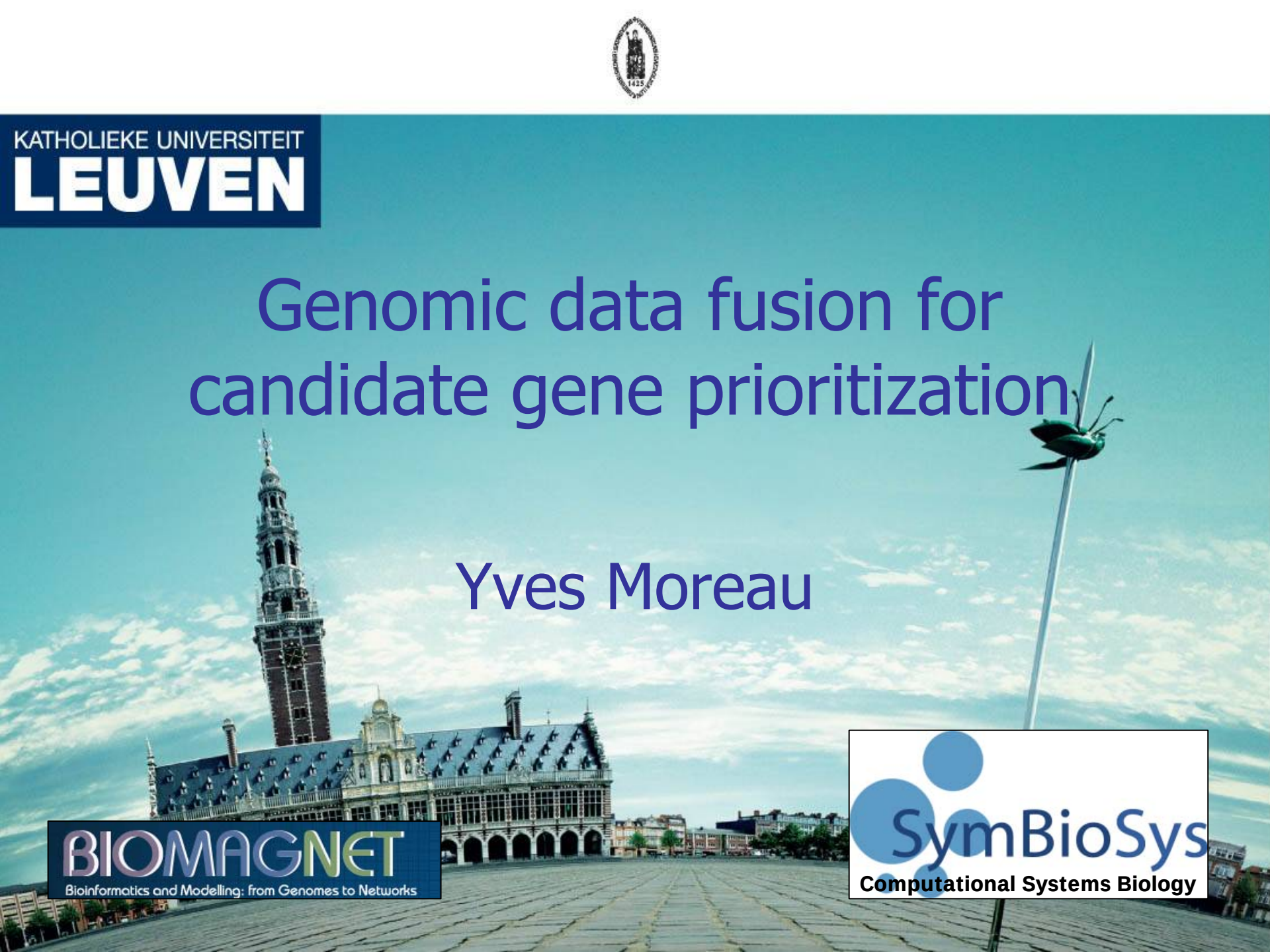
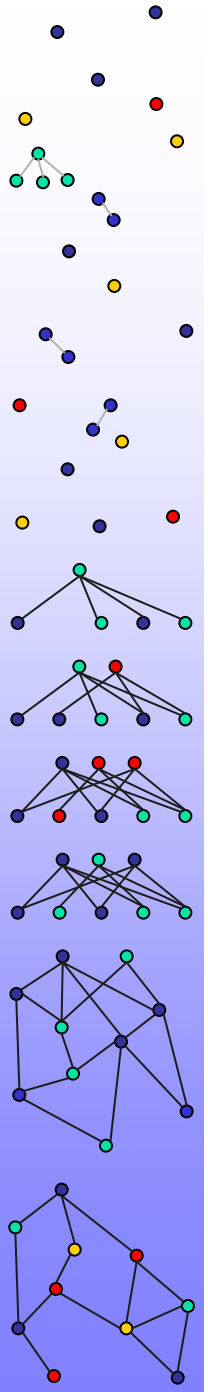


Genomic data fusion for candidate gene prioritization

Yves Moreau

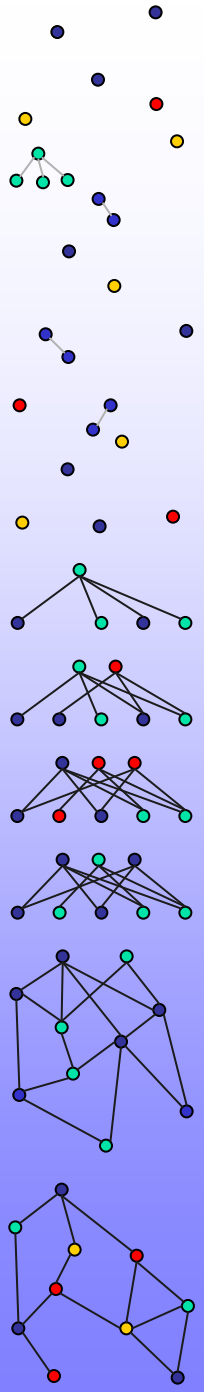
BIOMAGNET
Bioinformatics and Modelling: from Genomes to Networks

SymBioSys
Computational Systems Biology



Why I am the wrong man for the job

- Bioinformatician, not statistical geneticist
- Work on rare constitutional disorders using cytogenetic strategies
- Using ‘omics’ functional data beyond genetic variation
- and highly biased candidate gene strategies
- … but still hunting for genes



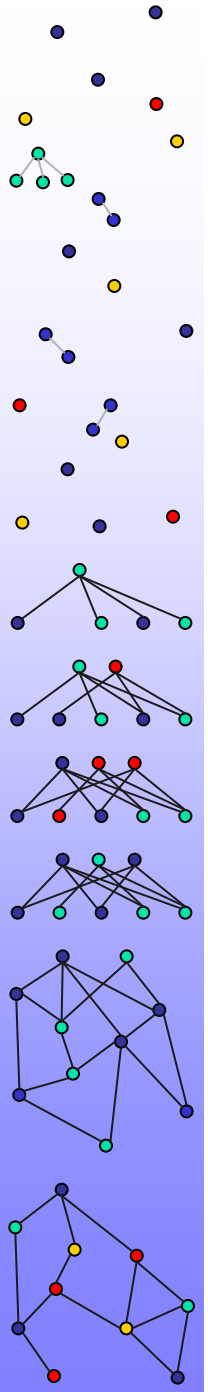
Why this talk is relevant

- Functional data is a rich source of information that can steer genetic studies
  - How to integrate functional data into genetic analysis strategies is still much of an open problem
- Results from genetic studies do not stand on their own but must eventually be integrated in the context of functional pathways
- The proposed strategies could be helpful for
  - Fishing genes in the "grey zone"
  - Selecting a subset of gene pairs for analysis of epistasis

Beyond the hairball

- Networks have become a central concept in biology
- Initial top-down analyses of omics data resulted in hairball description of gene or protein networks
  - High-level properties
    - Scale-free network
  - But what do we do with this?
- *Which methods are available to get actual biological predictions from these multiple sources of data?*

Yeast protein-protein interaction network
Jeong H. et al. Nature. 2001

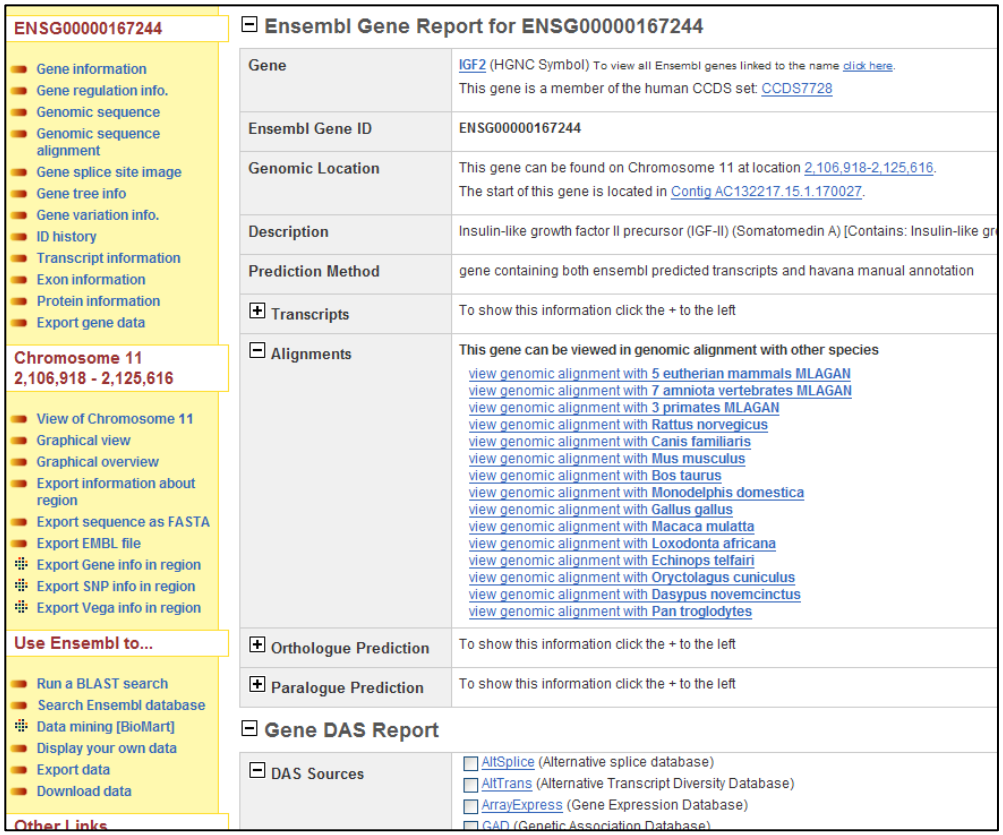


Omics data

- Many other sources of omics information and data are available to help us identify the most interesting candidates for further study
- ChIP chip
- Regulatory motifs
- Protein motifs
- Microarray compendia (Oncomine, ArrayExpress, GEO)
- Protein-protein interaction
- Gene Ontology
- KEGG

