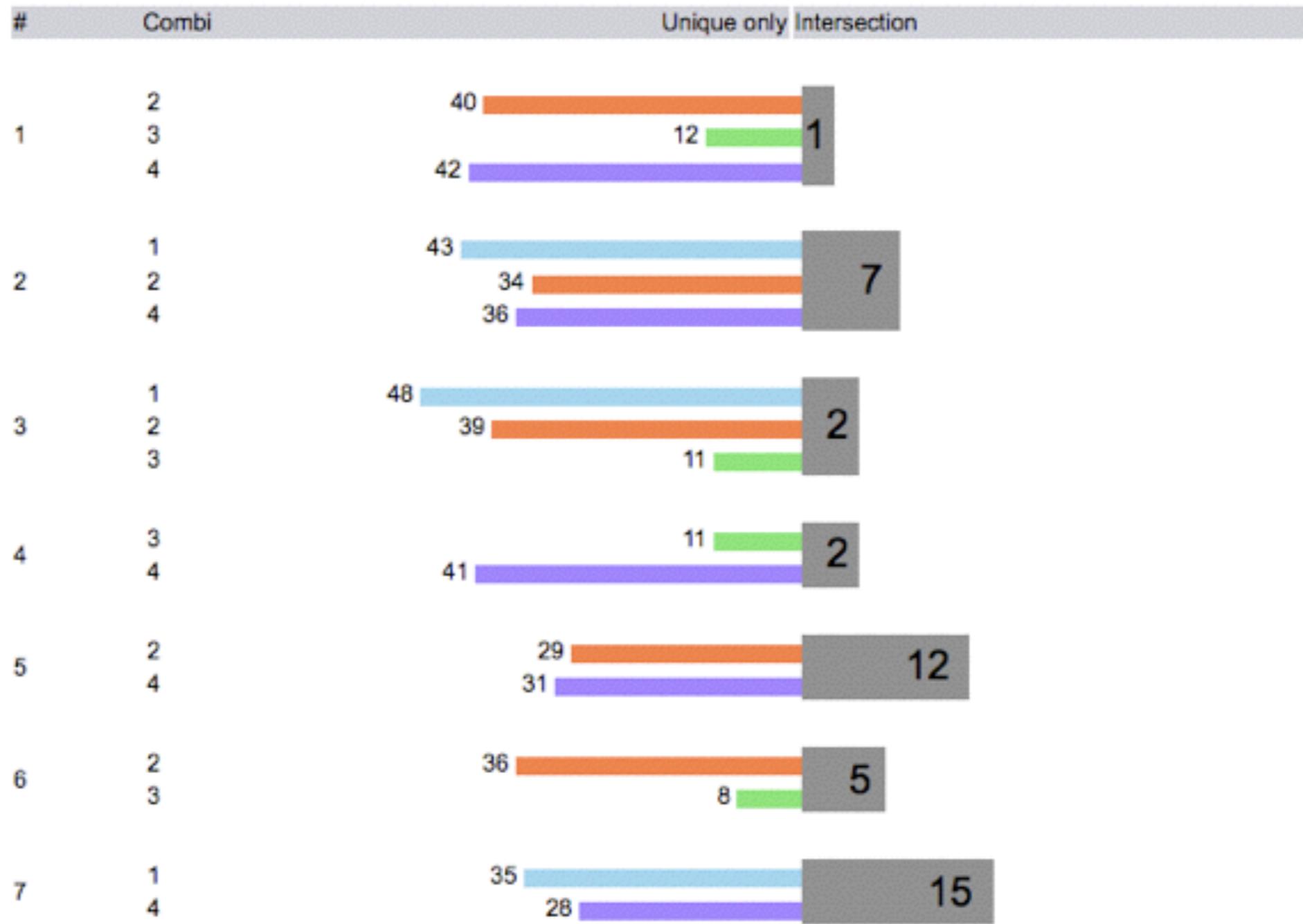
Combinations between maximum top 10 sets:

Legend	Click on color bars to explore analysis results																				
	<input checked="" type="checkbox"/> Expand All <input type="checkbox"/> Collapse All																				
Selected background	Whole genome																				
Input after conversion	<table border="1"><thead><tr><th>#</th><th>Name</th><th>Size</th><th>Organism</th></tr></thead><tbody><tr><td>1</td><td>exp1_human_ensembl_genes</td><td>50</td><td>human</td></tr><tr><td>2</td><td>exp2_mouse_entrezgenes</td><td>41</td><td>mouse->human</td></tr><tr><td>3</td><td>exp3_yeast_sgd_ids</td><td>13</td><td>yeast->human</td></tr><tr><td>4</td><td>exp4_pdf</td><td>43</td><td>human</td></tr></tbody></table>	#	Name	Size	Organism	1	exp1_human_ensembl_genes	50	human	2	exp2_mouse_entrezgenes	41	mouse->human	3	exp3_yeast_sgd_ids	13	yeast->human	4	exp4_pdf	43	human
#	Name	Size	Organism																		
1	exp1_human_ensembl_genes	50	human																		
2	exp2_mouse_entrezgenes	41	mouse->human																		
3	exp3_yeast_sgd_ids	13	yeast->human																		
4	exp4_pdf	43	human																		
Combinations	<input checked="" type="checkbox"/> Combinations between maximum top 10 sets: <div style="border: 1px solid #ccc; padding: 10px; width: fit-content; margin-left: auto; margin-right: auto;"><p>exp4_pdf Close</p><ul style="list-style-type: none">• Summary sheets• Homology clustering• Orthology Information• Clustering by domain architecture• Pathway Information• Chemistry Information• Gene ontology• Transcription factor binding site profiling• Patent Information• P-P, P-C Interactions• Visualization• Download gene list</div>																				

Input after conversion

#	Name	Size	Organism
1	exp1_human_ensembl_genes	50	human
2	exp2_mouse_entrezgenes	41	mouse->human
3	exp3_yeast_sgd_ids	13	yeast->human
4	exp4_pdf	43	human

Combinations between maximum top 10 sets:

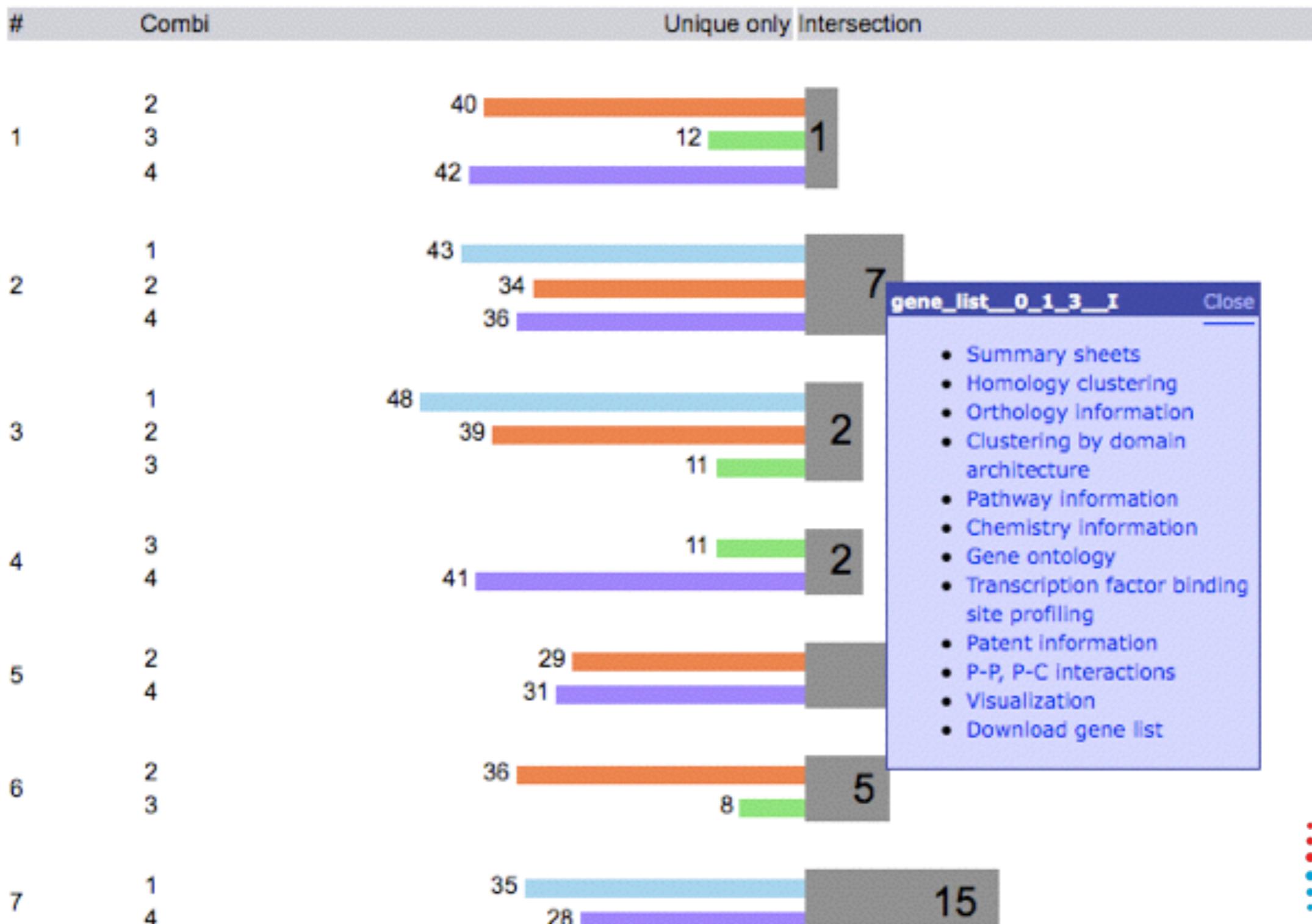


Combinations

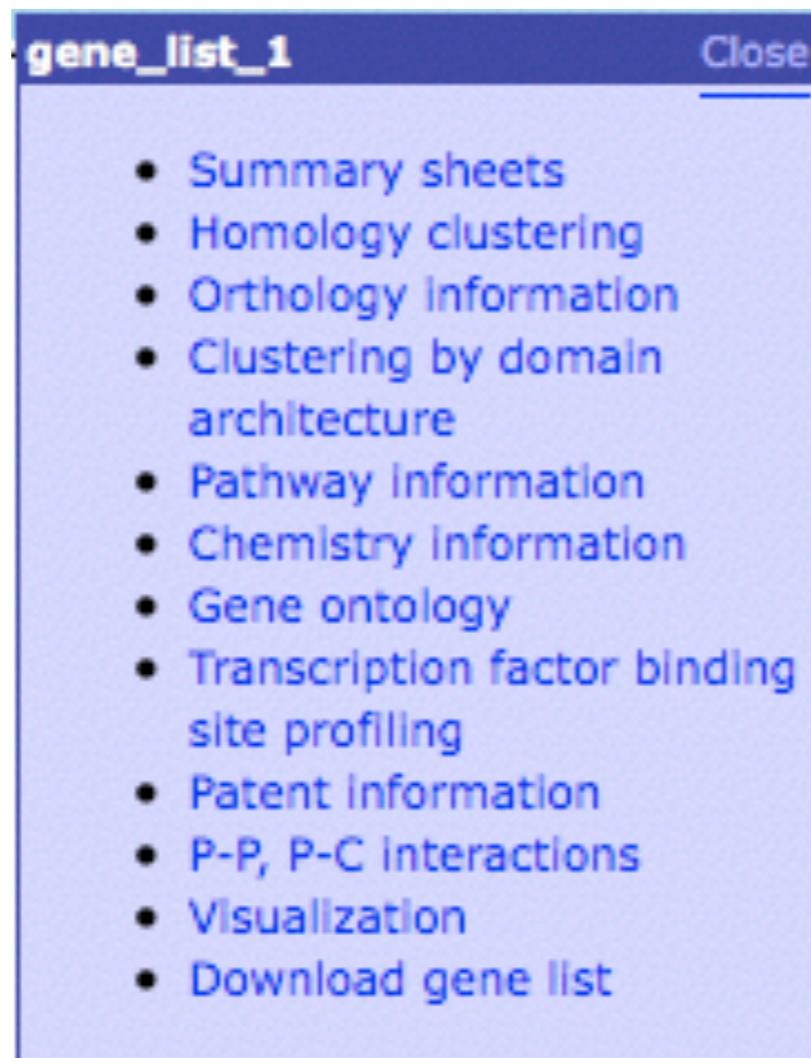
Input after conversion

#	Name	Size	Organism
1	exp1_human_ensembl_genes	50	human
2	exp2_mouse_entrezgenes	41	mouse->human
3	exp3_yeast_sgd_ids	13	yeast->human
4	exp4_pdf	43	human

Combinations between maximum top 10 sets:



Menu



List of genes/ proteins

		Sheets	Charts	SmartArt Gra
1	LPO_MOUSE	Erythropoietin		
2	EPOR_MOUSE	Erythropoietin receptor		
3	JAK2_MOUSE	Tyrosine-protein kinase JAK2		
4	PTN6_MOUSE	Tyrosine-protein phosphatase non-receptor type 6		
5	PTN11_MOUSE	Tyrosine-protein phosphatase non-receptor type 11		
6	GAB1_MOUSE	GAB2-associated-binding protein 1		
7	GAB2_MOUSE	GAB2-associated-binding protein 2		
8	SHIP1_MOUSE	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1		
9	SHIP2_MOUSE	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 2		
10	P55A_MOUSE	Phosphatidylinositol 3-kinase regulatory subunit alpha		
11	P55B_MOUSE	Phosphatidylinositol 3-kinase regulatory subunit beta		
12	P55G_MOUSE	Phosphatidylinositol 3-kinase regulatory subunit gamma		
13	PKCXA_MOUSE	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha isoform		
14	PKCXB_MOUSE	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta isoform		
15	PK3CG_MOUSE	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma isoform		
16	PK3CD_MOUSE	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta isoform		
17	PDPK1_MOUSE	3-phosphoinositide-dependent protein kinase 1		
18	AKT11_MOUSE	RAC-alpha serine/threonine-protein kinase		
19	AKT12_MOUSE	RAC-beta serine/threonine-protein kinase		
20	AKT13_MOUSE	RAC-gamma serine/threonine-protein kinase		
21	GSK3B_MOUSE	Glycogen synthase kinase-3 alpha		
22	GSK3B_MOUSE	Glycogen synthase kinase-3 beta		
23	FRAP_MOUSE	Serine/threonine-protein kinase mTOR		
24	LST8_MOUSE	Target of rapamycin complex subunit LST8		
25	RPTOR_MOUSE	Regulatory-associated protein of mTOR		
26	AKTS1_MOUSE	Proline-rich AKT1 substrate 1		
27	RICTR_MOUSE	Rapamycin-insensitive companion of mTOR		
28	RHEB_MOUSE	GTP-binding protein Rheb		
29	4EBP1_MOUSE	Eukaryotic translation initiation factor 4E-binding protein 1		
30	TSC1_MOUSE	Hamartin		
31	TSC2_MOUSE	Tuberin		
32	KS881_MOUSE	Ribosomal protein S8 kinase beta-1		
33	KS882_MOUSE	Ribosomal protein S8 kinase beta-2		
34	IF4E_MOUSE	Eukaryotic translation initiation factor 4E		
35	KPG2_MOUSE	Protein kinase C zeta type		
36	PR15A_MOUSE	Protein phosphatase 1 regulatory subunit 15A		
37	PP193C_MOUSE	Protein phosphatase 1 regulatory subunit 3C		
38	PLP1_MOUSE	PH domain leucine-rich repeat-containing protein phosphatase 1		
39	PML_MOUSE	Probable transcription factor PML		
40	2ASA_MOUSE	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit alpha isoform		
41	2ASG_MOUSE	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform		
42	2ASE_MOUSE	Serine/threonine-protein phosphatase 2A regulatory subunit epsilon isoform		
43	PP2AA_MOUSE	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform		
44	PP2AB_MOUSE	Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform		
45	Q9WYH4_MOUSE	Forkhead protein FKHF2		
46	FOXP1_MOUSE	Forkhead box protein O1		
47	CCND2_MOUSE	G1/S-specific cyclin-D2		
48	CCNG2_MOUSE	Cyclin-G2		
49	CCNE1_MOUSE	G1/S-specific cyclin-E1		
50	CDKN1B_MOUSE	Cyclin-dependent kinase inhibitor 1B		
51	CDKN1A_MOUSE	Cyclin-dependent kinase inhibitor 1		
52	CCNC_MOUSE	Cyclin-C		
53	EGR3_MOUSE	Early growth response protein 3		
54	CCDC45_MOUSE	CDC45-related protein		
55	BAD_MOUSE	Bcl2 antagonist of cell death		
56	MDM2_MOUSE	E3 ubiquitin-protein ligase Mdm2		
57	P53_MOUSE	Cellular tumor antigen p53		
58	4EBP1_MOUSE	Eukaryotic translation initiation factor 4E-binding protein 1		
59	MYC_MOUSE	Myo proto-oncogene protein		
60	PIM1_MOUSE	Proto-oncogene serine/threonine-protein kinase pim-1		
61	JUNB_MOUSE	Transcription factor jun-B		
62	ACV1B_MOUSE	activin A receptor, type IIB		
63	ACV1C_MOUSE	activin A receptor, type IC		
64	ACVR1_MOUSE	activin A receptor, type I		
65	AKT11_MOUSE	thymoma viral proto-oncogene 1		
66	ARBK1_MOUSE	adrenergic receptor kinase, beta 1		
67	ATF2_MOUSE	activating transcription factor 2		
68	ATFF2_MOUSE	activating transcription factor 2		
69	AVR2B_MOUSE	activin receptor III		
70	BAMBI_MOUSE	BMP and activin membrane-bound inhibitor, homolog (Xenopus laevis)		
71	BMP2_MOUSE	bone morphogenetic protein 2		
72	BMP4_MOUSE	bone morphogenetic protein 4		
73	BMP6_MOUSE	bone morphogenetic protein 6		
74	BMP7_MOUSE	bone morphogenetic protein 7		
75	BMPR2_MOUSE	bone morphogenetic protein receptor, type II (serine/threonine kinase)		
76	BMR1A_MOUSE	bone morphogenetic protein receptor, type 1A		
77	BMR1B_MOUSE	bone morphogenetic protein receptor, type 1B		
78	CER1_MOUSE	cerberus 1 homolog (Xenopus laevis)		
79	CHRD_MOUSE	chordin		
80	CITE1_MOUSE	Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 1		
81	CITE2_MOUSE	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2		
82	CRDL1_MOUSE	chordin-like 1		
83	CRDL2_MOUSE	chordin-like 2		
84	CREB1_MOUSE	cAMP responsive element binding protein 1		
85	CTDSP1_MOUSE	C-terminal binding protein 1		
86	CTDS1_MOUSE	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase 1		
87	CTDSL_MOUSE	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like		
88	CUL1_MOUSE	cullin 1		
89	DAB2_MOUSE	disabled homolog 2 (Drosophila)		

bioCompendium
The High-throughput experimental data analysis platform

Summary sheet information for given gene ENSG00000165029

Name: ABCA1
Description: ATP-binding cassette sub-family A member 1 (ATP-binding cassette transporter 1) (ABC1) (Cholesterol efflux regulatory protein)
[Source: UniProtKB/Swiss-Prot; Acc: Q96J77]

Protein Information from Dastyl2 DAS Client

Expand All | Collapse All

Gene Location:
Ensembl Genes:
EMBL:
EntrezGene:
RefSeq DNA:
UniGene:

Ensembl Protein:
Entrez Protein:
RefSeq Protein:
UniProt:

ID: A5068_HUMAN SubName: Full=ABCA1 protein;
B1A1M1_HUMAN SubName: Full=ATP-binding cassette, sub-family A (ABC1), member 1;
B1A2M1_HUMAN SubName: Full=ATP-binding cassette, sub-family A (ABC1), member 1, isoform CRA_a;
B2RUL2_HUMAN SubName: Full=ATP-binding cassette, sub-family A (ABC1), member 1;
Q5HCD2_HUMAN SubName: Full=ATP-binding cassette, sub-family A member 1 variant; Flags: Fragment;
Q5KQ33_HUMAN UNKNOWN
Q5H022_HUMAN SubName: Full=ATP-binding cassette transporter 1; Flags: Fragment;
Q5HT8_HUMAN UNKNOWN
Q5P93_HUMAN SubName: Full=ATP binding cassette transporter 1; Flags: Fragment;
Q5N78_HUMAN RecName: Full=ATP-binding cassette sub-family A member 1; AltName: Full=ATP-binding cassette transporter 1; Short=ABC1; AliasName: Full=Cholesterol efflux regulatory protein;

Human Protein Atlas
MiCoCheck
Link to MiCoCheckDB

SMART:
Hyperlink to SMART; Hyperlink to ENSEMBL

ABC1

InterPro:
Pfam:
PDB:
HSSP:
PSSM:
OMIM:
OMIM ID: 60048 Name: ATP-BINDING CASSETTE, SUBFAMILY A, MEMBER 1; ABCA1 AltName: ATP-BINDING CASSETTE TRANSPORTER 1 ABC TRANSPORTER 1 CHOLESTEROL EFFLUX REGULATORY PROTEIN; CERP CORONARY HEART DISEASE IN FAMILIAL HYPERCHOLESTEROLEMIA, INCLUDED HIGH DENSITY LIPOPROTEIN CHOLESTEROL LEVEL QUANTITATIVE TRAIT LOCUS 13, INCLUDED; HLOC213, INCLUDED

GO:
KEGG:
Panther:

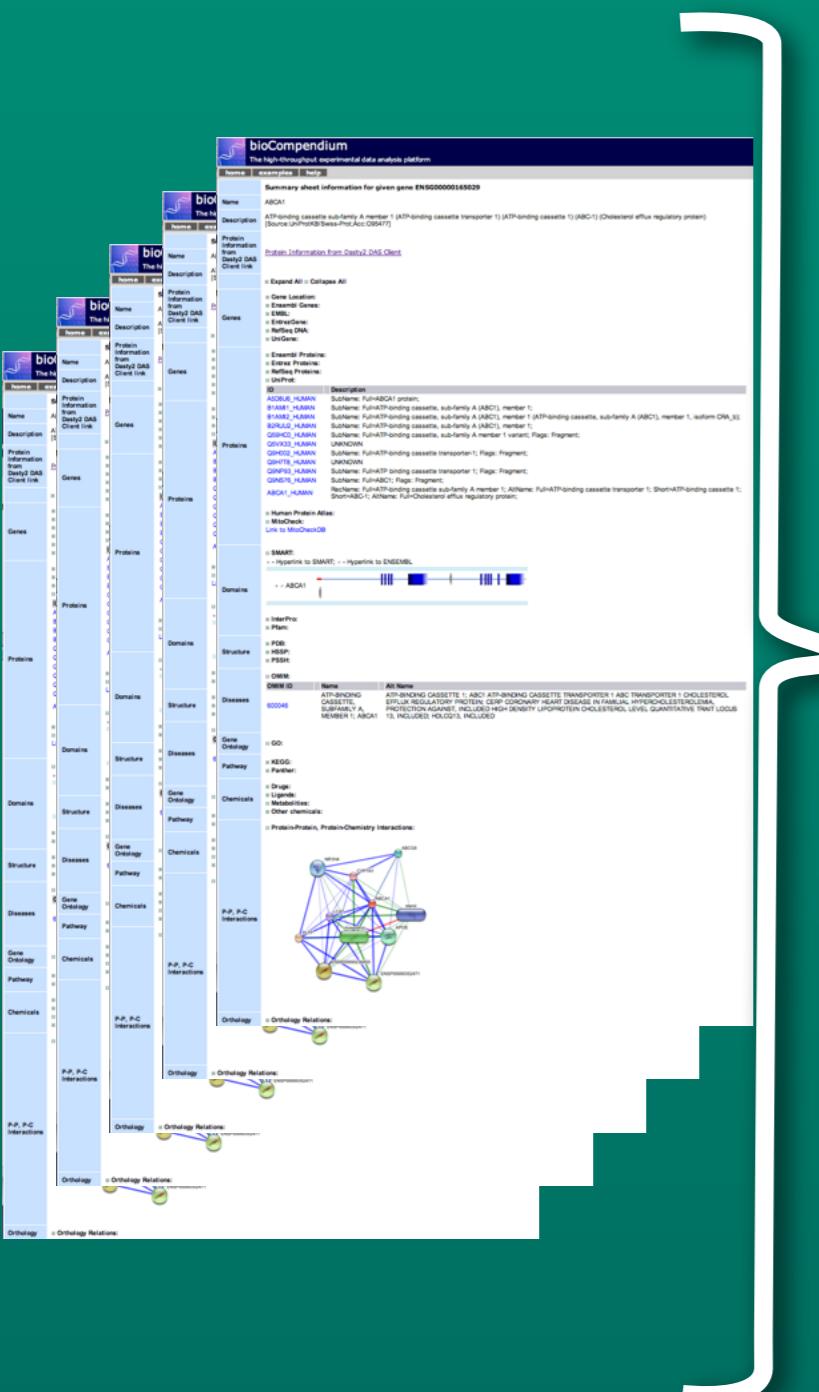
Pathway:
Gene Ontology:
Chemical:
Drugs:
UpGene:
Metabolites:
Other chemicals:

Protein-Protein, Protein-Chemistry Interactions:

Orthology Relations:
Orthology Relations:
Orthology Relations:
Orthology Relations:
Orthology Relations:

Single sequence information (approx. 40 subsections)

Clustering based on sequence similarity and domain architecture



Description Homology Clustering Details

No. of Clusters	46
Expand All □ Collapse All	

Clusters:

- # Cluster: 1 Myeloid zinc finger 1 (MZF-1) (Zinc finger protein 42) (Zinc finger alpha)
- # Cluster: 2 Phosphatidylinositol 3-kinase regulatory subunit beta (PI3-kinase zeta)
- # Cluster: 3 Ribosomal protein S6 kinase beta-1 [EC 2.7.11.1] (Ribosomal protein (p70(S6K)-alpha)) [Source:UniProtKB/Swiss-Prot;Acc:P23443]
- # Cluster: 4 Transforming growth factor beta-1 Precursor (TGF-beta-1) (LAP)
- # Cluster: 5 Transcription factor E2F5 (E2F-5) [Source:UniProtKB/Swiss-Prot;Acc:Q02464]

Gene Description

Smart domain cluster details

— Hyperlink to SMART; — Hyperlink to ENSEMBL

Cluster ID	Gene	Description	SMART ID	Start	Stop	Definition	Description
ZnF_C2H2	MZF1	SCAN	SM00355	541	563	zinc finger C2H2	Close
ZnF_C2H2	ZFP64						
ZnF_C2H2	EGR2						
ZnF_C2H2	KLF2						
ZnF_C2H2	YY1						
ZnF_C2H2	KLF10						
ZnF_C2H2	EGR3						
ZnF_C2H2	SP1						

Clustering based on sequence similarity and domain architecture

Description Homology Clustering Details

No. of Clusters: 46

- Expand All | Collapse All
- Smart domain cluster details

Cluster	Gene	Description
1	MZF1	Myeloid zinc finger 1 (Myb-Zip finger protein 42) (Myb) (Myb-like Zinc finger protein)
2	ZFP64	Proline-rich transcription factor ZFP64 (Zinc finger protein 64)
3	EGR2	Protein kinase C-inducible gene response protein 2 (EGR2)
4	KLF2	Transcription factor E2F1 (E2F1) (Dorsal Unpaired-Dense-Proline-rich)
5	YY1	Transcription factor YY1 (YY1)
6	KLF10	Transcription factor YY1-binding factor 10 (KLF10)
7	EGR3	Protein kinase C-inducible gene response protein 3 (EGR3)
8	SP1	AP-1 transcription factor subunit p65 (Activator Protein 1 Subunit p65)
9	YY2	Transcription factor YY2 (YY2)
10	YY3	Transcription factor YY3 (YY3)
11	YY4	Transcription factor YY4 (YY4)
12	YY5	Transcription factor YY5 (YY5)
13	YY6	Transcription factor YY6 (YY6)
14	YY7	Transcription factor YY7 (YY7)
15	YY8	Transcription factor YY8 (YY8)
16	YY9	Transcription factor YY9 (YY9)
17	YY10	Transcription factor YY10 (YY10)
18	YY11	Transcription factor YY11 (YY11)
19	YY12	Transcription factor YY12 (YY12)
20	YY13	Transcription factor YY13 (YY13)
21	YY14	Transcription factor YY14 (YY14)
22	YY15	Transcription factor YY15 (YY15)
23	YY16	Transcription factor YY16 (YY16)
24	YY17	Transcription factor YY17 (YY17)
25	YY18	Transcription factor YY18 (YY18)
26	YY19	Transcription factor YY19 (YY19)
27	YY20	Transcription factor YY20 (YY20)
28	YY21	Transcription factor YY21 (YY21)
29	YY22	Transcription factor YY22 (YY22)
30	YY23	Transcription factor YY23 (YY23)
31	YY24	Transcription factor YY24 (YY24)
32	YY25	Transcription factor YY25 (YY25)
33	YY26	Transcription factor YY26 (YY26)
34	YY27	Transcription factor YY27 (YY27)
35	YY28	Transcription factor YY28 (YY28)
36	YY29	Transcription factor YY29 (YY29)
37	YY30	Transcription factor YY30 (YY30)
38	YY31	Transcription factor YY31 (YY31)
39	YY32	Transcription factor YY32 (YY32)
40	YY33	Transcription factor YY33 (YY33)
41	YY34	Transcription factor YY34 (YY34)
42	YY35	Transcription factor YY35 (YY35)
43	YY36	Transcription factor YY36 (YY36)
44	YY37	Transcription factor YY37 (YY37)
45	YY38	Transcription factor YY38 (YY38)
46	YY39	Transcription factor YY39 (YY39)

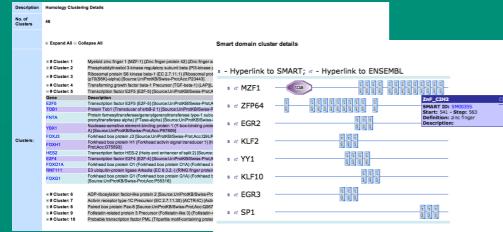
Pathway and GO terms enrichment

Molecular function:

Accession	Term	Adjusted P-Value
GO:0042625	ATPase activity, coupled to transmembrane movement of ions	6e-80
GO:0042623	ATPase activity, coupled	4e-54
GO:0016887	ATPase activity	1e-53
GO:0016817	hydrolase activity, acting on acid anhydrides	4e-38
GO:0022857	transmembrane transporter activity	1e-35
GO:0022892	substrate-specificity	2e-35
GO:0015078	hydrogen-e-	2e-27
GO:0008324	cation trans-	4e-27
GO:0046961	proton-trans-	7e-24
GO:0000287	magnesium	4e-23
GO:0016787	hydrolase	1e-22
GO:0032553	ribonucleot	4e-20
GO:0017076	purine nucl	2e-19
GO:0005388	calcium-tra	8e-18
GO:0008553	hydrogen-e-	3e-14
GO:0005391	sodium:pol	4e-09
GO:0032135	DNA insert	4e-09

Pathway:

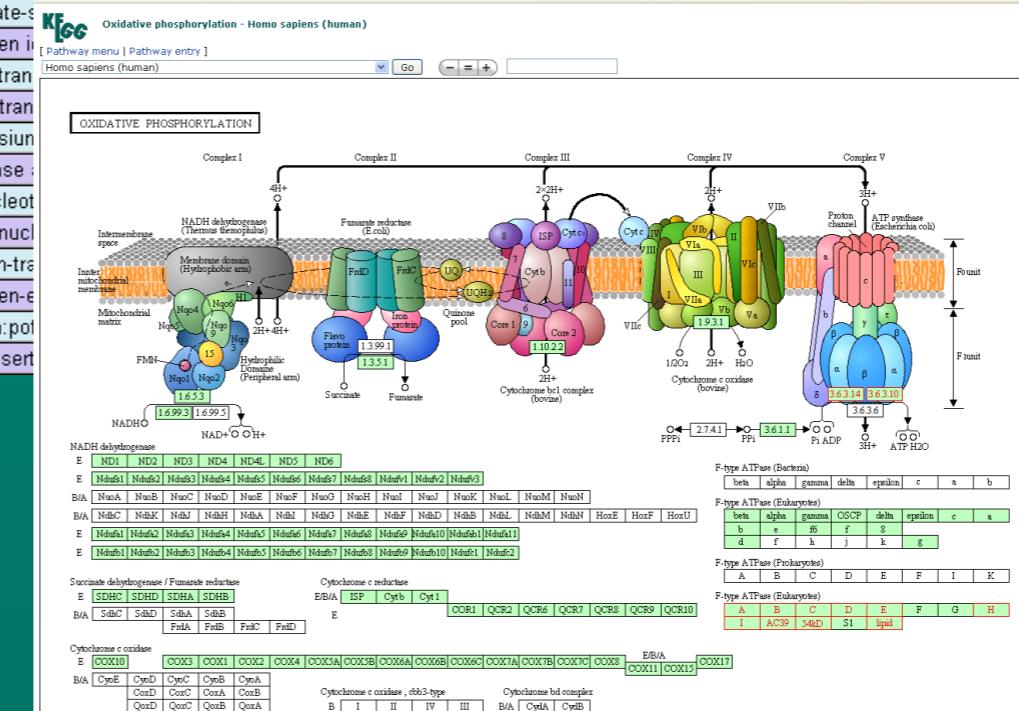
The diagram illustrates the mitochondrial electron transport chain. Electrons flow from NADH + H+ through Complex I (Rieske Iron-Sulfur Protein) and Complex II (Flavoprotein) to Complex III (Cytochrome bc1 complex). From Complex III, electrons enter the Cytochrome c pool, which then passes through Complex IV (Cytochrome c oxidase) to Complex V (ATP synthase). The energy from this transport drives the synthesis of ATP. The overall process is labeled "OXIDATIVE PHOSPHORYLATION".