Genome browsers

- UCSC genome browser — genome.ucsc.edu
- Ensembl — www.ensembl.org
- Federate many other information sources



Gene Ontology

- Gene Ontology www.geneontology.org

imprinting

Accession: GO:0006349
Ontology: biological_process
Synonyms: exact: DNA imprinting
Definition:
Heritable alterations in the activity of a gene that depend on whether it passed through the paternal or the maternal germline, but that are not encoded by DNA itself.
Definition Source:
GOC:ems
ISBN:0198506732
PMID:11498578
Comment: None

Term Context:
◉Term Ancestors ○Term Siblings
Submit

Term Lineage

Graphical

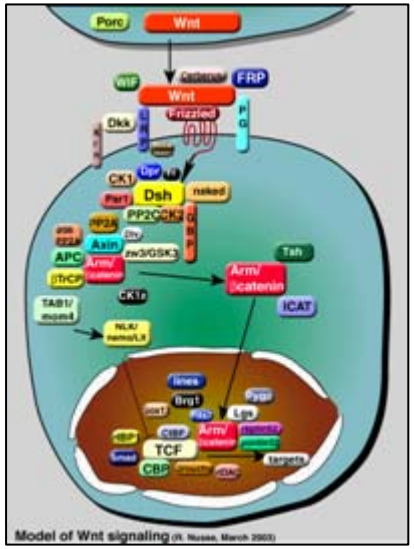
- all : all
  - GO:0008150 : biological_process
    - GO:0009987 : cellular process
      - GO:0050875 : cellular physiological process
        - GO:0044237 : cellular metabolism
          - GO:0006139 : nucleobase, nucleoside, nucleotide and nucleic acid metabolism
            - **GO:0006349 : imprinting**
    - GO:0007582 : physiological process
      - GO:0050875 : cellular physiological process
        - GO:0044237 : cellular metabolism
          - GO:0006139 : nucleobase, nucleoside, nucleotide and nucleic acid metabolism
            - **GO:0006349 : imprinting**

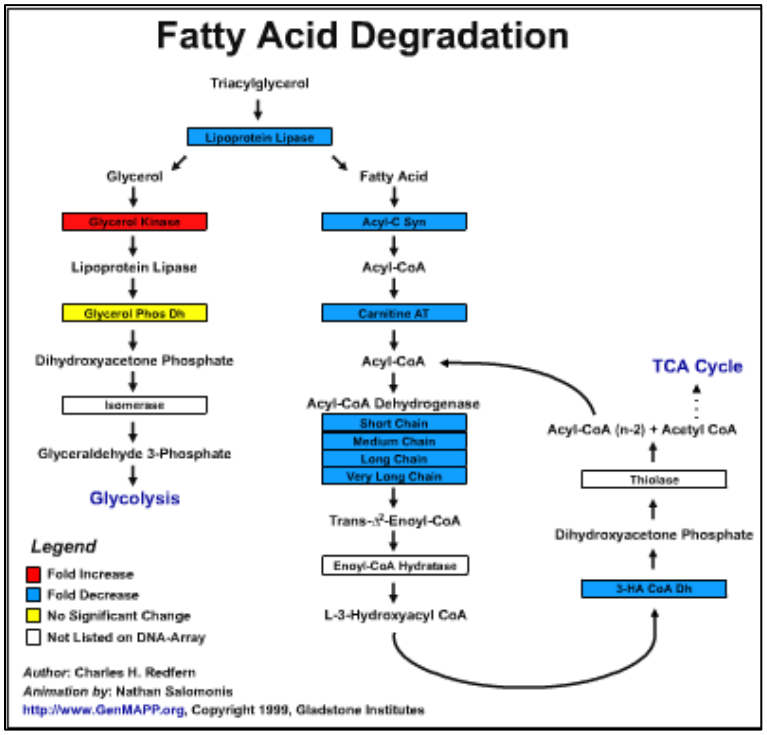
Qualifier	Symbol	Information	Source	Assigned By	Evidence	Reference
imprinting						
	DIRA3_HUMAN Sequence / GOst	DIRAS3, ARHI, NOEY2, RHOI: GTP-binding protein Di-Ras3, protein from *Homo sapiens*	UniProt	PINC	TAS	PMID:9874798
	DNM3A_HUMAN Sequence / GOst	DNMT3A: DNA, protein from *Homo sapiens*	UniProt	UniProt	ISS With UniProt:Q8IZV0	PMID:12138111
	DNM3L_HUMAN Sequence / GOst	DNMT3L: DNA, protein from *Homo sapiens*	UniProt	UniProt	NAS	PMID:12202768
	IGF2_HUMAN Sequence / GOst	IGF2, PP1446: Insulin-like growth factor II precursor, protein from *Homo sapiens*	UniProt	PINC	TAS	PMID:8968759
	Q8IZV0_HUMAN Sequence / GOst	DNMT3A2: DNA cytosine methyltransferase 3A2, protein from *Homo sapiens*	UniProt	UniProt	TAS	PMID:12138111

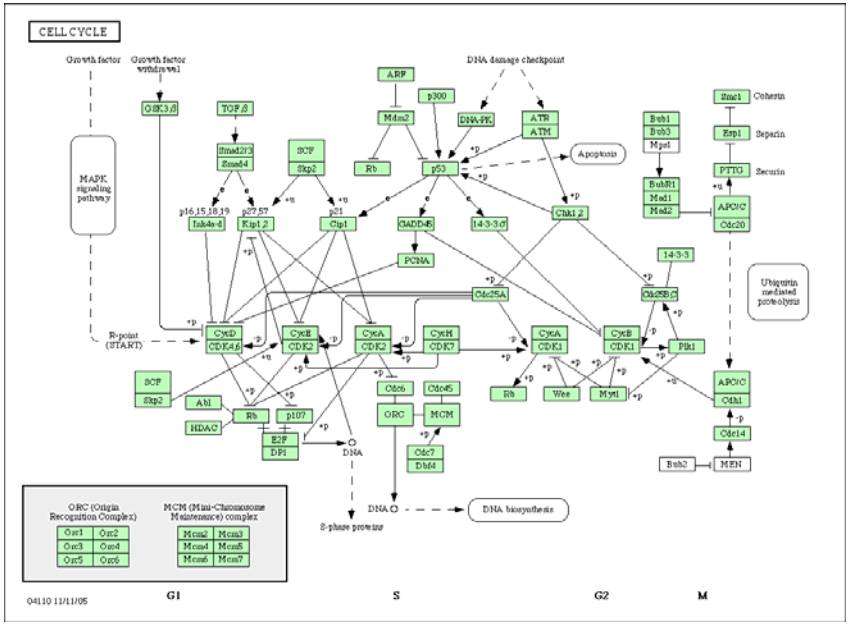
Check/Uncheck All | Get Detailed View | Submit.

Pathways

- Many databases of pathways: KEGG, GenMAPP, aMAZE, etc.

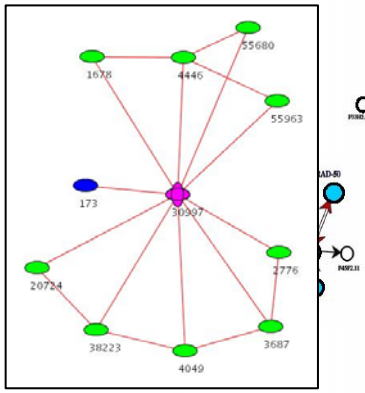




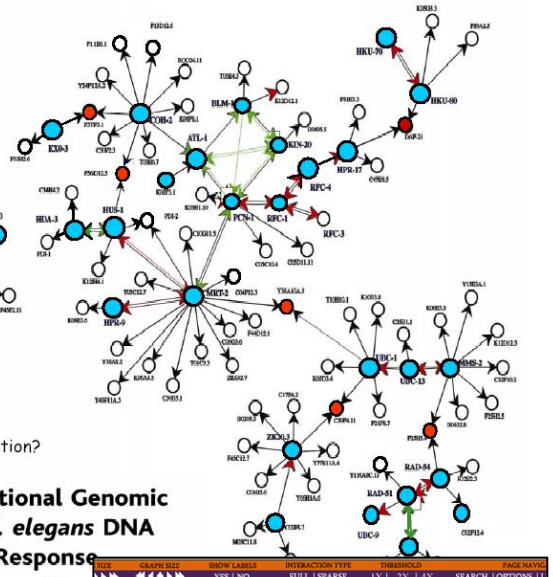


Protein-protein interaction

- Large databases of protein-protein interactions are becoming available
  - Yeast two-hybrid
  - Coimmunoprecipitation
- Data is getting cleaned and merged across organisms
  - Ulysses www.cisreg.ca
  - HiMAP www.himap.org