Pathway and GO terms enrichment

Molecular function:

Accession	Term	Adjusted P-Value
GO:0042625	ATPase activity, coupled to transmembrane movement of ions	6e-80
GO:0042623	ATPase activity, coupled	4e-54
GO:0016887	ATPase activity	1e-53
GO:0016817	hydrolase activity, acting on acid anhydrides	4e-38
GO:0022857	transmembrane transporter activity	1e-35
GO:0022892	substrate-specificity	2e-35
GO:0015078	hydrogen ion transmembrane transporter activity	2e-27
GO:0008324	cation transport	4e-27
GO:0046961	proton-translocating ATPase activity	7e-24
GO:0000287	magnesium ion transmembrane transporter activity	4e-23
GO:0016787	hydrolase activity	1e-22
GO:0032553	ribonucleotidreducere	4e-20
GO:0017076	purine nucleoside phosphorylase activity	2e-19
GO:0005388	calcium-translocating ATPase activity	8e-18
GO:0008553	hydrogen-ejecting ATPase activity	3e-14
GO:0005391	sodium:potassium translocating ATPase activity	4e-09
GO:0032135	DNA insertion	4e-09



Find: an in Next Previous Highlight all Match case

Clustering based on sequence similarity and domain architecture

Number of Clusters: 46

Homology Clustering Details

- Expand All | Collapse All
- Cluster 1: Myeloid zinc finger 1 (MZF1) (Zinc finger protein 42) (Zinc finger protein 42) (Zinc finger protein 42) (Zinc finger protein 42)
- Cluster 2: Zinc finger protein 58 (ZFP58) (Zinc finger protein 58)
- Cluster 3: Zinc finger protein 58-like 1 (ZFP58L1) (Zinc finger protein 58-like 1) (Zinc finger protein 58-like 1)
- Cluster 4: Transforming growth factor beta 1 (TGFbeta1) (Transforming growth factor beta 1) (Transforming growth factor beta 1)
- Cluster 5: Zinc finger protein 34 (ZFP34) (Zinc finger protein 34) (Zinc finger protein 34)
- Cluster 6: Zinc finger protein 34-like 1 (ZFP34L1) (Zinc finger protein 34-like 1) (Zinc finger protein 34-like 1)
- Cluster 7: Zinc finger protein 34-like 2 (ZFP34L2) (Zinc finger protein 34-like 2) (Zinc finger protein 34-like 2)
- Cluster 8: Zinc finger protein 34-like 3 (ZFP34L3) (Zinc finger protein 34-like 3) (Zinc finger protein 34-like 3)
- Cluster 9: Zinc finger protein 34-like 4 (ZFP34L4) (Zinc finger protein 34-like 4) (Zinc finger protein 34-like 4)
- Cluster 10: Zinc finger protein 34-like 5 (ZFP34L5) (Zinc finger protein 34-like 5) (Zinc finger protein 34-like 5)
- Cluster 11: Zinc finger protein 34-like 6 (ZFP34L6) (Zinc finger protein 34-like 6) (Zinc finger protein 34-like 6)
- Cluster 12: Zinc finger protein 34-like 7 (ZFP34L7) (Zinc finger protein 34-like 7) (Zinc finger protein 34-like 7)
- Cluster 13: Zinc finger protein 34-like 8 (ZFP34L8) (Zinc finger protein 34-like 8) (Zinc finger protein 34-like 8)
- Cluster 14: Zinc finger protein 34-like 9 (ZFP34L9) (Zinc finger protein 34-like 9) (Zinc finger protein 34-like 9)
- Cluster 15: Zinc finger protein 34-like 10 (ZFP34L10) (Zinc finger protein 34-like 10) (Zinc finger protein 34-like 10)
- Cluster 16: Zinc finger protein 34-like 11 (ZFP34L11) (Zinc finger protein 34-like 11) (Zinc finger protein 34-like 11)
- Cluster 17: Zinc finger protein 34-like 12 (ZFP34L12) (Zinc finger protein 34-like 12) (Zinc finger protein 34-like 12)
- Cluster 18: Zinc finger protein 34-like 13 (ZFP34L13) (Zinc finger protein 34-like 13) (Zinc finger protein 34-like 13)
- Cluster 19: Zinc finger protein 34-like 14 (ZFP34L14) (Zinc finger protein 34-like 14) (Zinc finger protein 34-like 14)
- Cluster 20: Zinc finger protein 34-like 15 (ZFP34L15) (Zinc finger protein 34-like 15) (Zinc finger protein 34-like 15)
- Cluster 21: Zinc finger protein 34-like 16 (ZFP34L16) (Zinc finger protein 34-like 16) (Zinc finger protein 34-like 16)
- Cluster 22: Zinc finger protein 34-like 17 (ZFP34L17) (Zinc finger protein 34-like 17) (Zinc finger protein 34-like 17)
- Cluster 23: Zinc finger protein 34-like 18 (ZFP34L18) (Zinc finger protein 34-like 18) (Zinc finger protein 34-like 18)
- Cluster 24: Zinc finger protein 34-like 19 (ZFP34L19) (Zinc finger protein 34-like 19) (Zinc finger protein 34-like 19)
- Cluster 25: Zinc finger protein 34-like 20 (ZFP34L20) (Zinc finger protein 34-like 20) (Zinc finger protein 34-like 20)
- Cluster 26: Zinc finger protein 34-like 21 (ZFP34L21) (Zinc finger protein 34-like 21) (Zinc finger protein 34-like 21)
- Cluster 27: Zinc finger protein 34-like 22 (ZFP34L22) (Zinc finger protein 34-like 22) (Zinc finger protein 34-like 22)
- Cluster 28: Zinc finger protein 34-like 23 (ZFP34L23) (Zinc finger protein 34-like 23) (Zinc finger protein 34-like 23)
- Cluster 29: Zinc finger protein 34-like 24 (ZFP34L24) (Zinc finger protein 34-like 24) (Zinc finger protein 34-like 24)
- Cluster 30: Zinc finger protein 34-like 25 (ZFP34L25) (Zinc finger protein 34-like 25) (Zinc finger protein 34-like 25)
- Cluster 31: Zinc finger protein 34-like 26 (ZFP34L26) (Zinc finger protein 34-like 26) (Zinc finger protein 34-like 26)
- Cluster 32: Zinc finger protein 34-like 27 (ZFP34L27) (Zinc finger protein 34-like 27) (Zinc finger protein 34-like 27)
- Cluster 33: Zinc finger protein 34-like 28 (ZFP34L28) (Zinc finger protein 34-like 28) (Zinc finger protein 34-like 28)
- Cluster 34: Zinc finger protein 34-like 29 (ZFP34L29) (Zinc finger protein 34-like 29) (Zinc finger protein 34-like 29)
- Cluster 35: Zinc finger protein 34-like 30 (ZFP34L30) (Zinc finger protein 34-like 30) (Zinc finger protein 34-like 30)
- Cluster 36: Zinc finger protein 34-like 31 (ZFP34L31) (Zinc finger protein 34-like 31) (Zinc finger protein 34-like 31)
- Cluster 37: Zinc finger protein 34-like 32 (ZFP34L32) (Zinc finger protein 34-like 32) (Zinc finger protein 34-like 32)
- Cluster 38: Zinc finger protein 34-like 33 (ZFP34L33) (Zinc finger protein 34-like 33) (Zinc finger protein 34-like 33)
- Cluster 39: Zinc finger protein 34-like 34 (ZFP34L34) (Zinc finger protein 34-like 34) (Zinc finger protein 34-like 34)
- Cluster 40: Zinc finger protein 34-like 35 (ZFP34L35) (Zinc finger protein 34-like 35) (Zinc finger protein 34-like 35)
- Cluster 41: Zinc finger protein 34-like 36 (ZFP34L36) (Zinc finger protein 34-like 36) (Zinc finger protein 34-like 36)
- Cluster 42: Zinc finger protein 34-like 37 (ZFP34L37) (Zinc finger protein 34-like 37) (Zinc finger protein 34-like 37)
- Cluster 43: Zinc finger protein 34-like 38 (ZFP34L38) (Zinc finger protein 34-like 38) (Zinc finger protein 34-like 38)
- Cluster 44: Zinc finger protein 34-like 39 (ZFP34L39) (Zinc finger protein 34-like 39) (Zinc finger protein 34-like 39)
- Cluster 45: Zinc finger protein 34-like 40 (ZFP34L40) (Zinc finger protein 34-like 40) (Zinc finger protein 34-like 40)
- Cluster 46: Zinc finger protein 34-like 41 (ZFP34L41) (Zinc finger protein 34-like 41) (Zinc finger protein 34-like 41)

Pathway and GO terms enrichment

Annotation	Tests	Adjusted P Value
Zinc finger protein 34-like 1	ATPase activity, coupled to transmembrane movement of ions	0.0000000000000001
Zinc finger protein 34-like 1	ATPase activity	0.0000000000000001
Zinc finger protein 34-like 1	Protein binding	0.0000000000000001
Zinc finger protein 34-like 1	Transmembrane transporter activity	0.0000000000000001
Zinc finger protein 34-like 1	Hydrogen bond formation	0.0000000000000001
Zinc finger protein 34-like 1	Protein binding	0.0000000000000001
Zinc finger protein 34-like 1	Chloride ion binding	0.0000000000000001
Zinc finger protein 34-like 1	ATPase activity, coupled to transmembrane movement of ions	0.0000000000000001
Zinc finger protein 34-like 1	ATPase activity	0.0000000000000001
Zinc finger protein 34-like 1	Protein binding	0.0000000000000001
Zinc finger protein 34-like 1	Transmembrane transporter activity	0.0000000000000001
Zinc finger protein 34-like 1	Hydrogen bond formation	0.0000000000000001
Zinc finger protein 34-like 1	Protein binding	0.0000000000000001
Zinc finger protein 34-like 1	Chloride ion binding	0.0000000000000001

Clinical trials and patent information

ClinicalTrials.gov

European Patent Office

espacenet 1998-2008

GENETICALLY ALTERED ANTIBODY-PRODUCING CELL LINES WITH IMPROVED ANTIBODY CHARACTERISTICS

Bibliographic data **Description** **Claims** **Mosaics** **Original document** **INPADOC legal status**

Publication numbers: WO2005023865 (A2)

Publication date: 2005-03-17

Inventor(s): GRASSO LUIGI [US]; NICOLAIDES NICHOLAS C [US]; SASS PHILIP M [US]

Applicant(s): MORPHOTEK INC [US]; GRASSO LUIGI [US]; NICOLAIDES NICHOLAS C [US]; SASS PHILIP M [US]

Classification:

- International:** C07H21/04; C07K16/40; C12N5/12; C12N5/16; C12N15/10; C07K16/00; C07K16/90; C07K16/42; C12N5/12; C12N5/16; C12N15/10; (PCT-7); C07K16/00
- European:** C12H15/10B1; C07H21/04; C07K16/00; C07K16/40; C12N5/12; C12N5/16; C12N15/10;

Application numbers: WO2004US229905 20040903

Priority number(s): US2003050071P 20030903; US20040933034 20040902

View INPADOC patent family **View list of citing documents** **View document in the European Register**

Abstract of WO 2005023865 (A2)

Dominant negative alleles of human mismatch repair genes can be used to generate hypermutable cells and organisms. Cells may be selected for expression of activation-induced cytidine deaminase (AID), stimulated to produce AID, or manipulated to express AID for further enhancement of hypermutability. These methods are useful for generating genetic diversity within immunoglobulin genes directed against an antigen of interest to produce altered antibodies with enhanced biochemical activity. Moreover, these methods are useful for generating antibody-producing cells with increased level of antibody production.

Data supplied from the espacenet database — Worldwide

Clustering based on sequence similarity and domain architecture

Description		Homology Clustering Details	
No. of Clusters	46		
			≡ Expand All Collapse All
			≡ Hyperlink to SMART; ≡ Hyperlink to ENSEMBL
Clusters:		<p>≡ Expanded All Collapse All</p> <p>≡ Cluster 1 Methyl zinc finger 1 (Zinc finger protein 42) [Zinc finger protein 42] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 2 Phosphotyrosine 3 kinase regulatory subunit delta (T3 kinase regulatory subunit delta) [T3 kinase regulatory subunit delta] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 3 Zinc finger protein 145 (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 4 Transcription factor E2F-1 (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 5 Transcription factor E2F-2 (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 6 Transcription factor E2F-3 (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 7 TSHZ3 (Thyroid hormone receptor beta-associated protein) [Thyroid hormone receptor beta-associated protein] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 8 FNTA (Finger protein 1A) [Finger protein 1A] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 9 Zinc finger protein 143 (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 10 AT (Alpha 1B/Protein kinase C eta/Protein kinase C eta) [Alpha 1B/Protein kinase C eta/Protein kinase C eta] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 11 FKBP12 (FK506-binding protein 12) [FK506-binding protein 12] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 12 FKBP12R (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 13 FKBP12R1 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 14 FKBP12R2 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 15 FKBP12R3 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 16 FKBP12R4 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 17 FKBP12R5 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 18 FKBP12R6 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 19 FKBP12R7 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 20 FKBP12R8 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 21 FKBP12R9 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 22 FKBP12R10 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 23 FKBP12R11 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 24 FKBP12R12 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 25 FKBP12R13 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 26 FKBP12R14 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 27 FKBP12R15 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 28 FKBP12R16 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 29 FKBP12R17 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 30 FKBP12R18 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 31 FKBP12R19 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 32 FKBP12R20 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 33 FKBP12R21 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 34 FKBP12R22 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 35 FKBP12R23 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 36 FKBP12R24 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 37 FKBP12R25 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 38 FKBP12R26 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 39 FKBP12R27 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 40 FKBP12R28 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 41 FKBP12R29 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 42 FKBP12R30 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 43 FKBP12R31 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 44 FKBP12R32 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 45 FKBP12R33 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p> <p>≡ Cluster 46 FKBP12R34 (FK506-binding protein 12-related) [FK506-binding protein 12-related] (Source: UniProtKB/Swiss-Prot; Organism: Human)</p>	
		<p>≡ MZF1</p> <p>≡ ZFP64</p> <p>≡ EGR2</p> <p>≡ KLF2</p> <p>≡ Y1</p> <p>≡ KLF10</p> <p>≡ EGR3</p> <p>≡ SP1</p>	<p>≡ EXP_CDS</p> <p>≡ SMART_3D</p> <p>≡ PDB_3D</p> <p>≡ Super_3D</p> <p>≡ Description: zinc Finger</p>

Pathway and GO terms enrichment

The figure shows a detailed view of a protein's molecular functions and associated GO terms. It also includes a large diagram of the protein complex, showing various subunits and their interactions.

Clinical trials and patent information

ClinicalTrials.gov

Full Text View | Table View | My Study Results Portal | Related Studies

Study of Live Attenuated Japanese Encephalitis Vaccine (Chenovirax™-JE) and Yellow Fever Vaccine (STAMARL™)

This study has been completed.

First Received: September 21, 2000; Last Update: September 21, 2000; No Change

Sponsored by: [redacted]
Information provided by: [redacted]

ClinicalTrials.gov Identifier: NCT00000000

Purpose

The purpose of this study is to obtain safety, tolerability, and immunogenicity data on the co-administration or sequential administration of Chenovirax™-JE and STAMARL™.

Safety:

- Obtain safety and tolerability data of a single, fixed dose of Chenovirax™-JE administered concurrently, one month before or one month after vaccination.

Immunogenicity:

- Assess the antibody response to a single, fixed dose of Chenovirax™-JE administered concurrently, one month before or one month after vaccination.
- Assess the durability of the immune response in adult volunteers 6 months after administration of Chenovirax™-JE and STAMARL™.

Condition	Intervention
Japanese Encephalitis	Single-dose live attenuated Japanese encephalitis virus - Yellow fever virus
Yellow Fever	Single-dose live attenuated Japanese encephalitis virus - Yellow fever virus

Study Type: Interventional
Study Design: Parallel Assignment, Randomized, Double Blind (Subject), Investigator, Outcomes Assessed, Active Control, Parallel Assignment
Official Title: Randomized, Double-blind, Phase I Evaluation of the Safety and Immunogenicity Following Administration of Live Attenuated JE and STAMARL™

Resource links provided by NCTI:

- MedlinePlus related topics: [Japanese Encephalitis](#)
- Drug Information available for: [Yellow Fever Vaccine](#)
- [USA.gov Resources](#)

Further study details as provided by Saudi Arabia:

Primary Outcome Measures:

- To provide information concerning the immunogenicity of Chenovirax™-JE and STAMARL™ after co-administration or sequential vaccination.
- To provide information concerning the safety after co-administration or sequential administration of Chenovirax™-JE and STAMARL™.

Secondary Outcome Measures:

- Assess the durability of the immune response in adult volunteers 6 months after administration of Chenovirax™-JE and STAMARL™.

Recruitment: [100](#) | [Open](#) | [Recruiting](#) | [Not yet recruiting](#) | [Completed](#)

[View all records](#) | [Print](#)

European Patent Office

Home | Content | English | Deutsch | Français | Help index | [Search](#)

Quick Search

Advanced Search

Number Search

Last result list

My patents list 0

Classification Search

Help

Quick Help

- > Why are some tabs deactivated for certain documents?
- > Why does a list of documents with the heading "Also published as" sometimes appear, and what are these documents?
- > What does A1, A2, A3 and B stand for after an EP publication number in the "Also published as" list?
- > What is a cited document?
- > What are citing documents?
- > What information will I find if I click on the link "View document in the European Register"?
- > Why do I sometimes find the abstract of a corresponding document?

Bibliographic data

Description | **Claims** | **Mosaics** | **Original document** | **INPADOC legal status**

Publication number: WO2005023865 (A2)

Publication date: 2005-03-17

Inventor(s): GRASSO LUIGI [US]; NICOLAIDES NICHOLAS C [US]; SASS PHILIP M [US]

Applicant(s): MORPHOTEK INC [US]; GRASSO LUIGI [US]; NICOLAIDES NICHOLAS C [US]; SASS PHILIP M [US]

Classification:

- International: C07H21/04; C07K16/00; C07K16/42; C12N5/12; C12N5/16; C12N5/18; C07H21/00; C07K16/00; C07K16/42; C12N5/12; C12N5/16; C12N5/18; (IPC1-7); C07K16/00
- European: C12N5/10B1; C07H21/04; C07K16/00; C07K16/42M10; C12N5/12

Application number: WO2004/US28905 20040903

Priority number(s): US20030600071P 20030903; US20040933034 20040902

Also published as:

- WO2005023865 (A3)
- US2005048621 (A1)
- JP2007525973 (T)
- EP1680448 (A2)
- CA2537881 (A1)

[more >](#)

Cited documents:

- WO20037967 (A1)
- WO200361363 (A2)
- XP000945402 (A)
- XP0002321656 (A)
- XP0002253537 (A)

[more >](#)

View INPADOC patent family

View list of citing documents

View document in the European Register

Report a data error here

Abstract of WO 2005023865 (A2)

Dominant negative alleles of human mismatch repair genes can be used to generate hypermutable cells and organisms. Cells may be selected for expression of activation-induced cytidine deaminase (AID), stimulated to produce AID, or manipulated to express AID for further enhancement of hypermutability. These methods are useful for generating genetic diversity within immunoglobulin genes directed against an antigen of interest to produce altered antibodies with enhanced biochemical activity. Moreover, these methods are useful for generating antibody-producing cells with increased level of antibody production.

Data supplied from the espacenet database — Worldwide

Clustering based on sequence similarity and domain architecture

The figure consists of two main parts. The top part is a hierarchical tree diagram titled "Homology Clustering Details" showing the relationship between 48 clusters. The bottom part is a detailed view of the "Smart domain cluster details" for cluster 1.

Hierarchical Tree (Top):

- Root: Homology Clustering Details
- Level 1: No. of Clusters (48)
- Level 2: Expand All / Collapse All
- Level 3: Cluster 1 (1 SMART domain) → Cluster 2 (1 SMART domain) → Cluster 3 (1 SMART domain) → Cluster 4 (1 SMART domain) → Cluster 5 (1 SMART domain) → Cluster 6 (1 SMART domain) → Cluster 7 (1 SMART domain) → Cluster 8 (1 SMART domain) → Cluster 9 (1 SMART domain) → Cluster 10 (1 SMART domain)
- Level 4: Cluster 11 (1 SMART domain) → Cluster 12 (1 SMART domain) → Cluster 13 (1 SMART domain) → Cluster 14 (1 SMART domain) → Cluster 15 (1 SMART domain) → Cluster 16 (1 SMART domain) → Cluster 17 (1 SMART domain) → Cluster 18 (1 SMART domain) → Cluster 19 (1 SMART domain) → Cluster 20 (1 SMART domain)
- Level 5: Cluster 21 (1 SMART domain) → Cluster 22 (1 SMART domain) → Cluster 23 (1 SMART domain) → Cluster 24 (1 SMART domain) → Cluster 25 (1 SMART domain) → Cluster 26 (1 SMART domain) → Cluster 27 (1 SMART domain) → Cluster 28 (1 SMART domain) → Cluster 29 (1 SMART domain) → Cluster 30 (1 SMART domain)
- Level 6: Cluster 31 (1 SMART domain) → Cluster 32 (1 SMART domain) → Cluster 33 (1 SMART domain) → Cluster 34 (1 SMART domain) → Cluster 35 (1 SMART domain) → Cluster 36 (1 SMART domain) → Cluster 37 (1 SMART domain) → Cluster 38 (1 SMART domain) → Cluster 39 (1 SMART domain) → Cluster 40 (1 SMART domain)
- Level 7: Cluster 41 (1 SMART domain) → Cluster 42 (1 SMART domain) → Cluster 43 (1 SMART domain) → Cluster 44 (1 SMART domain) → Cluster 45 (1 SMART domain) → Cluster 46 (1 SMART domain) → Cluster 47 (1 SMART domain) → Cluster 48 (1 SMART domain)

Smart domain cluster details (Bottom):

Cluster 1 (1 SMART domain) contains 1 SMART domain (zinc finger). The SMART domain is shown as a blue rectangle with a purple circle representing the zinc finger motif. Below the SMART domain, a legend indicates:

- Hyperlink to SMART: A solid blue line segment.
- Hyperlink to ENSEMBL: A dashed blue line segment.

 The SMART domain is associated with the following proteins:

- MZF1
- ZFP64
- EGR2
- KLF2
- YY1
- KLF10
- EGR3
- SP1

 The legend also includes:

- SMART ID: SMART00000000000000000000000000000000
- Description: zinc Finger
- Definition: Zinc finger

Pathway and GO terms enrichment

Clinical trials and patent information

A screenshot of a web browser showing search results for 'Clinical trials and patent information'. The top navigation bar includes links for 'Home', 'Search', 'Help', 'About', and 'Logout'. The main content area has a header 'Clinical trials and patent information' with a sub-header 'Only at ClinicalTrials.gov'. Below this is a search bar with placeholder text 'Search ClinicalTrials.gov' and a dropdown menu for 'Advanced Search'. A sidebar on the left lists 'Recent searches' and 'My profile'. The main search results are displayed in a table with columns: 'Title', 'Status', 'Condition', 'Intervention', 'Enrollment', 'Last updated', 'Last modified', and 'Last checked'. One result is highlighted: 'Genetic analysis of T-lymphocyte-producing cell lines with informed antibody characteristics' by 'CARTISAN GEN INC'. It shows '1 trial' with '1 study' listed, status 'Active', condition 'Immunotherapy', intervention 'Genetic engineering', enrollment 'Recruiting', last updated '2005-01-01', last modified '2005-01-01', and last checked '2005-01-01'. The right side of the page features a sidebar with links to 'espcinet', 'espcinet 2005-2006', 'espcinet 2004-2005', 'espcinet 2003-2004', 'espcinet 2002-2003', 'espcinet 2001-2002', 'espcinet 2000-2001', 'espcinet 1999-2000', and 'espcinet 1998-1999'. A footer at the bottom right says 'Data captured from the espcinet database - Worldwide'.

Chemistry information

- DrugBank, hmdb, STITCH, MATADOR, chEBI
 - PubChem, PDBLigandetc

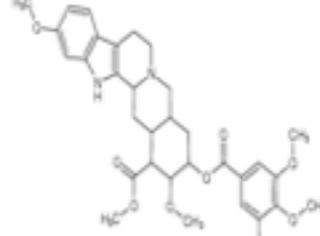
<input checked="" type="checkbox"/> AASS	Alpha-aminoadipic semialdehyde synthase, mitochondrial precursor Saccharopine dehydrogenase (EC 1.5.1.9) (SDH). [Source:Uniprot] <input checked="" type="checkbox"/> Drugs <input checked="" type="checkbox"/> Ligands <input checked="" type="checkbox"/> Metabolites <input checked="" type="checkbox"/> Others
<input checked="" type="checkbox"/> ABCA1	ATP-binding cassette sub-family A member 1 (ATP-binding cassette) [Source:Uniprot/SWISSPROT;Acc:095477] <input checked="" type="checkbox"/> Drugs <input checked="" type="checkbox"/> Ligands <input checked="" type="checkbox"/> Metabolites <input checked="" type="checkbox"/> Others

BilB011 Bil salt export pump (ATP-binding cassette sub-family B member 11). [Source Uniprot/SWISSPROT;Acc: Q854Q]

Drugs

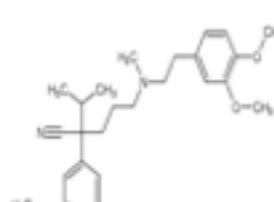
rot

Name: reserpine
 Source: [Mycobacterium tuberculosis](#)
 Uniprot id: [Q854Q](#)
 PubChem cat: [200](#)

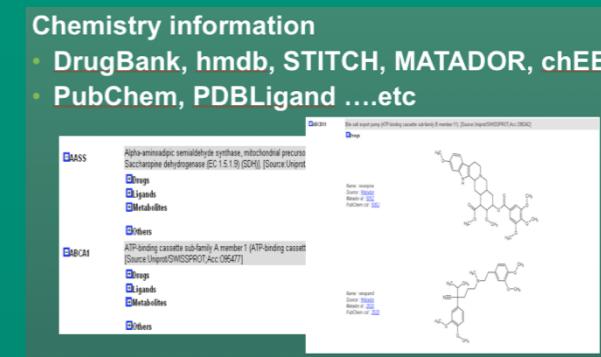
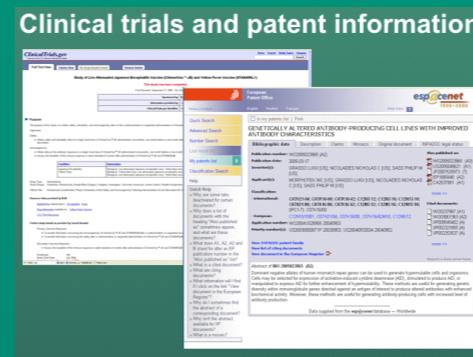
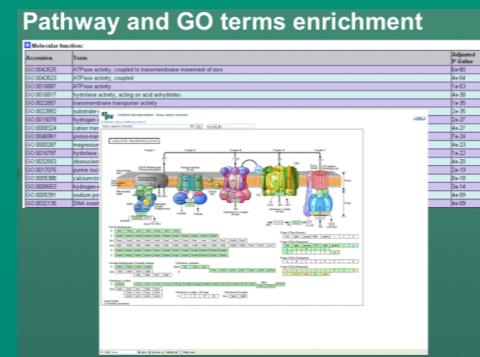
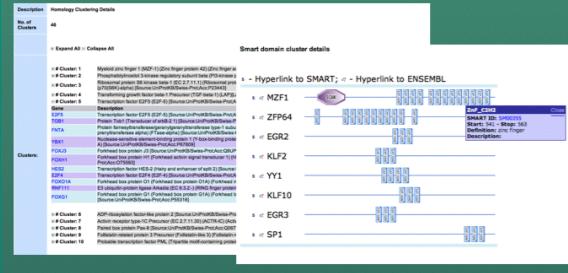


ett

Name: venusimil
 Source: [Mycobacterium tuberculosis](#)
 Uniprot id: [Q854Q](#)
 PubChem cat: [200](#)



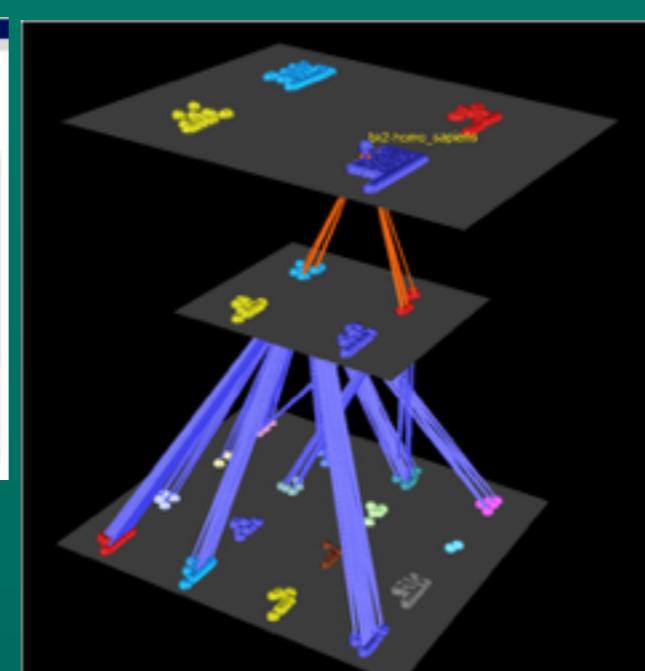
Clustering based on sequence similarity and domain architecture



Protein – protein, protein-chemical interactions



Visualization in Medusa & Arena3D



Clustering based on sequence similarity and domain architecture

Pathway and GO terms enrichment

The figure shows a screenshot of the UniProtKB search results for 'Molecular function'. The results list various biological processes and activities, each associated with a GO ID. A large, detailed diagram of a protein complex, likely a ribosome or similar macromolecular machine, is displayed in the center. The diagram illustrates the spatial arrangement of different proteins and RNA molecules within the complex.

Clinical trials and patent information

The screenshot shows a search result for a patent application. The title is "GENTERICALLY AL TERED ANTIBODY-PRODUCING CELL LINES WITH IMPROVED ANTICANcer ACTIVITY". The application number is PCT/US2003/023592, and the priority date is 2002-03-01. The inventors listed are KUHN, RICHARD J., HARRIS, ROBERT C., and REED, DAVID P. The abstract describes the invention of antibody-producing cell lines that have been genetically modified to produce antibodies with improved anticancer activity. The document includes several figures and tables, such as Figure 1 showing cell lines A2, A2L, and A2R, and Table 1 showing various parameters for these cell lines.

Chemistry information

- DrugBank, hmdb, STITCH, MATADOR, chEBI
 - PubChem, PDBLigandetc



Protein – protein, protein-chemical interactions

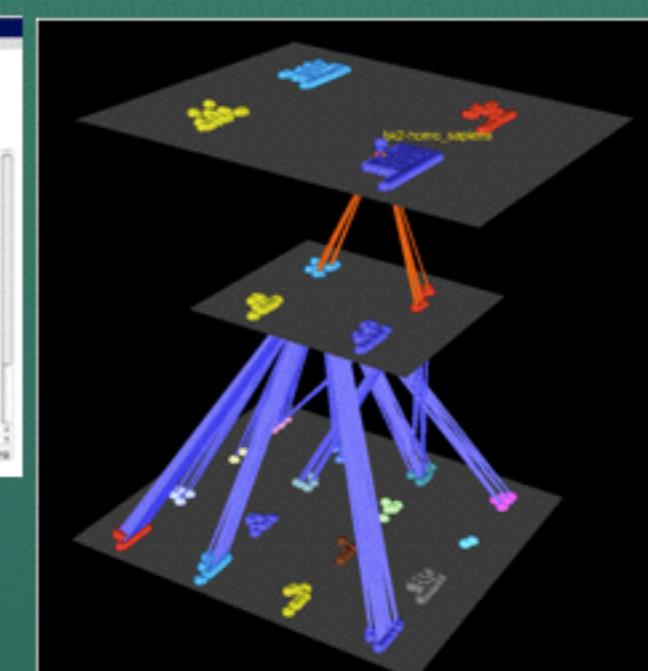
Visualization in Medusa & Arena3D

The diagram illustrates the relationships between various biological processes and diseases. Nodes are color-coded: black for Diseases, green for Genes, yellow for Chemical I Groups, and orange for Ligands. Edges represent interactions, with thickness indicating weight. Labels on edges provide specific interaction details.

Legend:

- Nodes
- Weights
- Wide edges on
- Mutiple edges on
- Labels

Layout: Chordal



Clustering based on sequence similarity and domain architecture

Pathway and GO terms enrichment

Clinical trials and patent information

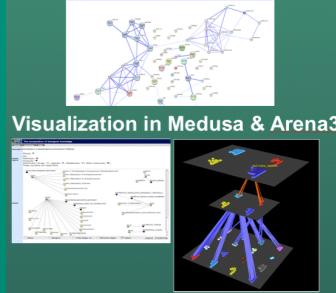
A screenshot of a web-based search interface titled "Clinical trials and patent information". The top navigation bar includes links for "Home", "Help", "Logout", and "Search". Below the title, there's a search bar with placeholder text "Search Clinical Trials and Patent Information". A sidebar on the left lists categories like "Patient", "Healthcare", "Researcher", and "Pharmaceutical". The main content area shows a search result for "Anti-VEGF monoclonal antibody" from "Genentech". It displays the antibody's name, its use as a "THERAPEUTIC MONOCLONAL ANTIBODY", and its target "VEGFR". A large orange button labeled "View details" is prominent. To the right, there's a sidebar with "GENETICALLY ALtered TIRED ANTIBODY PRODUCING CELL LINES WITH IMPROVED ANTIBODY CHARACTERISTICS" and a "View abstract" link. At the bottom, there's a "New search" section and a note about "Data supplied by the assignedee database - Webdatabe".

Chemistry information

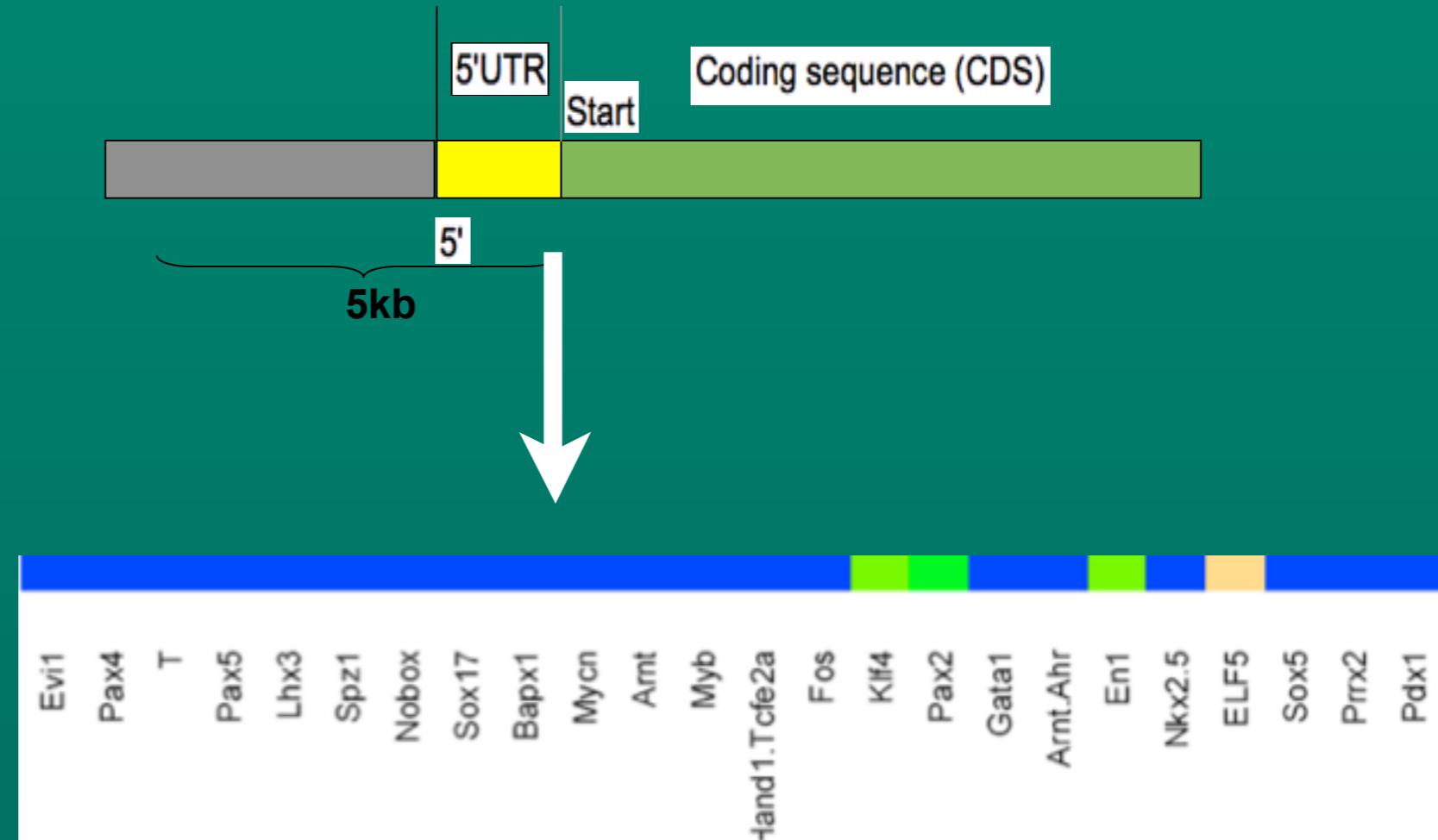
- DrugBank, hmdb, STITCH, MATADOR, chEBI
 - PubChem, PDBLigandetc