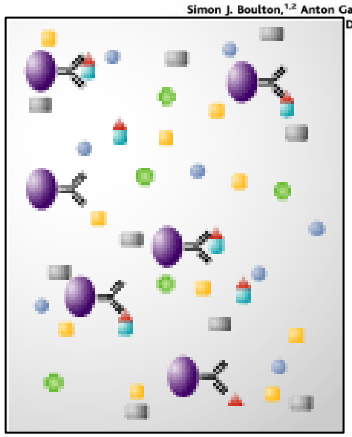


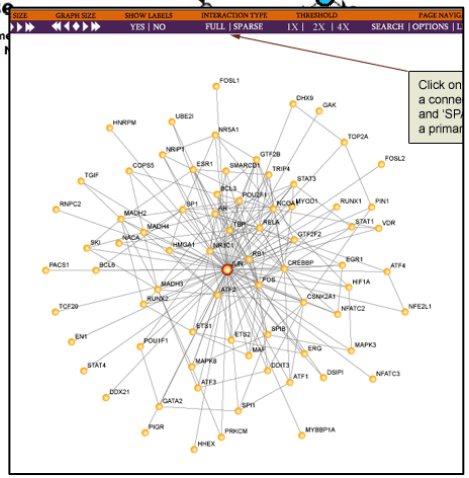
Functional connection?

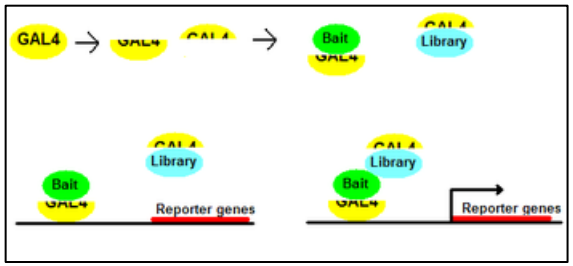


Combined Functional Genomic Maps of the *C. elegans* DNA Damage Response

Simon J. Boulton,[1,2] Anton Gartner,[3] Jérôme [illegible] David E. Hill,[1] [illegible]

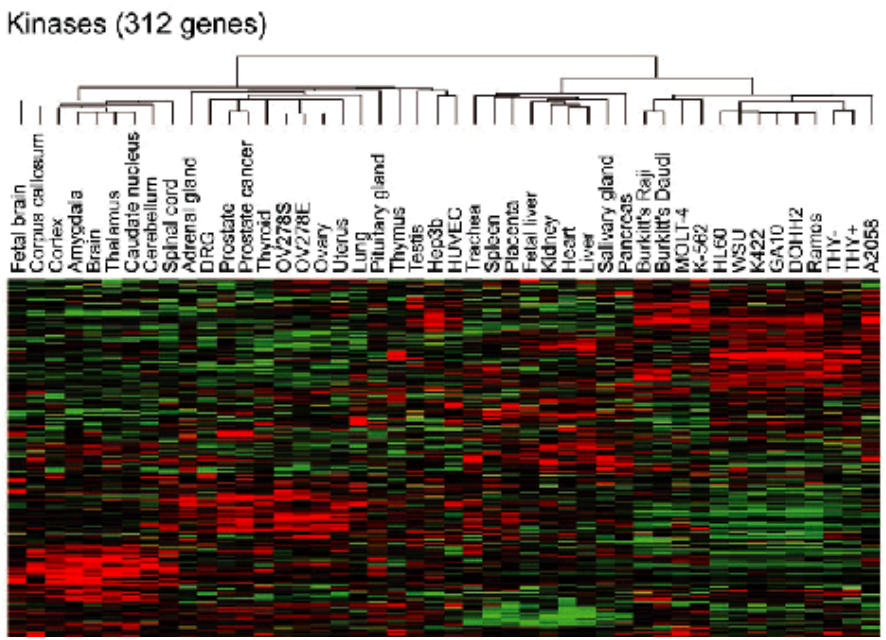






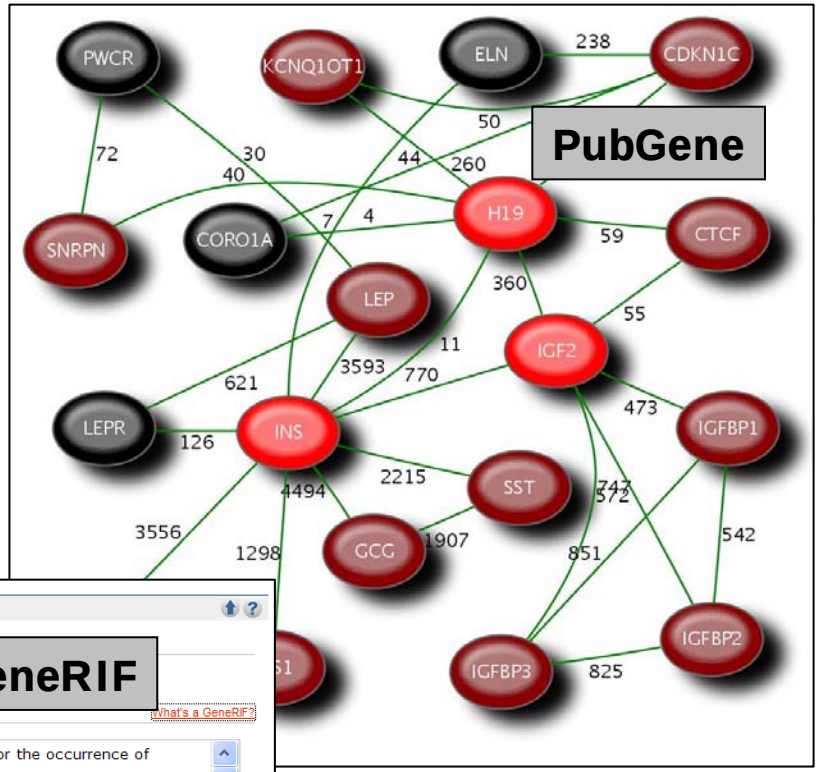
Microarray compendia

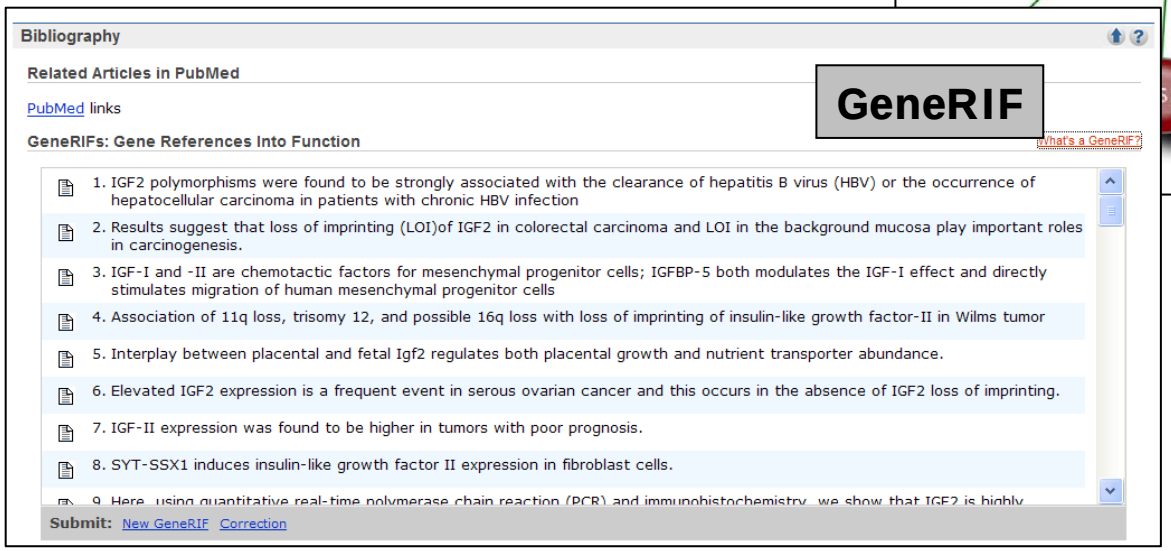
- Multiple large microarray data sets (compendia) are available that give a broad overview of general biological processes in different organisms
  - Su et al., Son et al., human and mouse tissues
  - Hughes et al., yeast mutants
  - Gasch et al., yeast stress
  - AtGenExpress, CAGE, Arabidopsis
- Available through microarray repositories
  - ArrayExpress
  - Gene Expression Omnibus



Literature abstracts

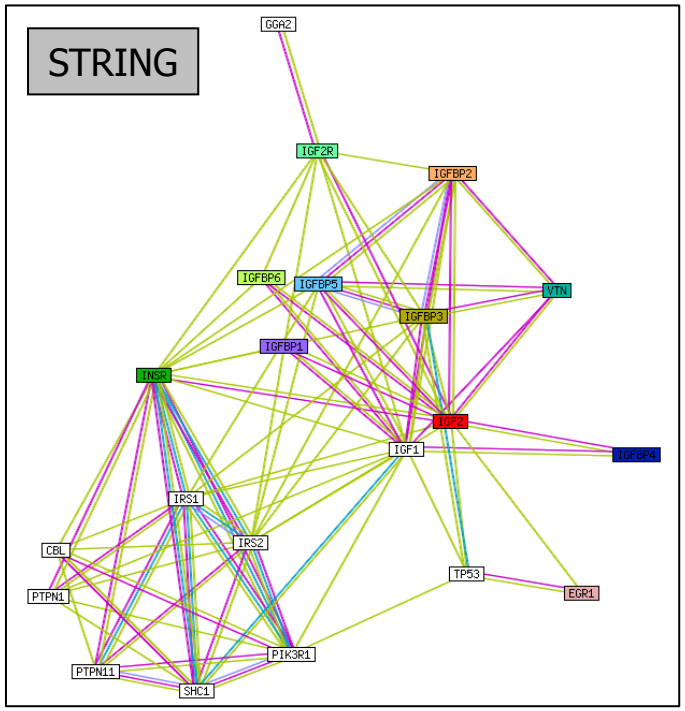
- PubMed
- EntrezGene GeneRIF
  www.ncbi.nlm.nih.gov/entrez/
- PubGene
  www.pubgene.org