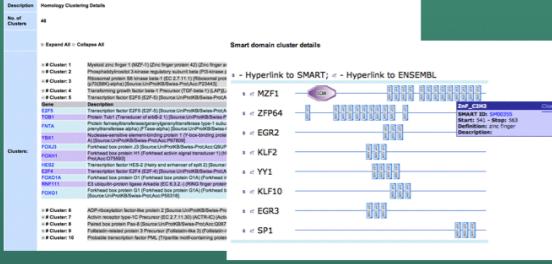
Protein – protein, protein-chemical interactions



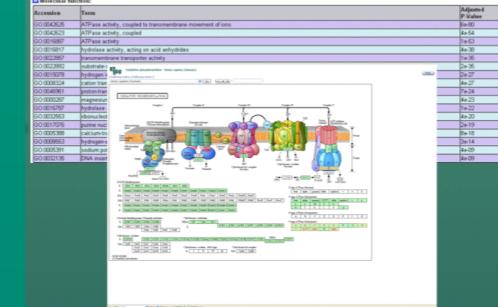
Transcription factor binding site (TFBS) profiling



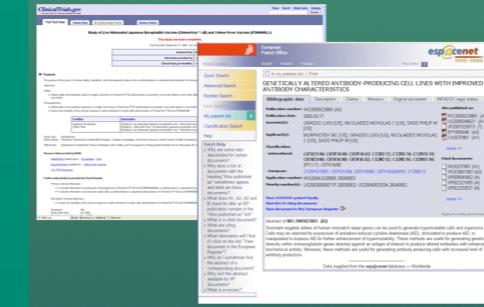
Clustering based on sequence similarity and domain architecture



Pathway and GO terms enrichment



Clinical trials and patent information

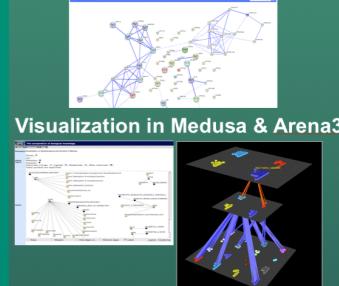


Chemistry information

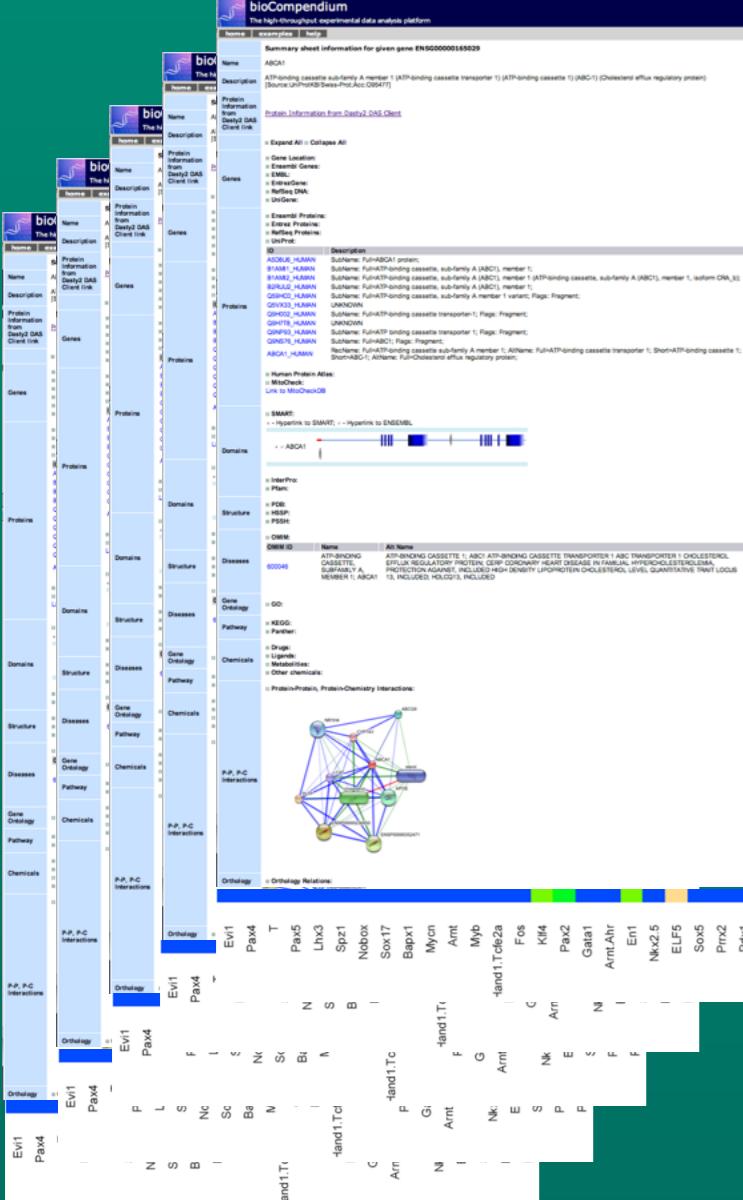
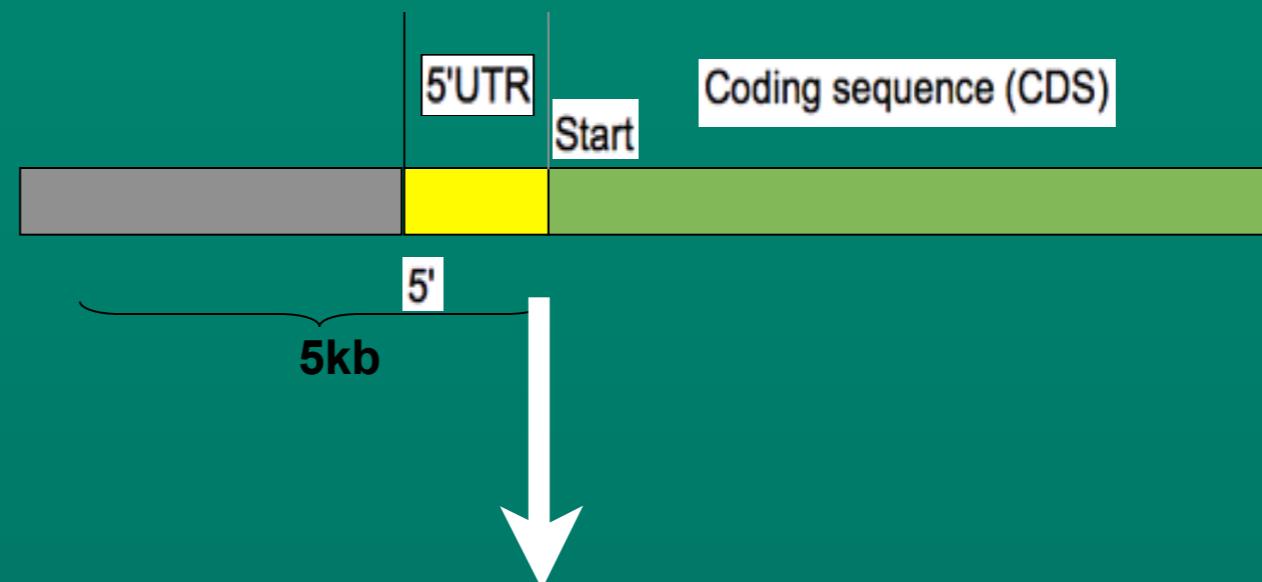
- DrugBank, hmdb, STITCH, MATADOR, chEBI
- PubChem, PDBLigandetc



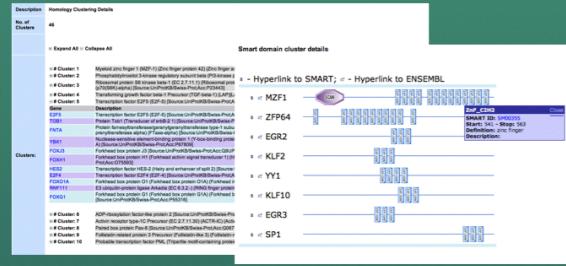
Protein – protein, protein-chemical interactions



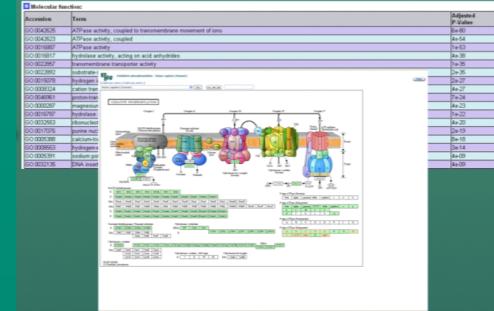
Transcription factor binding site (TFBS) profiling



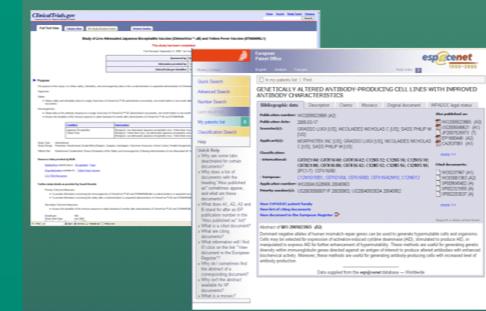
Clustering based on sequence similarity and domain architecture



Pathway and GO terms enrichment



Clinical trials and patent information

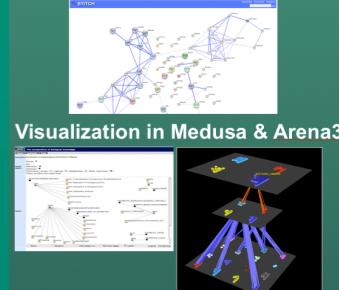


Chemistry information

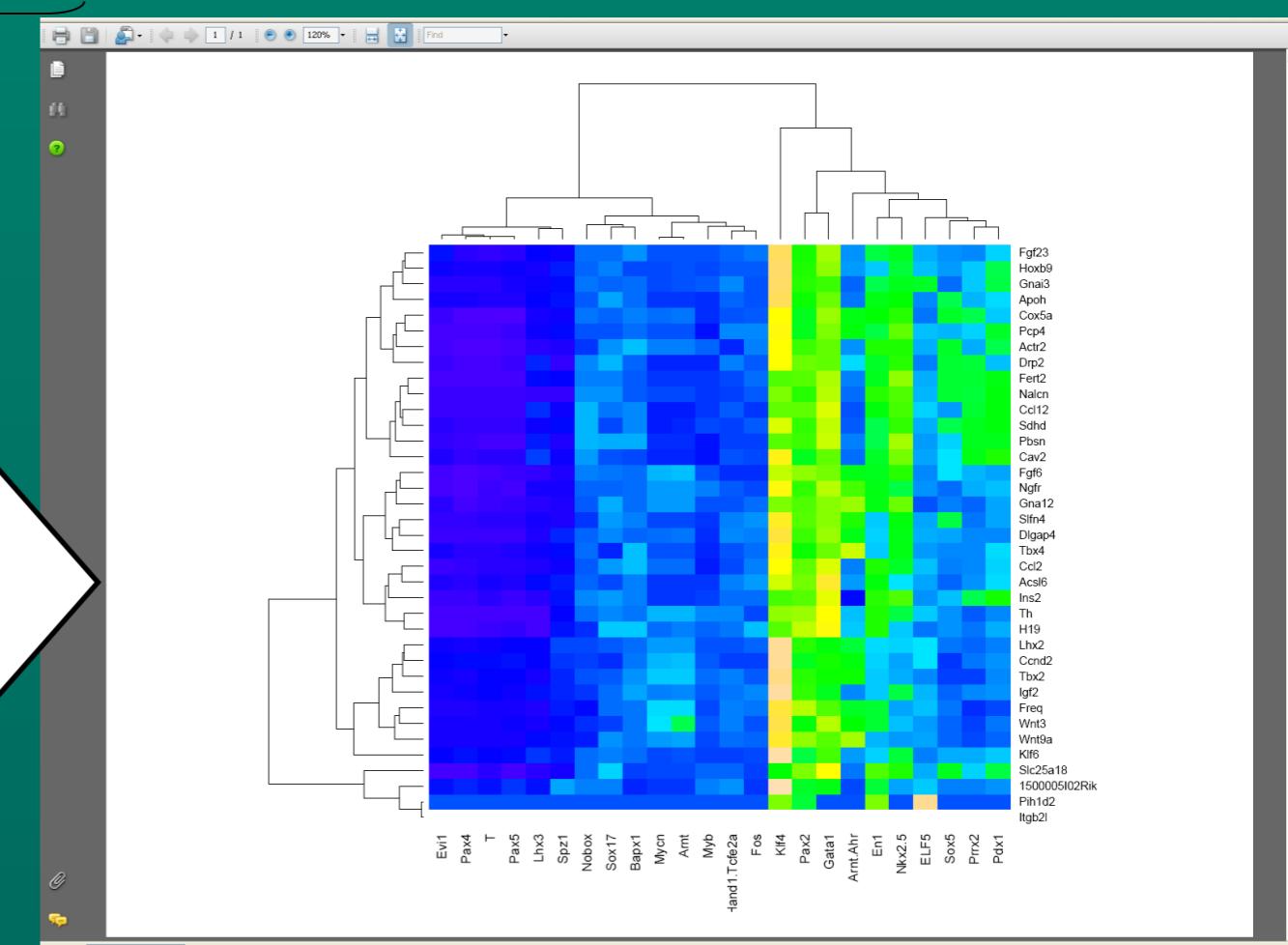
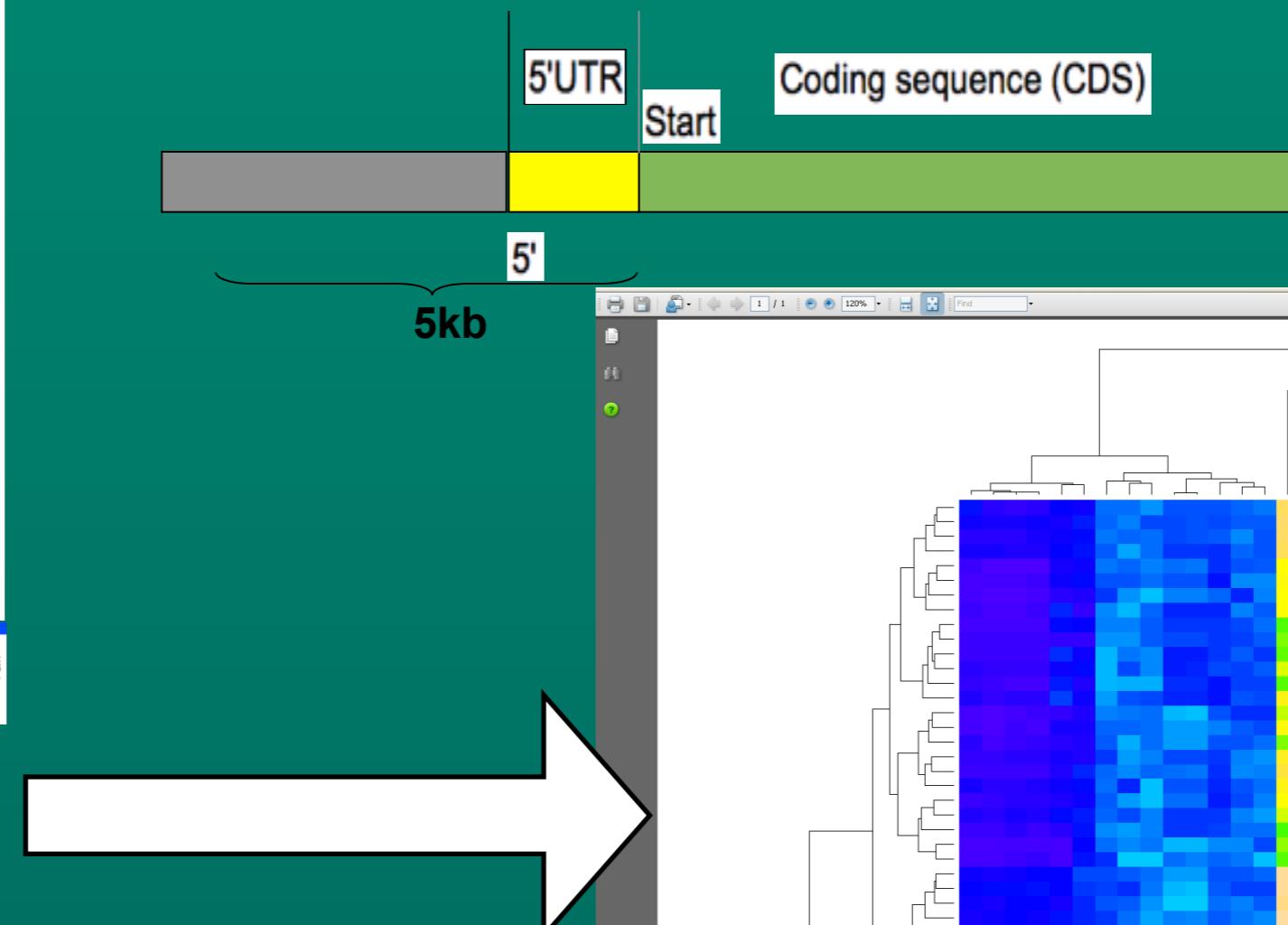
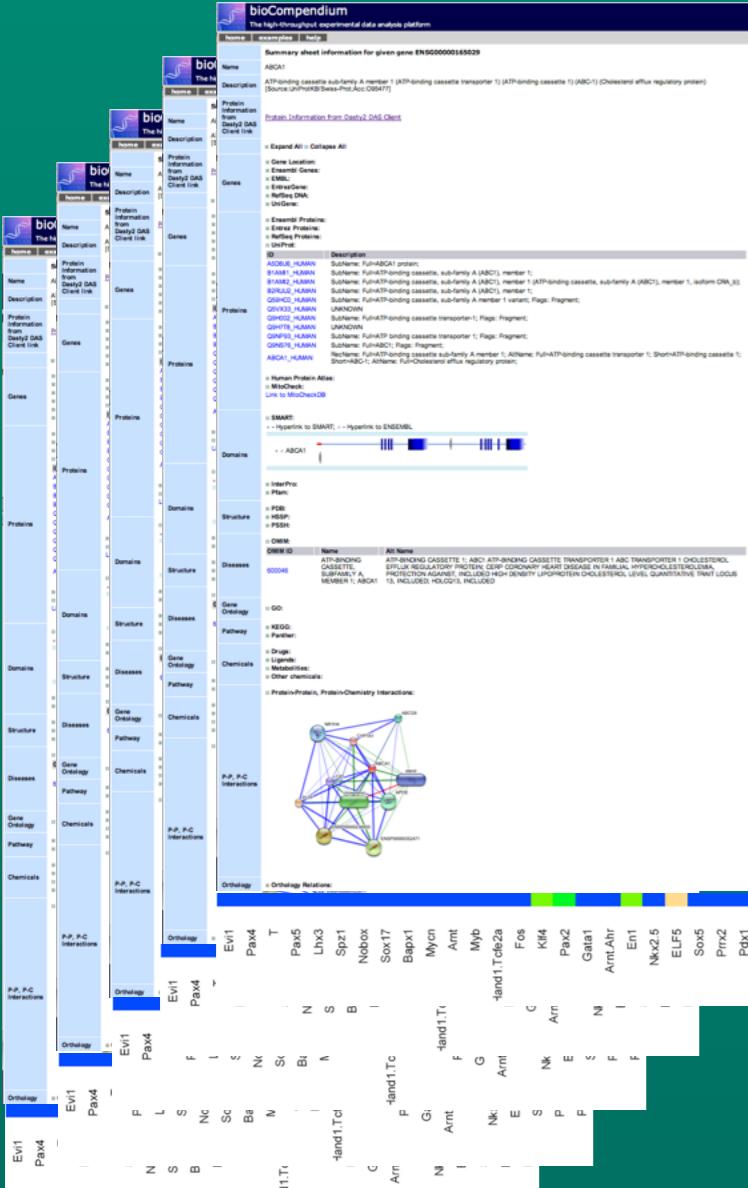
- DrugBank, hmdb, STITCH, MATADOR, chEBI
- PubChem, PDBLigandetc

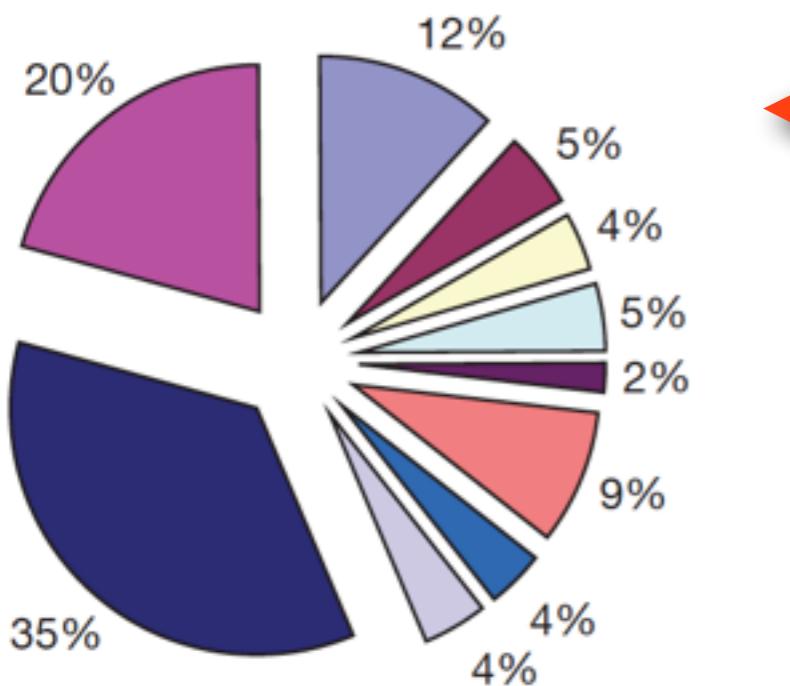


Protein – protein, protein-chemical interactions



Transcription factor binding site (TFBS) profiling



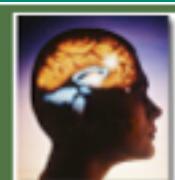


Gene Ontology (GO), Biological process annotations of the 572 validated mitotic hits



- Phenotypic profiling of the human genome by time-lapse microscopy reveals genes with functions in cell division, survival or migration
Ellenberg J. EMBL
(Neumann et.al., Nature, 2010)
- From experimental setup to bioinformatics: an RNAi screening platform to identify host factors involved in HIV-1 replication
Kräusslich H.G., Dep. of Infect. Diseases, Univ. Heidelberg
(Börner et.al., Biotech J., 2010)
- Human-gpDB: A database of human GPCRs, G-proteins, Effectors and their interactions
Hamodrakas S., Univ. Athens, Greece
(Satagopam et.al., Database(Oxford), 2010)
- Defective lamin A-Rb signaling in Hutchinson-Gilford Progeria Syndrome and reversal by farnesyltransferase inhibition
Djabali K., Columbia Univ. NYC
(Marji et.al., PLoS ONE 2010)
- Signaling in insulin-producing cells is altered when cells are grown in islets (3-D) compared to monolayer (2-D).
Bergsten P., Medical Cell Biology, Univ. Uppsala
- Cancer pain related targets
Kuner R., Institute of Pharmacology, Univ. Heidelberg

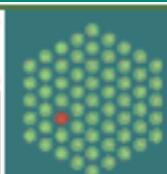
Use of BioCompendium



TAMAHUD
Bioinformatics data analysis



UNIVERSITY OF
CAMBRIDGE
TCPInnovationsLtd
SCIENCE INTO TECHNOLOGY



| SienaBiotech | University of Cambridge | TCP Innovations Limited | Bioalma | European Molecular Biology Laboratory |

Search gene-expression analysis results:

Legend

1. Click on analysis hyper link to see the analysis results
2. Select the check-boxes from different experiments to compare/analyze/enrich with [bioCompendium](#)

Expand All Collapse All

Huntington's Disease:

Select the data sets to be analyzed

[Go to bioCompendium](#)

Reference	Title	Organism	GEO acc	PubMed ID
<input checked="" type="checkbox"/> RF3	TAMAHUD screen	Human		
<input checked="" type="checkbox"/> Hodges , <i>Genes Brain Behav</i> 2007	Huntington's disease transgenic model: time course	Mouse	GDS2912	17696994

Analysis

[18wk R6/1 versus wildtype](#)

[18wk R6/1 versus wildtype](#)

[27wk R6/1 versus wildtype](#)

[27wk R6/1 versus wildtype](#)

P-Value cut off Fold change significant cut off No. of hits

0.05 0.0 104

0.01 0.0 6

0.05 0.0 1010

0.01 0.0 172

Links to GEO & PubMed

<input checked="" type="checkbox"/> Lee, <i>Plos Genet</i> 2007	Huntington's disease models	Mouse	GDS2911	17708681
<input checked="" type="checkbox"/> Apostol, <i>Hum Mol Genet</i> 2006	Huntingtin exon 1 protein overexpression	Rat	GDS1236	16330479
<input checked="" type="checkbox"/> Morton, <i>Eur J Neurosci</i> 2005	Huntington's disease and combination drug therapy	Mouse	GDS717	15787692

Analysis

[R6/2 control versus wild-type control](#)

[R6/2 control versus wild-type control](#)

P-Value cut off Fold change significant cut off No. of hits

0.01 0.0 23

0.05 0.0 250

Use of BioCompendium

The screenshot shows the BLUEGECKO workspace interface. At the top, there is a toolbar with various icons and a tab bar showing several open projects: "Garuda...", "Provisio...", "bioCom...", "bioCom...", "Species ...", and "Work...". The main area is titled "BLUEGECKO" and contains two sections: "Available data" and "Available actions".

Available data: A button labeled "Upload new data".

Available actions: A list of actions grouped under "Data manipulation":

- Data mining (highlighted in blue)
- Machine learning
- Sequence processing
- Visualization

Details for the "Data mining" action:

- Find functionally related genes
- GO terms analysis
- Gene ID conversion
- Get relevant protein interaction graph

Jobs status: Buttons for "Refresh" and "Clear failed jobs".

bioCompendium: A button at the bottom right of the workspace area, which is circled in red.

Exercises part 3

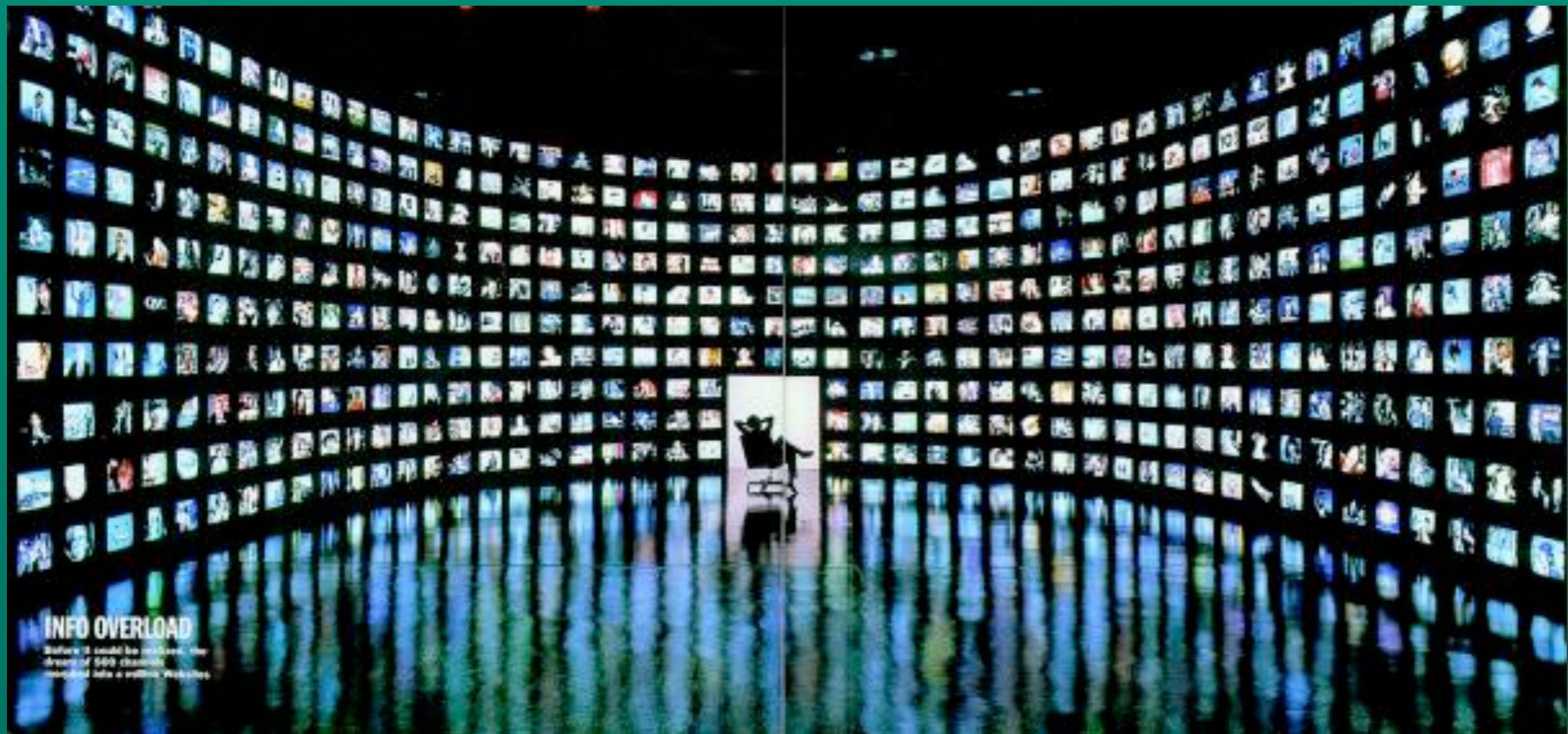
Parse EMBL records

Parsing EMBL records

- Write a perl parser to process given file ‘EMBL_raw.txt’ and produce a tab delimited file with ‘Accession Number’, ‘Organism’ and ‘Description’ information as it is in the file ‘EMBL_parsed.txt’

Reflect demo

Information overload

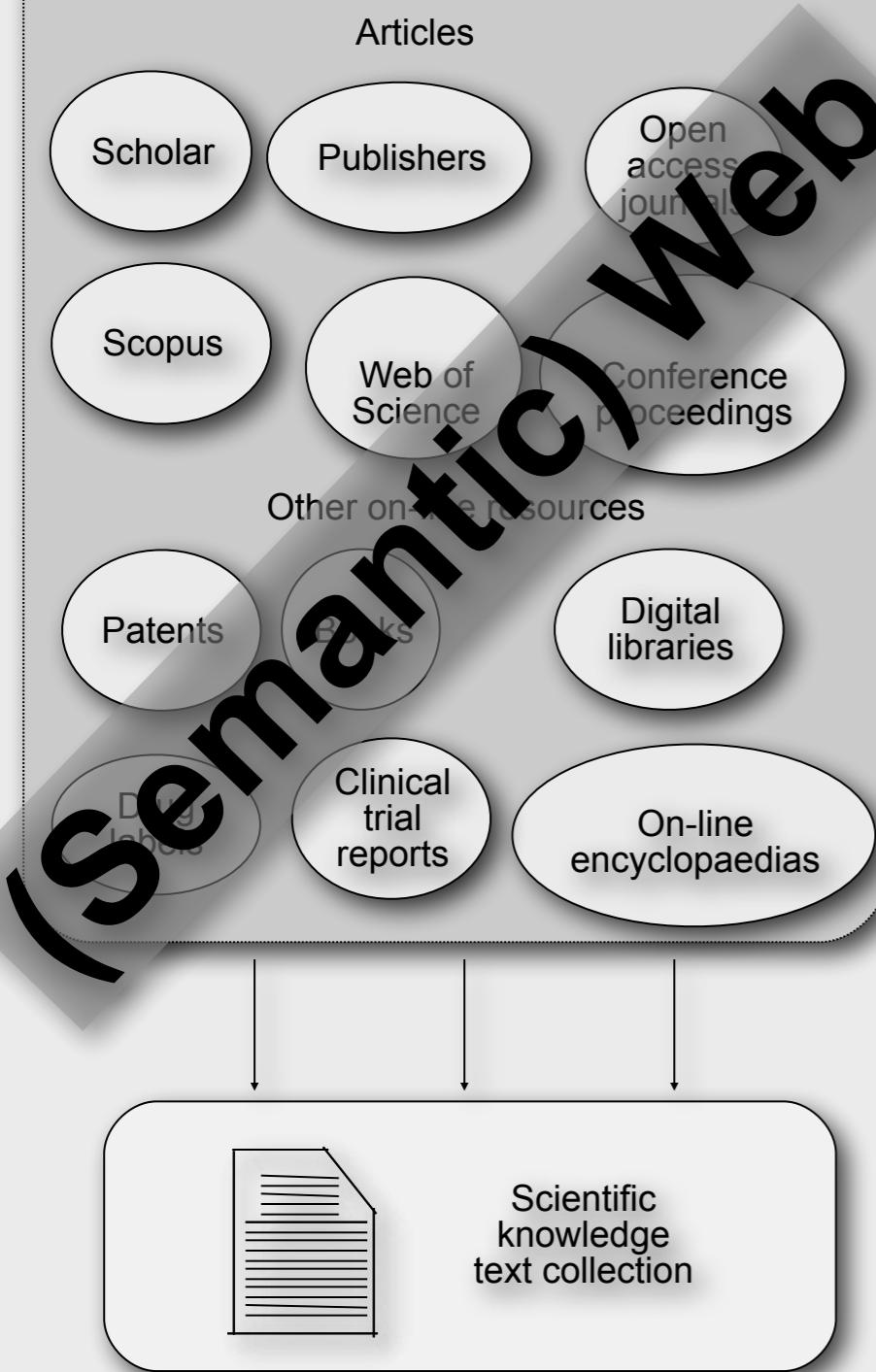


INFO OVERLOAD

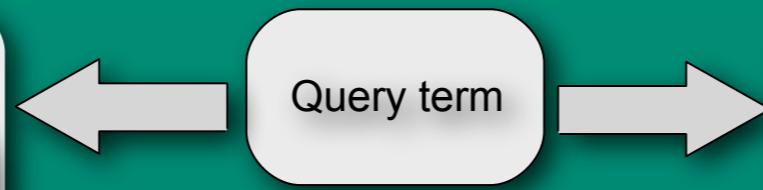
Before it could be reduced, the
stream of 500 channels
contained 100 million messages.