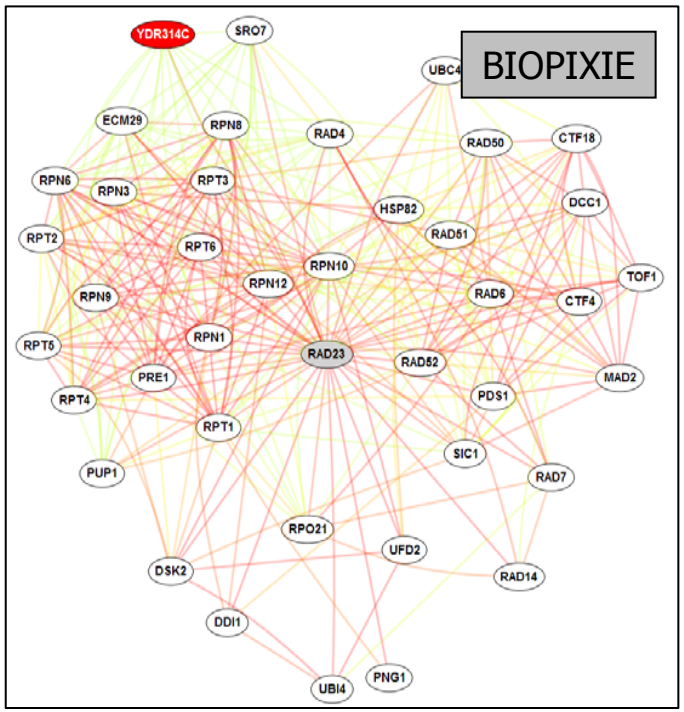


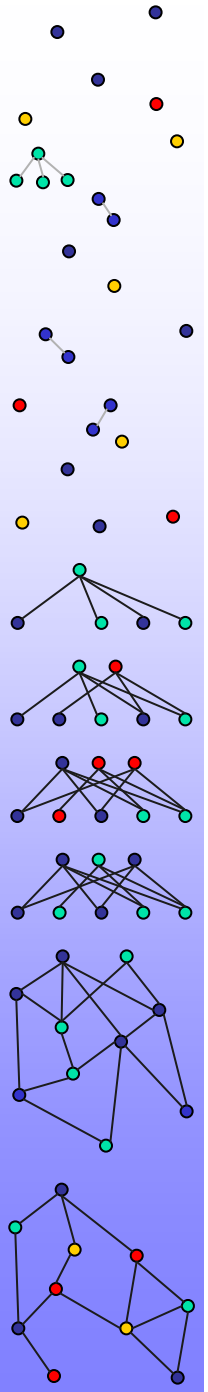


Multisource networks

- Some tools integrate multiple types of data to browse a network of genes
- BioPIXIE (yeast) pixie.princeton.edu
- STRING string.embl.de

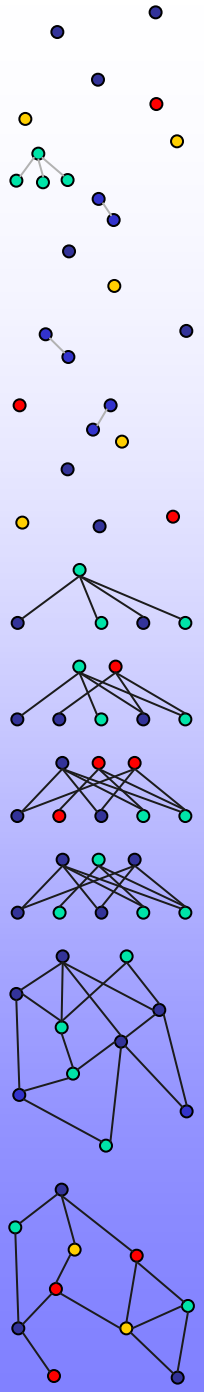






So much data...

So little time...



Candidate gene prioritization

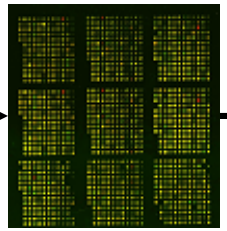
Array CGH: from diagnosis to gene discovery

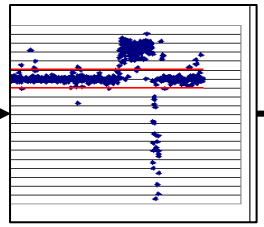
Patients with congenital & acquired disorders

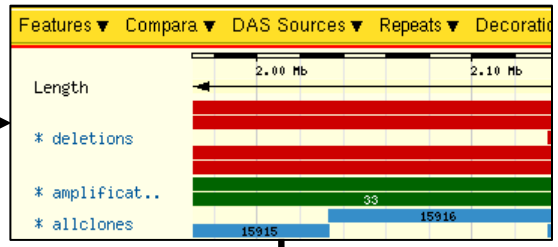
CGH microarrays Molecular karyotyping

Statistical analysis

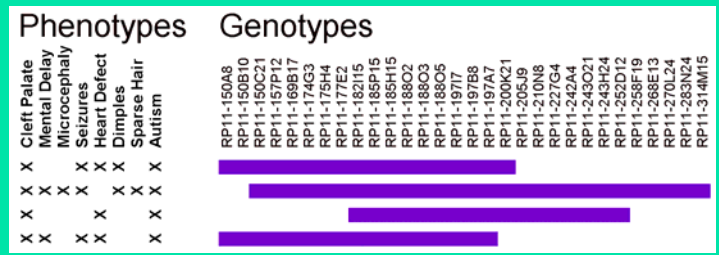
Location of chromosomal imbalances







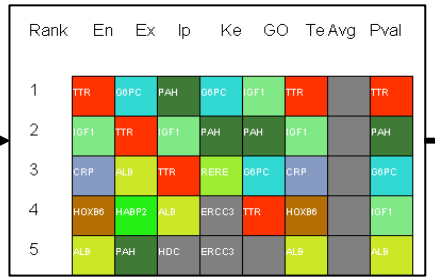
Databasing



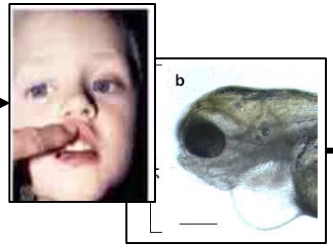
- Map chromosomal abnormalities
- Improved diagnosis

Discover new disease causing genes and explain their function

Prioritized candidate genes

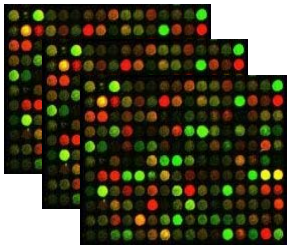


Validation

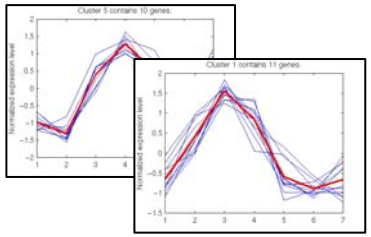


Candidate gene prioritization

High-throughput genomics

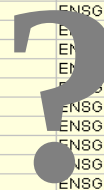


Data analysis

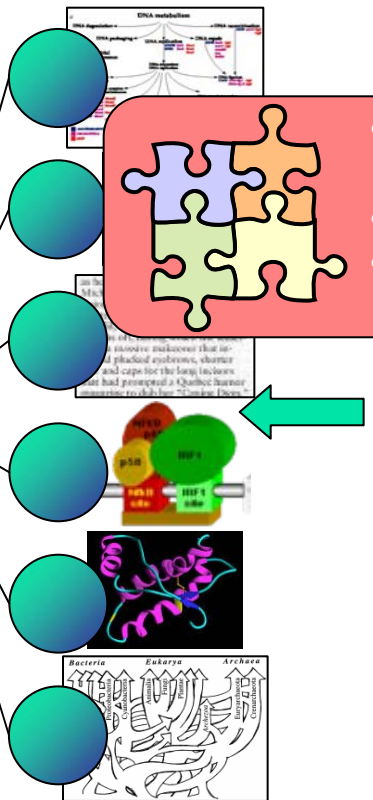


Candidate genes

Name	Ensembl
TTR	ENSG00000118271
PAH	ENSG00000171759
G6PC	ENSG00000131482
IGF1	ENSG00000017427
ALB	ENSG00000163631
CRP	ENSG00000132693
HABP2	ENSG00000148702
IF	ENSG00000138799
FST	ENSG00000134363
ARAF1	ENSG00000078061
HMGA2	ENSG00000149948
C9	ENSG00000113600
PCBP2	ENSG00000111406
HOXB6	ENSG00000108511
RERE	ENSG00000142599
HOXA11	ENSG00000005073
GRIN2A|GRIN2B	ENSG00000150086
SIM1	ENSG00000112246
C14orf10	ENSG00000092020
STX8	ENSG00000170310
MSH5	ENSG00000096474
CRH	ENSG00000147571
MID1	ENSG00000101871
TGFB3	ENSG00000184508
MMP3|MMP1	ENSG00000149968