Bridging the silos

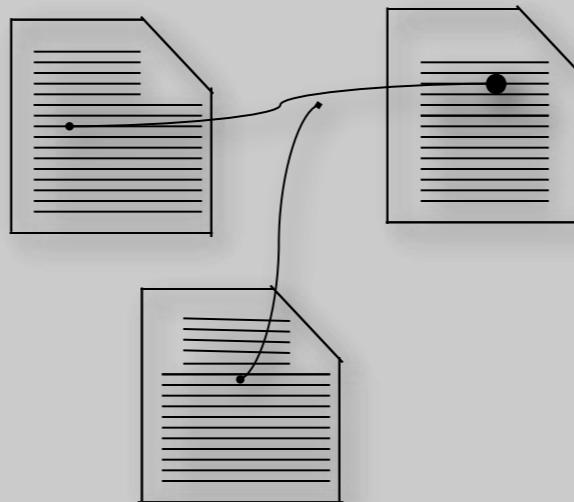
Automated web mining



Query term

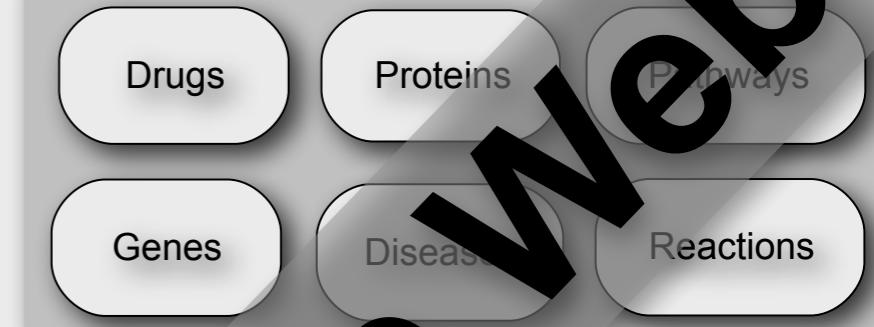


Knowledge discovery -
Relationship prediction



Text
mining
engine

Database queries



Query expansion based on:

Sequence homology

Structural similarity

Synonyms,
symbols,
identifiers,
terminological
annotation
collection

Document
collection

Primary Design Criteria

- Installing and using should be fast and easy
- Interactive (speed)

<http://reflect.ws>



Pafilis, O'Donoghue, Jensen, Horn, Kuhn, Brown, Schneider
Reflect: Augmented Browsing for the Life Scientist
Nature Biotechnology, 27, 508-510, 2009

Molecular portrait of cisplatin i... +

www.ncbi.nlm.nih.gov/pubmed?term=17711579

Most Visited Getting Started Latest Headlines Apple Yahoo! Google Maps YouTube Wikipedia News Popular

NCBI Resources How To My NCBI Sign In

PubMed.gov US National Library of Medicine National Institutes of Health

PubMed 17711579[uid] Search RSS Save search Limits Advanced Help

Display Settings: Abstract Send to: Read free full text at BioMed Central FREE full text article in PubMed Central

Mol Cancer. 2007 Aug 21;6:53.

Molecular portrait of cisplatin induced response in human testis cancer cell lines based on gene expression profiles.

Duale N, Lindeman B, Komada M, Olsen AK, Andreassen A, Soderlund EJ, Brunborg G.
Department of Chemical Toxicology, Division of Environmental Medicine, Norwegian Institute of Public Health, Oslo, Norway. nur.duale@fhi.no

Abstract

BACKGROUND: Testicular germ cell tumors (TGCTs) respond well to cisplatin-based chemotherapy and show a low incidence of acquired resistance compared to most somatic tumors. The reasons for these specific characteristics are not known in detail but seem to be multifactorial. We have studied gene expression profiles of testicular and colon cancer derived cell lines treated with cisplatin. The main goal of this study was to identify novel gene expression profiles with their functional categories and the biochemical pathways that are associated with TGCT cells' response to cisplatin.

RESULTS: Genes that were differentially expressed between the TGCT cell lines vs the (somatic) HCT116 cell line, after cisplatin treatment, were identified using the significance analysis of microarrays (SAM) method. The response of TGCT cells was strikingly different from that of HCT116, and we identified 1794 genes that were differentially expressed. Functional classification of these genes showed that they participate in a variety of different and widely distributed functional categories and biochemical pathways. Database mining showed significant association of genes ($n = 41$) induced by cisplatin in our study, and genes previously reported to be expressed in differentiated TGCT cells. We identified 37 p53-responsive genes that were altered after cisplatin exposure. We also identified 40 target genes for two microRNAs, hsa-mir-372 and 373 that may interfere with p53 signalling in TGCTs. The tumor suppressor genes NEO1 and LATS2, and the estrogen receptor gene ESR1, all have binding sites for p53 and hsa-mir-372/373. NEO1 and LATS2 were down-regulated in TGCT cells following cisplatin exposure, while ESR1 was up-regulated in TGCT cells. Cisplatin-induced genes associated with terminal growth arrest through senescence were identified, indicating associations which were not previously described for TGCT cells.

CONCLUSION: By linking our gene expression data to publicly available databases and literature, we provide a global pattern of cisplatin induced cellular response that is specific for testicular cancer cell lines. We have identified cisplatin-responsive functional classes and pathways, such as the angiogenesis, Wnt, integrin, and cadherin signaling pathways. The identification of differentially expressed genes in this study may contribute to a better understanding of the unusual sensitivity of TGCT to some DNA-damaging agents.

PMID: 17711579 [PubMed - indexed for MEDLINE] PMCID: PMC1988831 Free PMC Article

Images from this publication. See all images (5) Free text

Related citations

- Role of MEK/ERK pathway in the MAD2-mediated cisplatin sen [Br J Cancer. 2006]
- Expression of p53, Bcl-2 and Bax in cisplatin-induced apoptosis in testicula [Br J Cancer. 1998]
- Differentiation of human embryonal carcinomas in vitro and in vivo reveals exq [Cancer Res. 2005]
- Review Testicular germ cell tumours: the paradigm of chemc [Int J Biochem Cell Biol. 2005]
- Review Chromosomes and expression in human testicular germ-cell tum [Ann N Y Acad Sci. 2007]

See reviews... See all...

Cited by 7 PubMed Central articles

- Alterations in gene expression profiles correlated with cisplatin cytotoxicity ir [Genet Mol Biol. 2010]
- Micro-RNA expression in cisplatin resistant germ cell tumor cell lines. [Mol Cancer. 2011]
- Review miRNA in pluripotent stem cells. [Regen Med. 2010]

See all...

Related information

Related Citations

Molecular portrait of cisplatin i...



www.ncbi.nlm.nih.gov/pubmed?term=17711579



Reflect



Google

Most Visited

Getting Started

Latest Headlines

Apple

Yahoo!

Google Maps

YouTube

Wikipedia

News

Popular

NCBI

Resources

How To

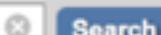
My NCBI Sign In

PubMed.gov

US National Library of Medicine
National Institutes of Health

PubMed

17711579[uid]



Search



RSS Save search

Limits

Advanced

Help

Display Settings: Abstract

Send to:

Read free full text at
BioMed Central**FREE** full text article
in PubMed Central

Mol Cancer. 2007 Aug 21;6:53.

Molecular portrait of cisplatin induced response in human testis cancer cell lines based on gene expression profiles.

Duale N, Lindeman B, Komada M, Olsen AK, Andreassen A, Soderlund EJ, Brunborg G.

Department of Chemical Toxicology, Division of Environmental Medicine, Norwegian Institute of Public Health, Oslo, Norway. nur.duale@fhi.no

Abstract

BACKGROUND: Testicular germ cell tumors (TGCTs) respond well to cisplatin-based chemotherapy and show a low incidence of acquired resistance compared to most somatic tumors. The reasons for these specific characteristics are not known in detail but seem to be multifactorial. We have studied gene expression profiles of testicular and colon cancer derived cell lines treated with cisplatin. The main goal of this study was to identify novel gene expression profiles with their functional categories and the biochemical pathways that are associated with TGCT cells' response to cisplatin.

RESULTS: Genes that were differentially expressed between the TGCT cell lines vs the (somatic) HCT116 cell line, after cisplatin treatment, were identified using the significance analysis of microarrays (SAM) method. The response of TGCT cells was strikingly different from that of HCT116, and we identified 1794 genes that were differentially expressed. Functional classification of these genes showed that they participate in a variety of different and widely distributed functional categories and biochemical pathways. Database mining showed significant association of genes (n = 41) induced by cisplatin in our study, and genes previously reported to be expressed in differentiated TGCT cells. We identified 37 p53-responsive genes that were altered after cisplatin exposure. We also identified 40 target genes for two microRNAs, hsa-mir-372 and 373 that may interfere with p53 signaling in TGCTs. The tumor suppressor genes NEO1 and LATS2, and the estrogen receptor gene ESR1, all have binding sites for p53 and hsa-mir-372/373. NEO1 and LATS2 were down-regulated in TGCT cells following cisplatin exposure, while ESR1 was up-regulated in TGCT cells. Cisplatin-induced genes associated with terminal growth arrest through senescence were identified, indicating associations which were not previously described for TGCT cells.

CONCLUSION: By linking our gene expression data to publicly available databases and literature, we provide a global pattern of cisplatin induced cellular response that is specific for testicular cancer cell lines. We have identified cisplatin-responsive functional classes and pathways, such as the angiogenesis, Wnt, integrin, and cadherin signaling pathways. The identification of differentially expressed genes in this study may contribute to a better understanding of the unusual sensitivity of TGCT to some DNA-damaging agents.

PMID: 17711579 [PubMed - Indexed for MEDLINE] PMCID: PMC1988831 Free PMC Article

See reviews...

See all...

Cited by 7 PubMed Central articles

Alterations in gene expression profiles correlated with cisplatin cytotoxicity in [Genet Mol Biol. 2010]

Micro-RNA expression in cisplatin resistant germ cell tumor cell lines. [Mol Cancer. 2011]

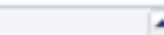
Review miRNA in pluripotent stem cells. [Regen Med. 2010]

See all...

Related Information

Related Citations

Images from this publication. See all images (5) Free text



Molecular portrait of cisplatin induced response in human testis cancer cell lines based on gene expression profiles.

Duale N, Lindeman B, Komada M, Olsen AK, Andreassen A, Soderlund EJ, Brunborg G.

Department of Chemical Toxicology, Division of Environmental Medicine, Norwegian Institute of Public Health, Oslo, Norway. nur.duale@fhi.no

Abstract

BACKGROUND: Testicular germ cell tumors (TGCTs) respond well to cisplatin-based chemotherapy and show a low incidence of acquired resistance compared to most somatic tumors. The reasons for these specific characteristics are not known in detail but seem to be multifactorial. We have studied gene expression profiles of testicular and colon cancer derived cell lines treated with cisplatin. The main goal of this study was to identify novel gene expression profiles with their functional categories and the biochemical pathways that are associated with TGCT cells' response to cisplatin.

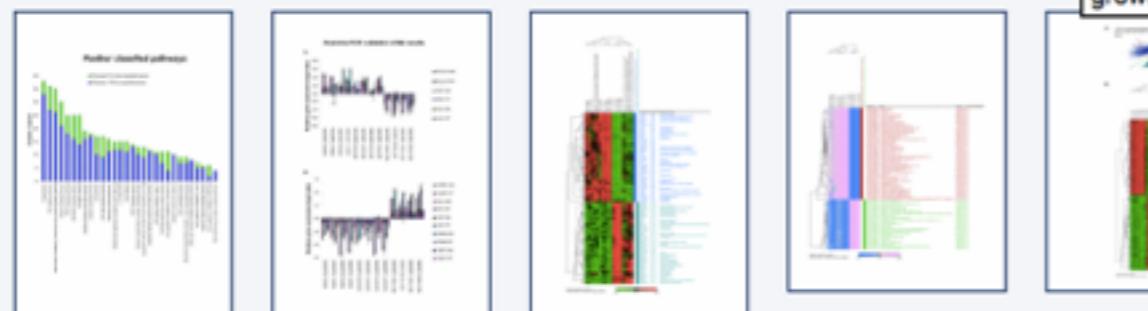
RESULTS: Genes that were differentially expressed between the TGCT cell lines vs the (somatic) HCT116 cell line, after cisplatin treatment, were identified using the significance analysis of microarrays (SAM) method. The response of TGCT cells was strikingly different from that of HCT116, and we identified 1794 genes that were differentially expressed. Functional classification of these genes showed that they participate in a variety of different and widely distributed functional categories and biochemical pathways. Database mining showed significant association of genes ($n = 41$) induced by cisplatin in our study, and genes previously reported to be expressed in different that were altered after cisplatin exposure. We also identified 40 target genes for two microRNAs that regulate p53 signaling in TGCTs. The tumor suppressor genes NEO1 and LATS2, and the estrogen receptor hsa-mir-372/373. NEO1 and LATS2 were down-regulated in TGCT cells following cisplatin exposure. Cisplatin-induced genes associated with terminal growth arrest through senescence were identified and described for TGCT cells.

CONCLUSION: By linking our gene expression data to publicly available databases and literature, we have identified a cellular response that is specific for testicular cancer cell lines. We have identified cisplatin-induced pathways such as angiogenesis, Wnt, integrin, and cadherin signaling pathways. The identification of differentially expressed genes will better understanding of the unusual sensitivity of TGCT to some DNA-damaging agents.

PMID: 17711579 [PubMed - Indexed for MEDLINE] PMCID: PMC1988831

[Free PMC Article](#)

Images from this publication. [See all images \(5\)](#) [Free text](#)



Reflect - p53

Protein Wikipedia Add About

TP53 (ENSP00000269305) ▾ H. sapiens Edit

p53; p53 tumor suppressor; tumor suppressor p53; Trp53; Mutant p53; TP53; P04637, Sequence, Domains, Structure, Locus, Literature

MEEPQSDPSVEPPLSQETFSDLWKLLPENVLSPILPSQAMDDLMI

[View sequence](#)

[View Domains](#)

[View Structure](#)

[View Locus](#)

[View Literature](#)

This protein acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on physiological circumstances

7 PubMed Central articles

... in gene expression profiles correlated with cisplatin cytotoxicity in [Genet Mol Biol. 2010]

... expression in cisplatin resistant germ cell lines. [Mol Cancer. 2011]

mRNA in pluripotent stem cells. [Regen Med. 2010]

[See all...](#)

Information

Citations

Compound (MeSH Keyword)

GEO DataSets

References for this PMC Article

Substance (MeSH Keyword)

Free in PMC

Cited in PMC

Matched terms

Sequence identifier(s)

Synonyms

Domains in SMART

Draggable Sequence window

Protein description

Representative 3D structure In PDBsum

Genomic location In ENSEMBL

Major interaction Partners in STITCH

Related Medline Abstracts in iHOP

Green indicates Known subcellular location

Organism(s)

Edit synonyms or description

Domains – mouse over to show name

Sequence Scroll bars

Organism

TP53

Protein

ENSP00000269305 P53_HUMAN H. sapiens [edit]

Antigen NY CO 13; Phosphoprotein p53; Tumor suppressor p53; A2522

Domains, Sequence, Structure, Locus, Literature

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPPLPSQAMDDLMQLS

3D structure

Network graph

Subcellular localization

Organism image

Text: Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and

This figure is a screenshot of a protein information page for TP53. The top navigation bar includes 'Close (Esc)', 'Help', and '[edit]'. The main content area shows the protein identifier ENSP00000269305 and its genomic location P53_HUMAN in H. sapiens. Below this is a list of synonyms: Antigen NY CO 13; Phosphoprotein p53; Tumor suppressor p53; A2522. A 'Domains, Sequence, Structure, Locus, Literature' link is present. The sequence MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPPLPSQAMDDLMQLS is shown with a green highlight. A red box highlights the first few amino acids. A blue box highlights the last few. A grey box covers the middle. A 'Draggable Sequence window' is indicated by arrows. Below the sequence are four icons: a 3D structure, a network graph, a subcellular localization icon (green), and an organism image. A large text box at the bottom contains a description of TP53's function. Labels on the left side point to specific features: 'Matched terms' points to the protein identifier; 'Sequence identifier(s)' points to the ENSP ID; 'Synonyms' points to the synonym list; 'Domains in SMART' points to the domain section; 'Draggable Sequence window' points to the sequence bar; 'Protein description' points to the text below the 3D structure; 'Representative 3D structure In PDBsum' points to the 3D structure icon; 'Genomic location In ENSEMBL' points to the genomic location section; 'Major interaction Partners in STITCH' points to the network graph; 'Related Medline Abstracts in iHOP' points to the literature section; 'Organism(s)' points to the organism image; 'Edit synonyms or description' points to the edit link; 'Domains – mouse over to show name' points to the domain icons; 'Sequence Scroll bars' points to the sequence scroll bars; and 'Organism' points to the organism image.

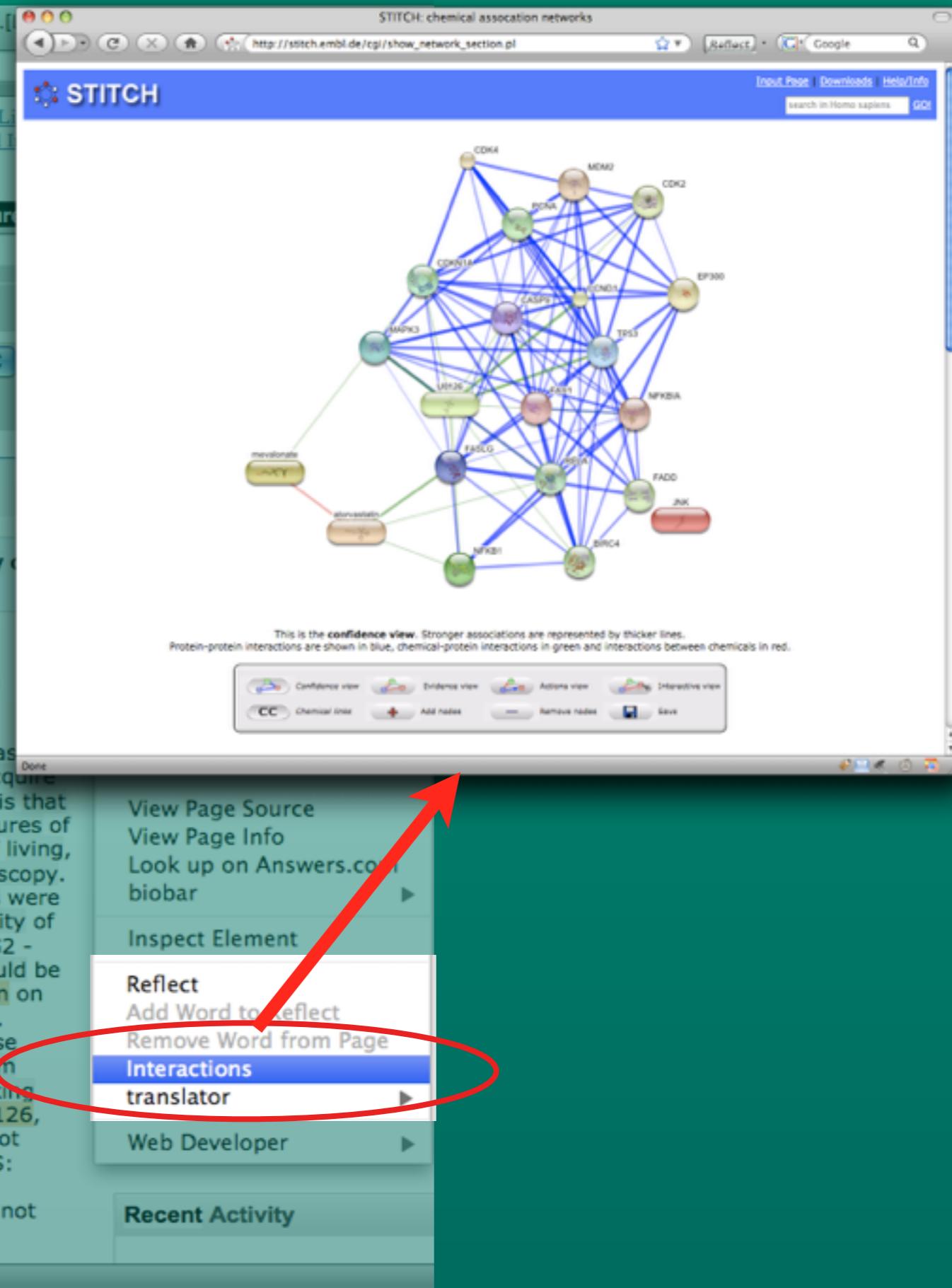
Choose interactions

Atorvastatin induces apoptosis by a caspase-9-dependent pathway: an in vitro study on activated rat hepatic stellate cells.

Aprigliano I, Dudas J, Ramadori G, Saile B.

Department of Internal Medicine, Section of Gastroenterology and Endocrinology, University of Göttingen, Göttingen, Germany.

BACKGROUND: Statins are shown to have cholesterol-independent properties such as anti-inflammation and immunomodulation. Activated hepatic stellate cells (HSCs) acquire the capacity to synthesize matrix proteins in damaged liver. We tested the hypothesis that atorvastatin may be capable of inducing apoptosis in HSCs. **METHODS:** Primary cultures of rat HSCs were exposed to atorvastatin, mevalonic acid and U0126. Quantification of living, apoptotic and necrotic HSCs was performed by flow cytometry and laser-scan microscopy. Cell-cycle analysis was performed by flow cytometry. Pro- and anti-apoptotic factors were investigated by Western blot and electrophoresis mobility shift assay. Protease activity of caspases was calculated using a colorimetric kit. **RESULTS:** Atorvastatin leads to a G2 - arrest and induces apoptosis in activated HSCs. Atorvastatin-mediated apoptosis could be blocked by co-administration of mevalonic acid and U0126. No effects of atorvastatin on gene expression of CD95, CD95L, NF-kappaB, p53 and p21WAF1 could be observed. Atorvastatin - induced apoptosis in activated HSCs is related to an increased protease activity of caspase-9 and -3. Gene expression of the major proteins of the bcl-system shows that truncated Bid is involved in apoptosis mediated by atorvastatin. By blocking the extracellular signal-regulated protein kinase (ERK1 / 2) activation by adding U0126, we could prevent the apoptosis induced by atorvastatin. By Western blot we could not detect any change in the activation of c-jun N-terminal kinase (JNK). **CONCLUSIONS:** Atorvastatin induces apoptosis in activated HSCs acting through an ERK-dependent cleavage of Bid and a highly increased protease activity of caspase-9 and -3. JNK is not involved in atorvastatin-mediated apoptosis in HSCs.



OnTheFly (<http://onthefly.embl.de>)

File

Nr.		Size	Type
1.	bcr753-S2.txt	53 Kb	
2.	EVIDENCE FOR SHOWING GENE-PROTEIN NAME SU...	710 Kb	
3.	PIPE-Dehner.pdf	984 Kb	
4.	SteinTFinal.xls	267 Kb	

Reflect Network Summary

Saccharomyces cerevisiae

summary

Files: 4 Files selected: 1 Total size: 1.97 Mb

Organism: *Saccharomyces cerevisiae*

Drag n' Drop

NCBI PubMed

All Databases PubMed Nucleotide

Search PubMed for [PubMed]

Limits Preview/Index History Clipboard

Display AbstractPlus Show 20

All: 1 Review: 0

1: [Cell. 2008 Jun 13;133\(6\):1006-18.](#)

Comment in: [Cell. 2008 Jun 13;133\(6\):958-61.](#)

Chemokine signaling via the CXCR2 receptor reinforces senescence.

Acosta JC, O'Loghlen A, Banito A, Guijarro MV, Augert A, Raguz S, Fumagalli M, Da Costa M, Brown C, Popov N, Takatsu Y, Melamed J, d'Adda di Fagagna F, Bernard D, Hernando E, Gil J.

Cell Proliferation Group, MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College, Hammersmith Campus, W12 0NN London, UK.

Cells enter senescence, a state of stable proliferation arrest, in response to both cellular stresses, including telomere erosion, DNA damage, and oncogenic signaling, which acts as a barrier against malignant transformation *in vivo*. To identify genes controlling senescence, we conducted an unbiased screen for small hairpin RNAs that extend the life span of primary human fibroblasts. Here, we report that knocking down the chemokine receptor CXCR2 (IL8RB) alleviates both *telomerase*-induced senescence (OIS) and diminishes the DNA-damage response. Conversely, ectopic expression of CXCR2 results in premature senescence via a p53-dependent mechanism. Cells undergoing OIS secrete multiple CXCR2-binding chemokines, which are regulated by the NF-kappaB and C/EBPbeta transcription factors and coordinately induce CXCR2 expression. CXCR2 upregulation is also observed in preneoplastic lesions *in vivo*. These results suggest that senescent cells activate a self-reinforcing network in which CXCR2-binding chemokines reinforce growth arrest.

PMID: 18555777 [PubMed - indexed for MEDLINE]

Reflect Add 'Word' to Reflect Remove 'JUN' from Page Interactions translator Web Developer

NCBI PubMed

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journal

Search PubMed for [PubMed]

Limits Preview/Index History Clipboard Details

All: 1 Review: 0

1: [Cell. 2008 Jun 13;133\(6\):1006-18.](#)

Comment in: [Cell. 2008 Jun 13;133\(6\):958-61.](#)

Chemokine signaling via the CXCR2 receptor reinforces senescence.

Acosta JC, O'Loghlen A, Banito A, Guijarro MV, Augert A, Raguz S, Fumagalli M, Da Costa M, Brown C, Popov N, Takatsu Y, Melamed J, d'Adda di Fagagna F, Bernard D, Hernando E, Gil J.

Cell Proliferation Group, MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College, Hammersmith Campus, W12 0NN London, UK.

Cells enter senescence, a state of stable proliferation arrest, in response to both cellular stresses, including telomere erosion, DNA damage, and oncogenic signaling, which acts as a barrier against malignant transformation *in vivo*. To identify genes controlling senescence, we conducted an unbiased screen for small hairpin RNAs that extend the life span of primary human fibroblasts. Here, we report that knocking down the chemokine receptor CXCR2 (IL8RB) alleviates both *telomerase*-induced senescence (OIS) and diminishes the DNA-damage response. Conversely, ectopic expression of CXCR2 results in premature senescence via a p53-dependent mechanism. Cells undergoing OIS secrete multiple CXCR2-binding chemokines, which are regulated by the NF-kappaB and C/EBPbeta transcription factors and coordinately induce CXCR2 expression. CXCR2 upregulation is also observed in preneoplastic lesions *in vivo*. These results suggest that senescent cells activate a self-reinforcing network in which CXCR2-binding chemokines reinforce growth arrest.

PMID: 18555777 [PubMed - indexed for MEDLINE]

Reflect Add 'NF-kappaB' to Reflect Remove 'NF-kappaB' from Page Interactions

Translate "NF-kappaB" Explain "NF-kappaB" translator Web Developer

NF-kappaB

Protein Wikipedia Add Synonym Help

Add "NF-kappaB" as a:

Type: ▼ Protein Synonym

Identifier or Name:

Organism:

Matches: IKBKG (ENSG00000073009)

Add to Reflect



The Elsevier Grand Challenge

Knowledge Enhancement in the Life Sciences

Home
About the Challenge
>
Finalist Round >
Semifinalist Round
News >



CRN

www.channelweb.co.uk

[Click here to print](#)

EMBL and DERI triumph in Elsevier Grand Challenge

Launched by the global healthcare and scientific publisher in June 2008, the competition invited researchers to prototype tools dealing with the ever-increasing amount of online life sciences information

IWR News Desk, [Information World Review](#) 24 Apr 2009

The Winners Now Announced!



First prize winner (\$35,000)

Sean I. O'Donoghue, Lars J. Jensen, Heiko Horn, Evangelos Pafilis, Michael Kuhn, Nigel P. Brown and Reinhard Schneider, EMBL Germany, for their project Reflect: Automated Annotation of Scientific Terms

[>> Play presentation \(webcast\)](#)



ELSEVIER

<http://www.elsevier.com>

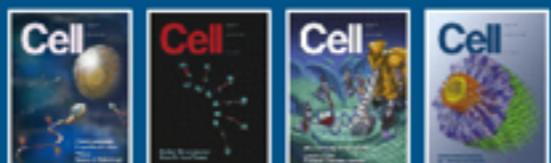
Survey Question:
“Should Cell Press offer Reflect annotations on all of its research articles?”

Feedback: 80% voted “Yes”

Elsevier Pilots New Research Tool ‘Reflect’ In Its Premier Life Science Journal *Cell*

‘Reflect’ links scientific terms to information rich content

Amsterdam, 13 November 2009 – Elsevier, the leading publisher of scientific, technical and medical information products and services, today announced that the innovative research tool ‘Reflect’, winner of Elsevier’s Grand Challenge 2009, will be piloted on the research articles in the November 12th issue of *Cell*. The ‘Reflect’ tool identifies the proteins, genes and small molecules mentioned in the *Cell* articles, and generates pop-up windows containing relevant contextual information, with additional links, about those entities.



FuGENE® HD Transfection Reagent
Now available from Promega [Learn More](#)

Article Outline

Summary
Introduction
Results and Discussion
Experimental Procedures
Acknowledgments
Accession Numbers
Supplemental Data
References
Publication Information

Copyright © 2009 Elsevier Inc. All rights reserved.
Cell, Volume 139, Issue 6, 1084-1095, 11 December 2009

doi:10.1016/j.cell.2009.11.015

[Previous Article](#) [Table of Contents](#) [Next Article](#)

Article

The Structural Basis for mRNA Recognition and Cleavage by the Ribosome-Dependent Endonuclease RelE

Cajetan Neubauer^{1,4}, Yong-Gui Gao^{1,4}, Kasper R. Andersen^{2,4}, Christine M. Dunham^{1,5}, Ann C. Kelley¹, Jendrik Hentschel¹, Kenn Gerdes³, V. Ramakrishnan¹, and Ditlev E. Brodersen²,

¹ MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK

² Department of Molecular Biology, Aarhus University, DK-8000 Aarhus C, Denmark

³ Institute for Cell and Molecular Biosciences, The Medical School, University of Newcastle, Newcastle NE2 4HH, UK

Corresponding author

Corresponding author

⁴ These authors contributed equally to the work

⁵ Present address: Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

Summary

Translational control is widely used to adjust gene expression levels. During the stringent response in bacteria, mRNA is degraded on the ribosome by the ribosome-dependent endonuclease, RelE. The molecular basis for recognition of the ribosome and mRNA by RelE and the mechanism of cleavage are unknown. Here, we present crystal structures of *E. coli* RelE in isolation (2.5 Å) and bound to programmed *Thermus thermophilus* 70S ribosomes before (3.3 Å) and after (3.6 Å) cleavage. RelE occupies the A site and causes cleavage of mRNA after the second nucleotide of the codon by reorienting and activating the mRNA for 2'-OH-induced hydrolysis. Stacking of A site codon bases with conserved residues in RelE and 16S rRNA explains the requirement for the ribosome in catalysis and the subtle sequence specificity of the reaction. These structures provide detailed insight into the translational regulation on the bacterial ribosome by mRNA cleavage.

INTRODUCTION

Rapid adaptation to environmental stress is vital for free-living bacteria. During deprivation of nutrients, uncharged transfer RNAs (tRNAs) bind to the stalled ribosomes and catalyzes synthesis of the signal nucleotide, (P)ppGpp. This alarmone regulates the stringent response, a far-reaching adaptation (Potrykus and Cashel, 2008). The stringent response also leads to activation of RelE, an effective inhibitor of protein synthesis. Under normal physiological conditions, RelB, and is inactive (Christensen and Gerdes, 2003; Christensen et al., 2001; Galvani et al., 2001; Gotfredsen and Gerdes, 1998; Li et al., 2009; Overgaard et al., 2009). Degradation of RelB by Lon protease, RelE is able to bind the ribosome and specifically cleave messenger RNA (mRNA) in the A site (Christensen and Gerdes, 2003; Pedersen et al., 2003).

Such toxin-antitoxin pairs are very common in bacteria, and the RelE superfamily also encompasses HigB, YoeB, YafQ, and YhaV, associated with a variety of functions (Christensen and Gerdes, 2003; Grady and Hayes, 2003; Prysak et al., 2009; Schmidt et al., 2007). The crystal structure of YoeB showed that its fold and catalytic mechanism are similar to RelE (Li et al., 2009; Takagi et al., 2005) but lacks the conserved catalytic histidine and glutamic acid. RelE has an intrinsic nuclease activity of the ribosome (Garza-Sanchez et al., 2008; Hayes and Sauer, 2003; Kamada and Hanaoka, 2005; Li et al., 2009; Suncharoen et al., 2009). Pausing ribosomes are recovered for rescue by tmRNA in the absence of stringent response factors (Hayes and Sauer, 2003; Kamada and Hanaoka, 2005). RelE-induced cleavage is most likely a result of the combined action of RelE-like endonucleases and exonucleases like RNase II (Garza-Sanchez et al., 2009). RelE-induced cleavage occurs at the second nucleotide of the codon, although it is occasionally also seen after the third nucleotide and, upon peptide release, even in the E site (Pedersen et al., 2003). The mRNA cleavage efficiency, with the UAG and UGA stop codons and sense codons like UCG and CAG, among the most efficiently cleaved (Pedersen et al., 2003). Together,

RelA

Protein Chemical Help
relA (b2784) *E. coli* Edit
(P)ppGpp synthetase; ppGpp synthetase I; ATP:GTP 3'-pyrophosphotransferase
Literature Sequence Structure Locus Domains
MVAVRSAHINKAGEFDPEKNIASLGITSQKSCECLAETHAYCLOOTQ(4)
No information available
(P)ppGpp synthetase /GTP pyrophosphokinase; In eubacteria ppGpp (guanosine 3'-diphosphate 5'-diphosphate) is a mediator of the

Reflect

Reflect highlights proteins and small molecules and provides links to information-rich summaries.

Cell is using Reflect on a trial basis.
[Tell us what you think.](#)

[Click here to turn off Reflect.](#)

Add/View Comments (0)

Article Information

[PDF \(1796 kb\)](#)

[Full Text with Large Figures](#)

[Supplemental Data](#)

[Export Citation](#)

[Request permission](#)

PubMed

[Articles by Cajetan Neubauer](#)

[Articles by Ditlev E. Brodersen](#)

Related Articles

[RAG Proteins Shepherd](#)

[Double-Strand Breaks to a Specific Plat...](#)

[Joining-Deficient RAG1](#)

[Mutants Block V\(D\)J Recombination In...](#)

[Separation-of-Function Mutants Reveal Critical Roles for RAG...](#)

[...more](#)

Get Bookmark

[Add Bookmark](#)

Related Resources from Featured Advertisers

View datasheets, protocols and more for chromatin & nuclear signaling products

[Abcam](#)

QuikChange® site directed and random mutagenesis kits

[Agilent \(Stratagene Products\)](#)

[More»](#)



Jackson Lab Finds Text-Mining Software Can Speed Some Database Curation Tasks

January 08, 2010

Newsletter: [BioInform](#)
[BioInform - January 8, 2010](#)

By Vivien Marx

Researchers at the Jackson Laboratory have found that integrating text-mining tools into the Mouse Genome Informatics biocuration workflow has improved the productivity of staffers by around 50 percent for certain steps in the curation workflow.

The team found that, when combined, OntheFly and Reflect "offer functionality we are able to use in-house," Dowell said, adding that the MGI team has begun using them along with ProMiner.....

"Right now we are seeing about a 50 percent increase in the productivity of our staff indexers,"

One of the major difficulties in biomedical literature-based curation is the unambiguous identification of biological and chemical entities within text. Reflect and OnTheFly both offer a significant degree of assistance to curators in this area, providing candidate database identifiers and useful supplementary information on such entities which speeds their identification by curators and the creation of a curated entry. Reflect is also useful as an authoring tool. Curators can examine tagged versions of their own UniProt entries and quickly identify inconsistencies in nomenclature which when corrected would both significantly aid comprehension by human users and facilitate automatic tagging of the text entities by Reflect