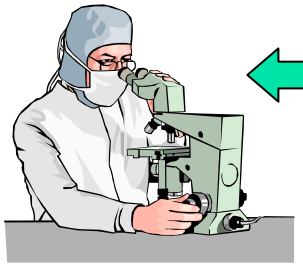
Information sources



Candidate prioritization

Rank	En	Ex	Ip	Ke	GO	Te	Avg	Pval
1	TTR	G6PC	PAH	G6PC	IGF1	TTR		TTR
2	IGF1	TTR	IGF1	PAH	PAH	IGF1		PAH
3	CRP	ALB	TTR	RERE	G6PC	CRP		G6PC
4	HOXB6	HABP2	ALB	ERCC3	TTR	HOXB6		IGF1
5	ALB	PAH	HDC	ERCC3		ALB		ALB
6	NR4A2	IF	TLL2	ANKRD3		HMGA2		CRP
7	PAH		C1QR1	ARAF1	HDC	NR4A2		
8	HOXA11	IGF1	G6PC	PKD2	F13A1	PAH		IF
9	NFYA	CRP	HABP2	MMR1	KCNN3	HOXA11	C13orf7	FST
10	C9	ARAF1	IF	HDC	CLIC1	NFYA	TTR	ARAF1

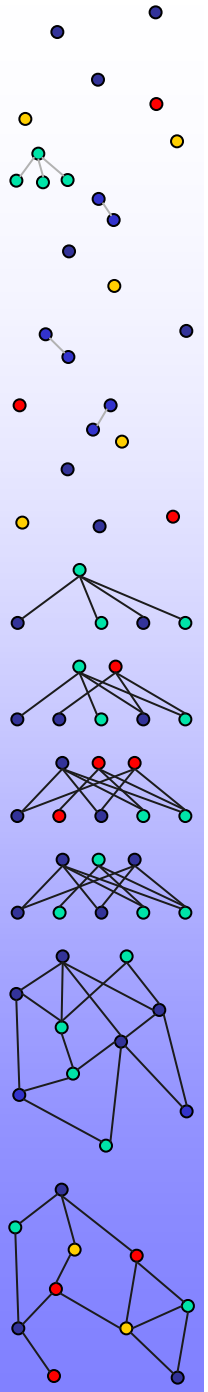
Validation



- Identify key genes and their function
- Emerging method
- Integration of multiple types of information

Prioritization by example

- Several cardiac abnormalities mapped to 3p22-25
  - Atrioventricular septal defect
  - Dilated cardiomyopathy
  - Brugada syndrome
- Candidate genes ("test set")
  - 3p22-25, 210 genes
- Known genes ("training set")
  - 10-15 genes: NKX2.5, GATA4, TBX5, TBX1, JAG1, THRAP, CFC1, ZFPM2, PTPN11, SEMA3E
  - Congenital heart defects (CHD)
- High scoring genes
  - ACVR2, SHOX2 - linked to heterotaxy and Turner syndrome (often associated with CHD)
  - Plexin-A1 - reported as essential for chick cardiac morphogenesis
  - Wnt5A, Wnt7A – neural crest guidance



Known T2D and obesity genes

- Type II diabetes
  - 21 known genes in OMIM
  - 118 known genes in GAD
- Obesity
  - 20 known genes in OMIM
  - 80 known genes in GAD
- Manually curated gene set from Elbers *et al.*, 2007
  - ACDC, ADRA2A, ADRA2B, ADRB1, ADRB2, ADRB3, LEP, LEPR, NR3C1, UCP1, UCP2, UCP3, PPARG, KCNJ11, and TCF7L2

Examples of prioritizations

- Prioritizations of control regions (from Elbers *et al.*, 2007):
  - 6q22-24: 220 candidates
  - 12q24: 327 candidates
  - 20q12-13: 357 candidates
- Other prioritizations (predictions):
  - 8p21.3: 106 candidates (from Tiffin *et al.*, 2006)
  - 4p16.1: 63 candidates (from Sandhu *et al.*, 2008)

Region 6q22-24: 220 candidates

SNP (rs648007) associated with T2D in African-Americans.

- Upregulation of ENPP1 transcription in liver and brain of diabetic rabbits compared with controls.
- ENPP1 121Q allele predicts susceptibility to T2D in south Asians and Caucasians.
- The Q allele of K121Q and the T allele of rs997509 were found to be associated with T2D in obese subjects from Poland.
- A risk haplotype was found to be associated with childhood obesity, adult morbid and moderate obesity and T2D.

Eller *et al.* (2006), Abate *et al.* (2005), Bochensky *et al.* (2006), Meyre *et al.* (2005)

Region 12q24: 327 candidates

Responsible for MODY, an uncommon monogenetic form of early onset T2D.

NCOR2 has an important role in the adipocyte by inhibiting adipocyte differentiation via repression of PPAR-g activity.

Key component in the reverse cholesterol transport pathway. Genetically associated with differences in insulin sensitivity in healthy subjects

McCarthy *et al.* (2006), Cohen *et al.* (2006), Perez-Martinez *et al.* (2005)

Region 20q12-13: 357 candidates

Responsible for MODY, an uncommon monogenetic form of early onset T2D.

Targeted disruption of the GNAS gene in mice leads to distinct phenotypes in heterozygotes, depending on whether the maternal (m-/+) or paternal (+/p-) allele is mutated. m-/+ mice become obese, whereas +/p- mice are thinner than normal. Both
m-/+ and +/p- mice have greater sensitivity to insulin, with low to normal fasting glucose levels, low fasting insulin levels, improved glucose tolerance and exaggerated hypoglycaemic response to administered insulin.

McCarthy *et al.* (2006), Yu *et al.* (2001)

Examples of prioritizations.

- Prioritizations of control regions (from Elbers *et al.*, 2007):
  - 6q22-24: 220 candidates
  - 12q24: 327 candidates
  - 20q12-13: 357 candidates
- Other prioritizations (predictions):
  - 8p21.3: 106 candidates (from Tiffin *et al.*, 2006)
  - 4p16.1: 63 candidates (from Sandhu *et al.*, 2008)

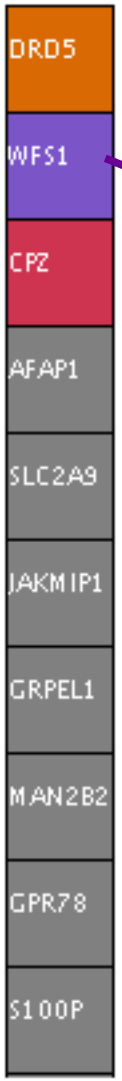
Region 8p21.3: 106 candidates

Candidate reported by Tiffin et al in 2006 in their study on T2D and obesity.

- TGFB downregulates LPL expression viaSP1/SP3 TFBS
- LPL activation regulates ACAA2 and ECHS1 in a rat model.
- PPARA, a T2D susceptibility gene, regulates LPL.
- Direct binding of LPL to VLDLR.
- Increased LPL activity increases the propensity for obesity and insulin resistance in mouse.

Tiffin *et al.* (2006), Irvine *et al.* (2005), Doi *et al.* (2005), Nagashima *et al.* (2005), Laplante *et al.* (2003), Schoonjans *et al.* (1996), Takahashi *et al.* (2004), Roberts *et al.* (2002), Kim *et al.* (2001), Duivenvoorden *et al.* (2005)

Region 4p16.1: 63 candidates

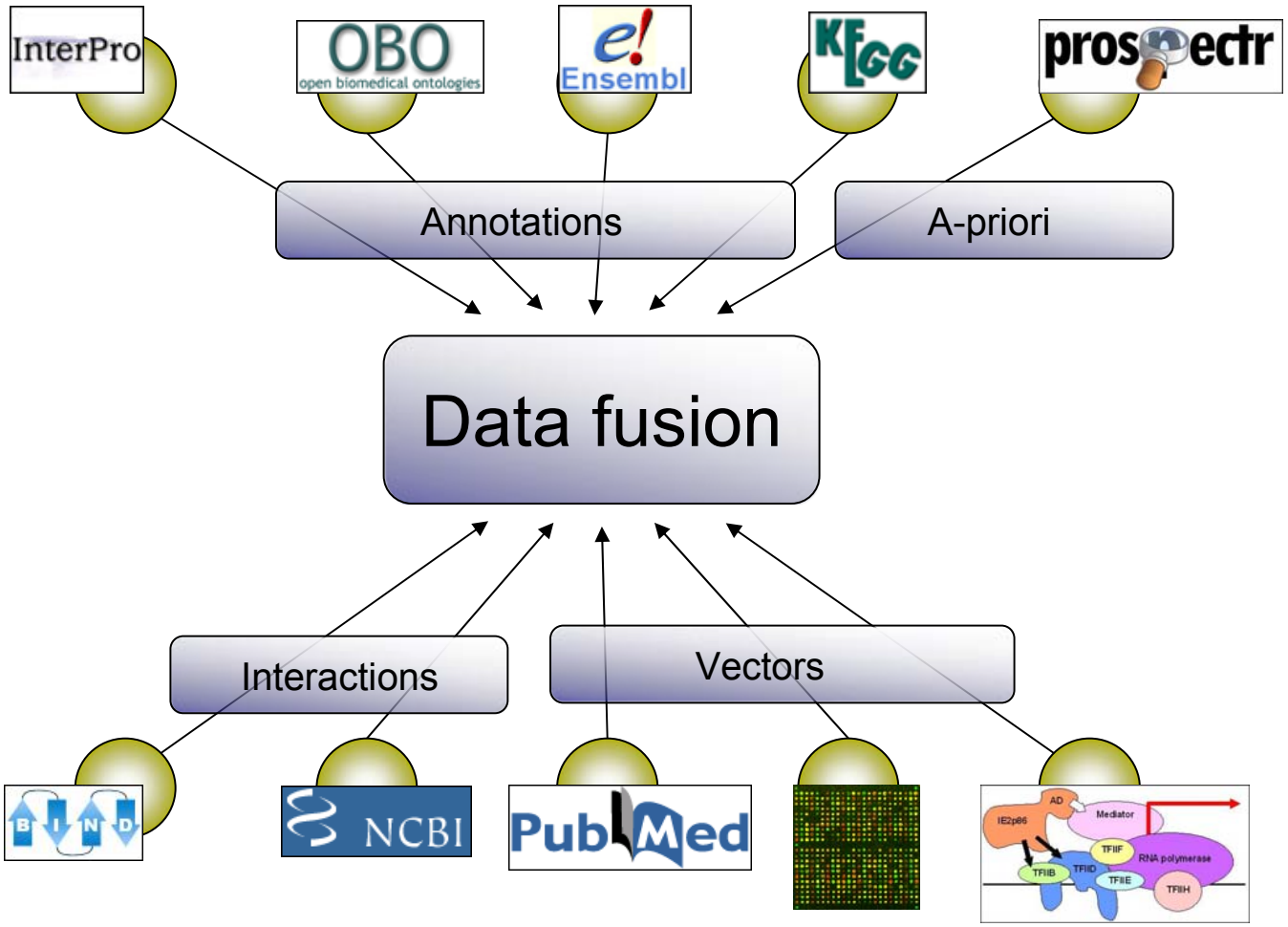


Candidate reported by Sandhu et al in 2008.

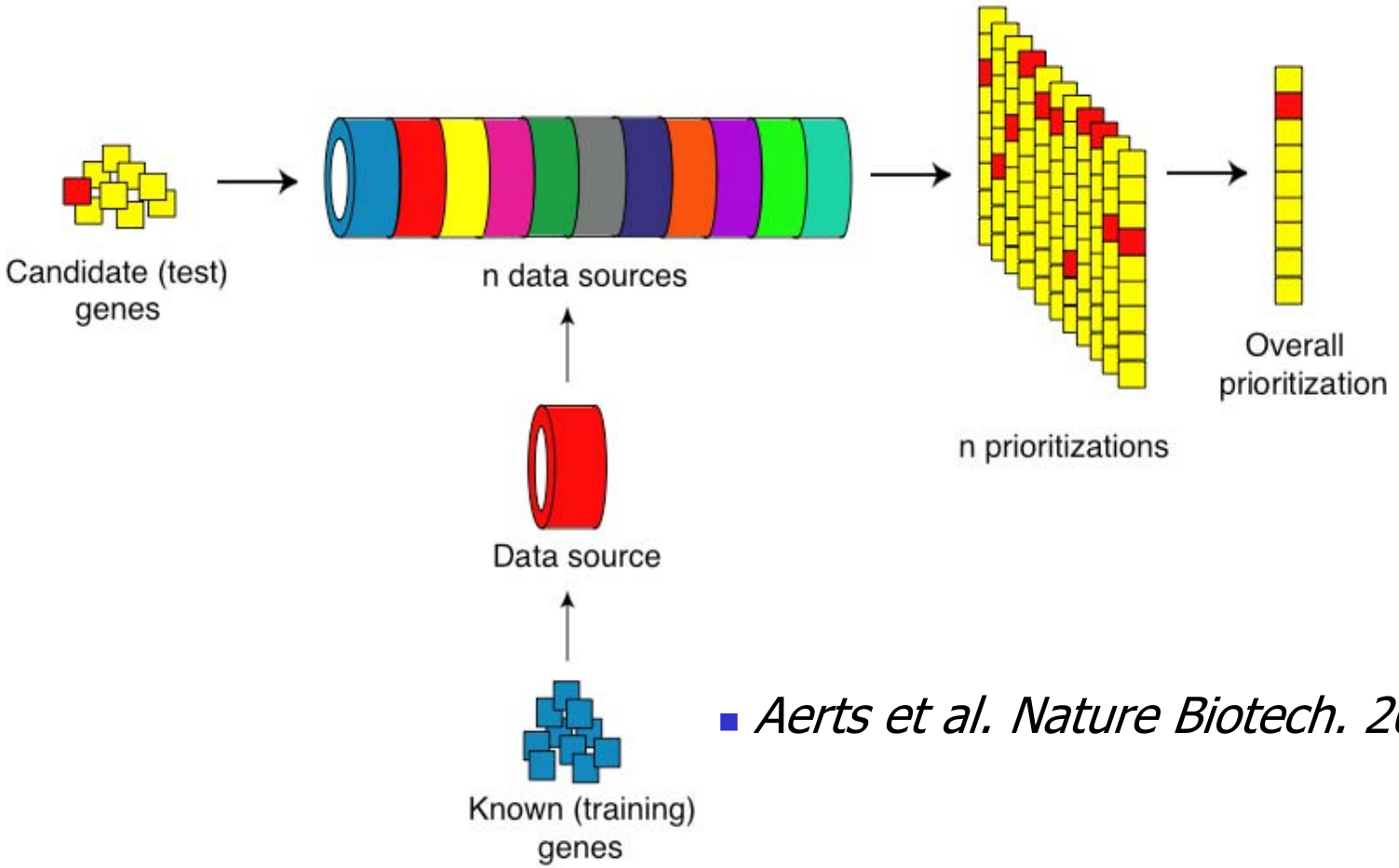
- Association between SNPs located in WFS1 and risk of T2D (rs10010131, rs6446482, rs752854, rs734312)
- Mutations cause Wolfram syndrome (characterized by diabetes insipidus, juvenile-onset non-autoimmune diabetes mellitus, optic atrophy and deafness).
- Disruption of WFS1 in mice causes overt diabetes or impaired glucose tolerance.
- Both humans and mice deficient in Wolframin show pancreatic beta cell loss.

Sandhu *et al.* (2008), Inoue *et al.* (1998), Strom *et al.* (1998), Riggs *et al.* (2005), Ishihara *et al.* (2004), Karasik *et al.* (1989), Yamada *et al.* (2006)

Multiple sources of information

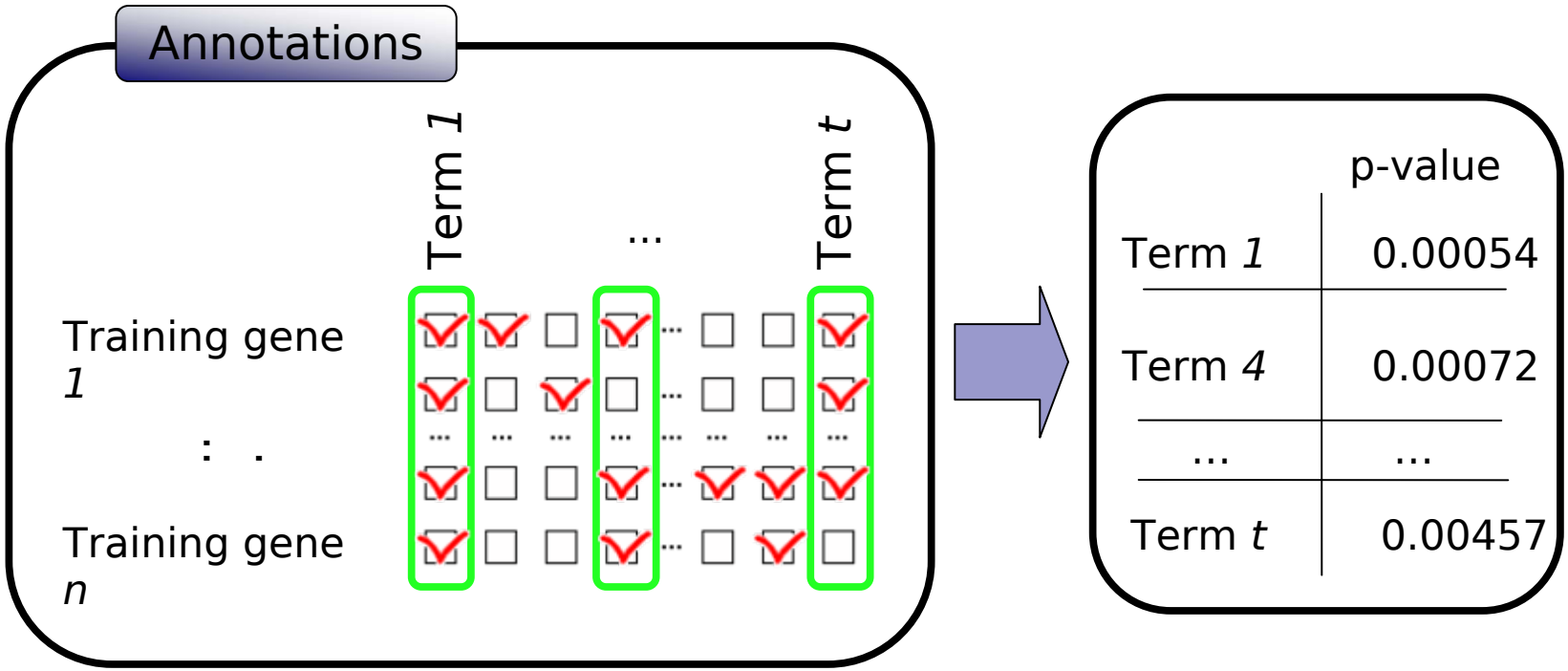


Data fusion with order statistics



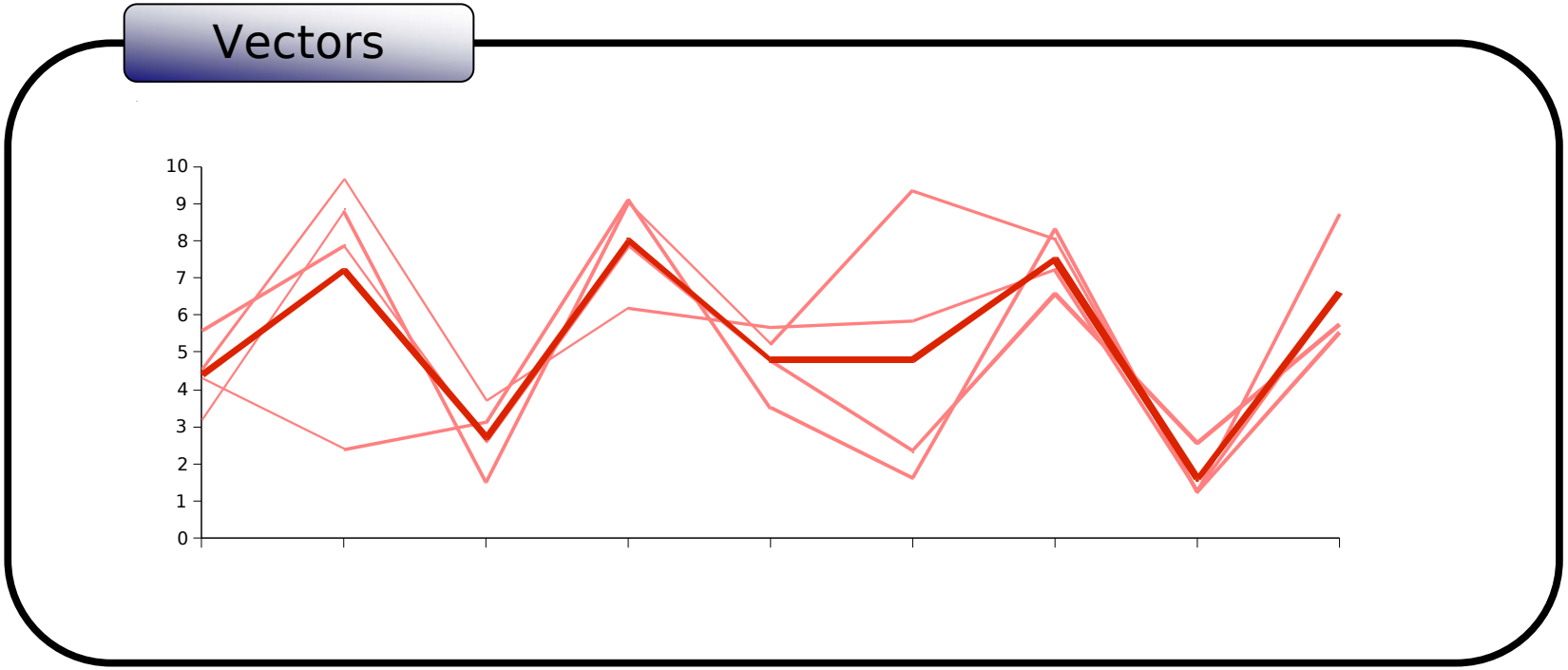
- *Aerts et al. Nature Biotech. 2006*

Training of an attribute submodel



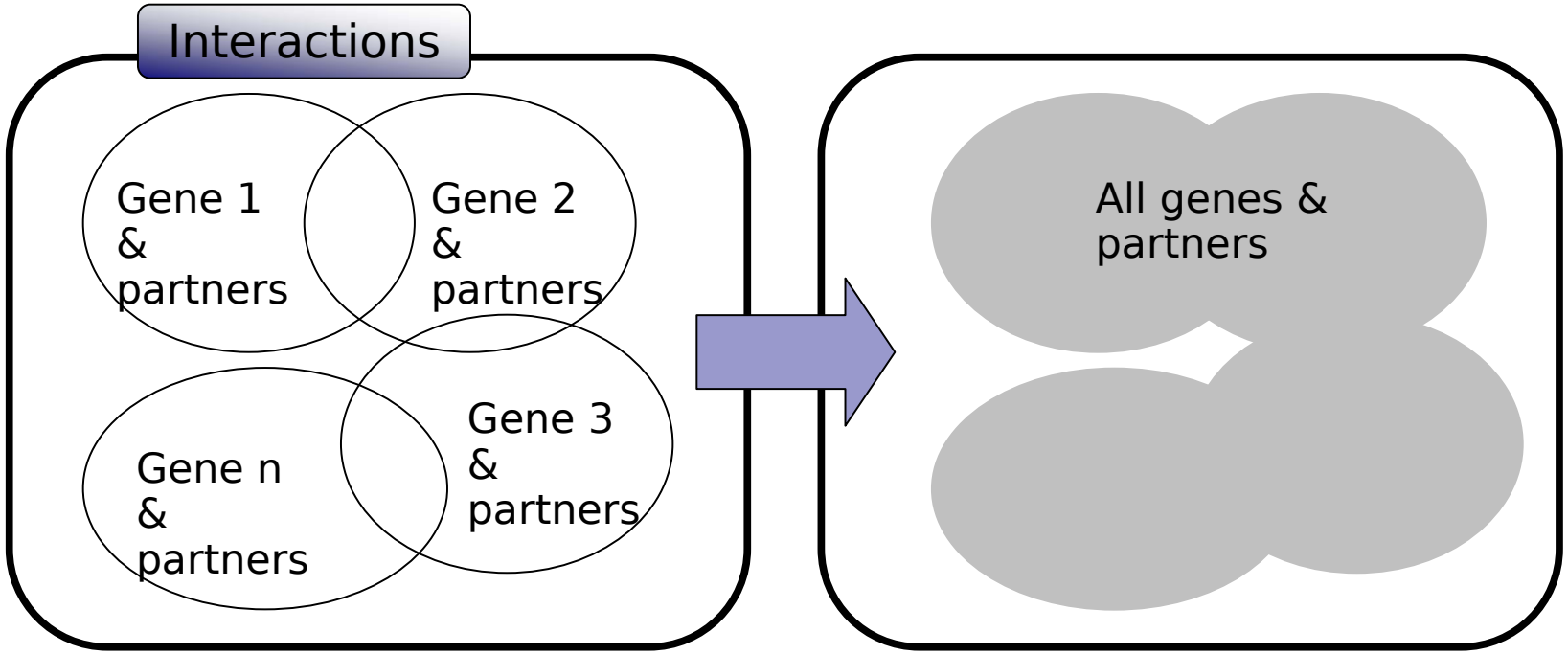
- A term is over-represented if its frequency inside the training set is significantly larger than its frequency over the genome
- Gene Ontology, Interpro, KEGG & EST submodels

Training of a vector submodel



- A collection of profiles (here numerical vectors) can be represented by the average profile
- Microarray, motif & text submodels

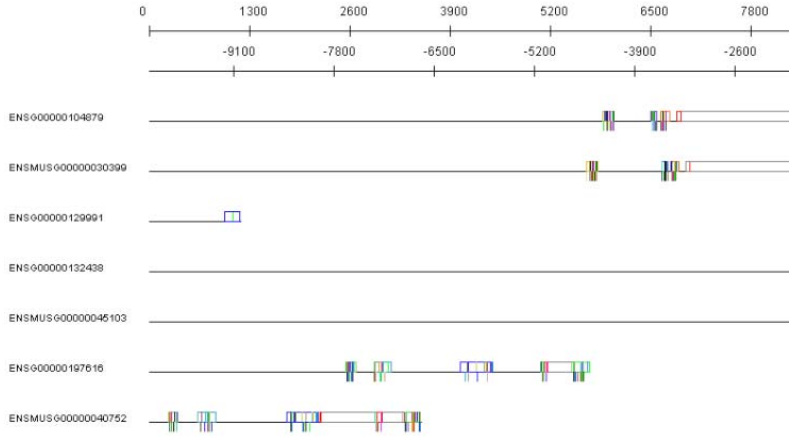
Training of a set submodel



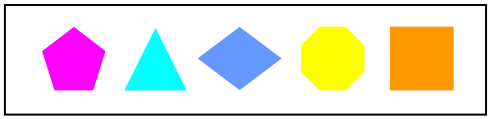
- We group together all gene partners in one set
- BIND protein-protein interaction submodels

Other submodels

- Disease probabilities
  - Phylogenetic score of conservation
  - Precomputed score
- BLAST
  - Lowest BLAST score
- Cis-regulatory module
  - Combinatorial model of transcriptional regulation



ModuleSearcher



p,v

211 bp

Order statistics

- Given a set of n ordered rank ratios for gene i

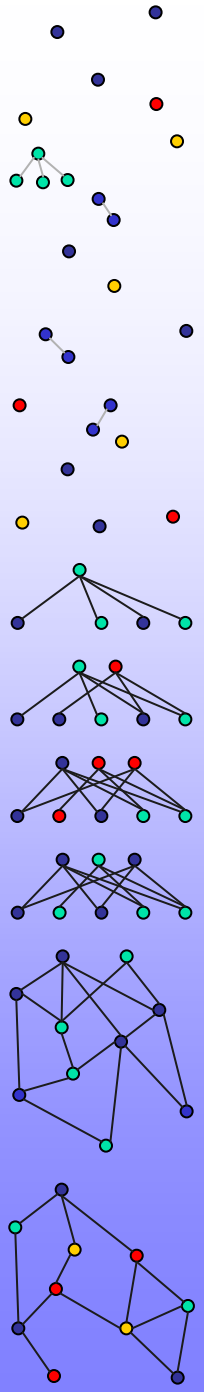
$$(9/100;\ 4/120;\ 30/150;\ 30/50;\ 2/10;\ 80/80)$$
$$\rightarrow (0.09;\ 0.03;\ 0.2;\ 0.5;\ 0.2;\ 0.3)$$
$$\rightarrow (0.03;\ 0.09;\ 0.2;\ 0.2;\ 0.3;\ 0.5;\ 0.6;\ 1)$$

- What is the probability of getting these rank ratios or better by chance alone?
- "How many rank vectors does my vector strictly dominate?"
- Joint probability density function of all n order statistics

$$Q(r_1, r_2, \ldots, r_n) = n! \int_0^{r_1} \int_{s_1}^{r_2} \ldots \int_{s_{n-1}}^{r_n} ds_n ds_{n-1} \ldots ds_1$$

- Recursive formula of complexity O(n^2)

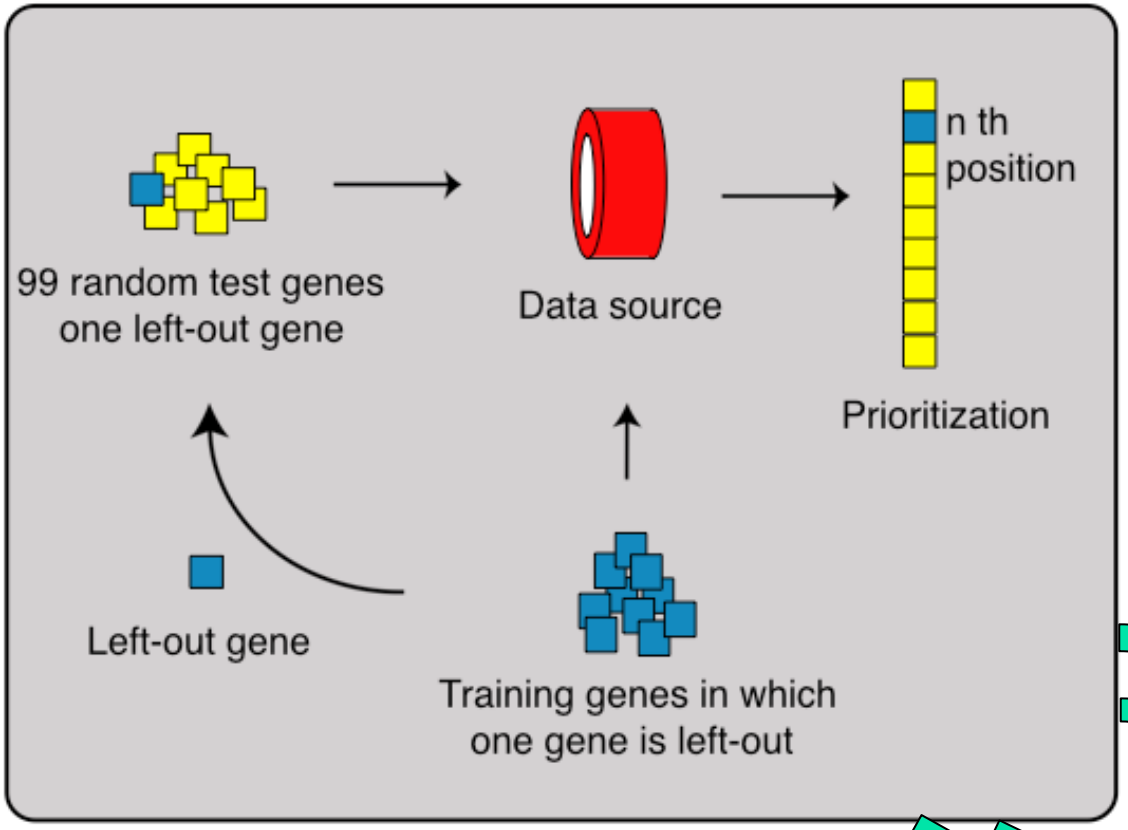
$$V_k = \int \ldots \int = \sum_{i=1}^{k-1} (-1)^{i-1} \frac{V_{k-i}}{i!} r_{n-k-1}^i, \quad V_0 = 1$$



OMIM & GO cross-validation

- Diseases
  - Alzheimer's disease, amyotrophic lateral sclerosis (ALS), anemia, breast cancer, cardiomyopathy, cataract, charcot-marie-tooth disease, colorectal cancer, deafness, diabetes, dystonia, Ehlers-Danlos, epilepsy, hemolytic anemia, ichthyosis, leukemia, lymphoma, mental retardation, muscular dystrophy, myopathy, neuropathy, obesity, Parkinson's disease, retinitis pigmentosa, spastic paraplegia, spinocerebellar ataxia, usher syndrome, xeroderma pigmentosum, Zellweger syndrome
- Pathways
  - Wnt pathway members (GO:0016055: Wnt receptor signaling pathway)
  - Notch pathway members (GO:0007219: Notch signaling pathway)
  - EGFR pathway members (GO:0007173: epidermal growth factor receptor signaling pathway)

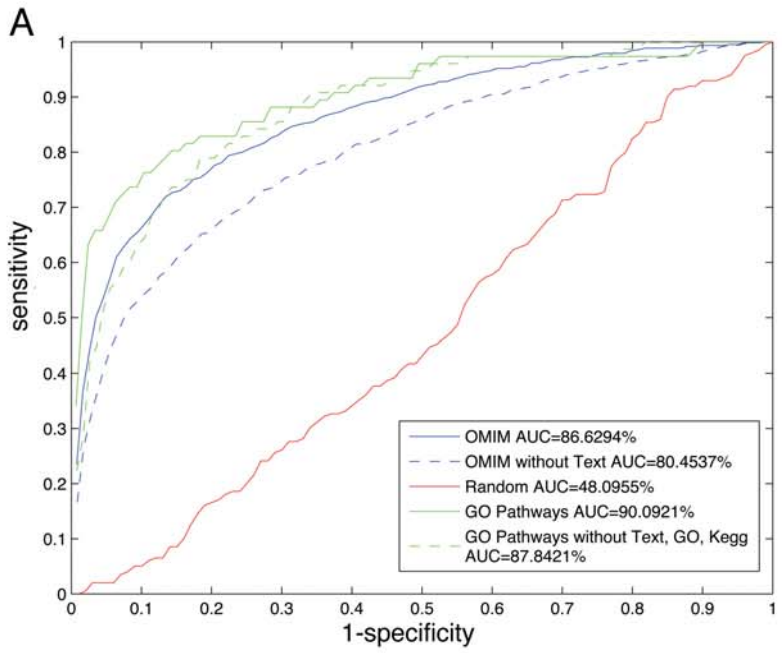
Cross-validation

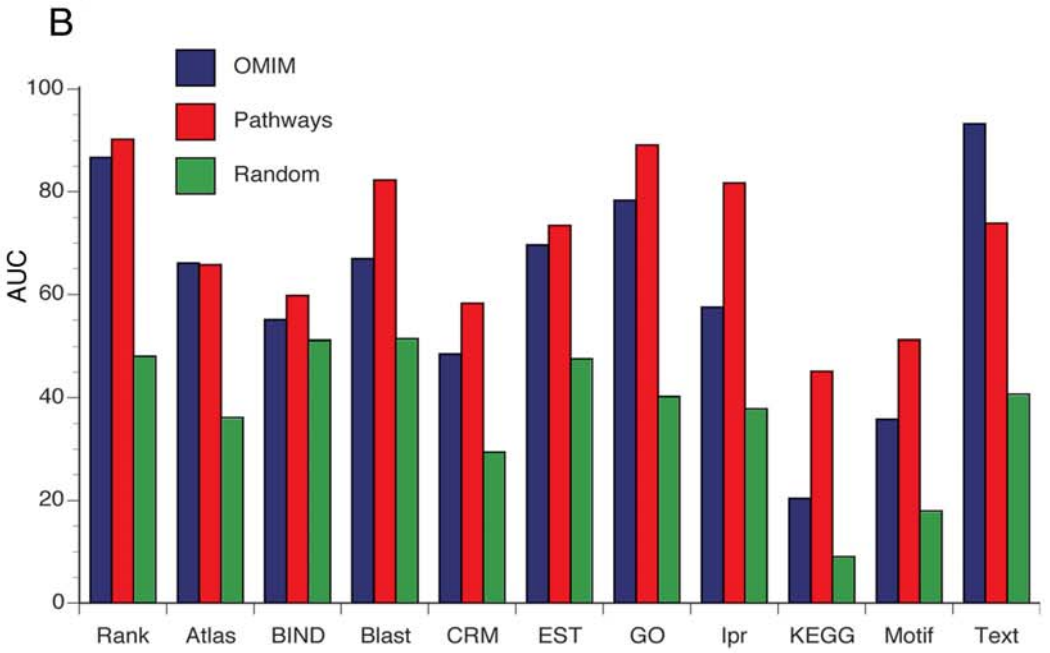


Repeat
- For each gene
- For each disease or pathway

Compute average rank

Rank ROC curves

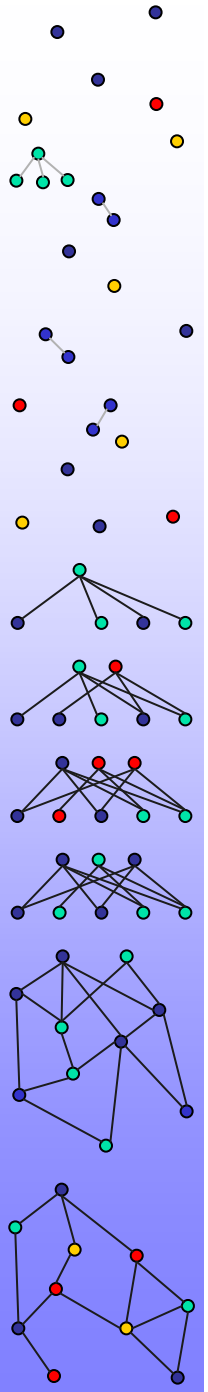




Evaluation on monogenic diseases + text model

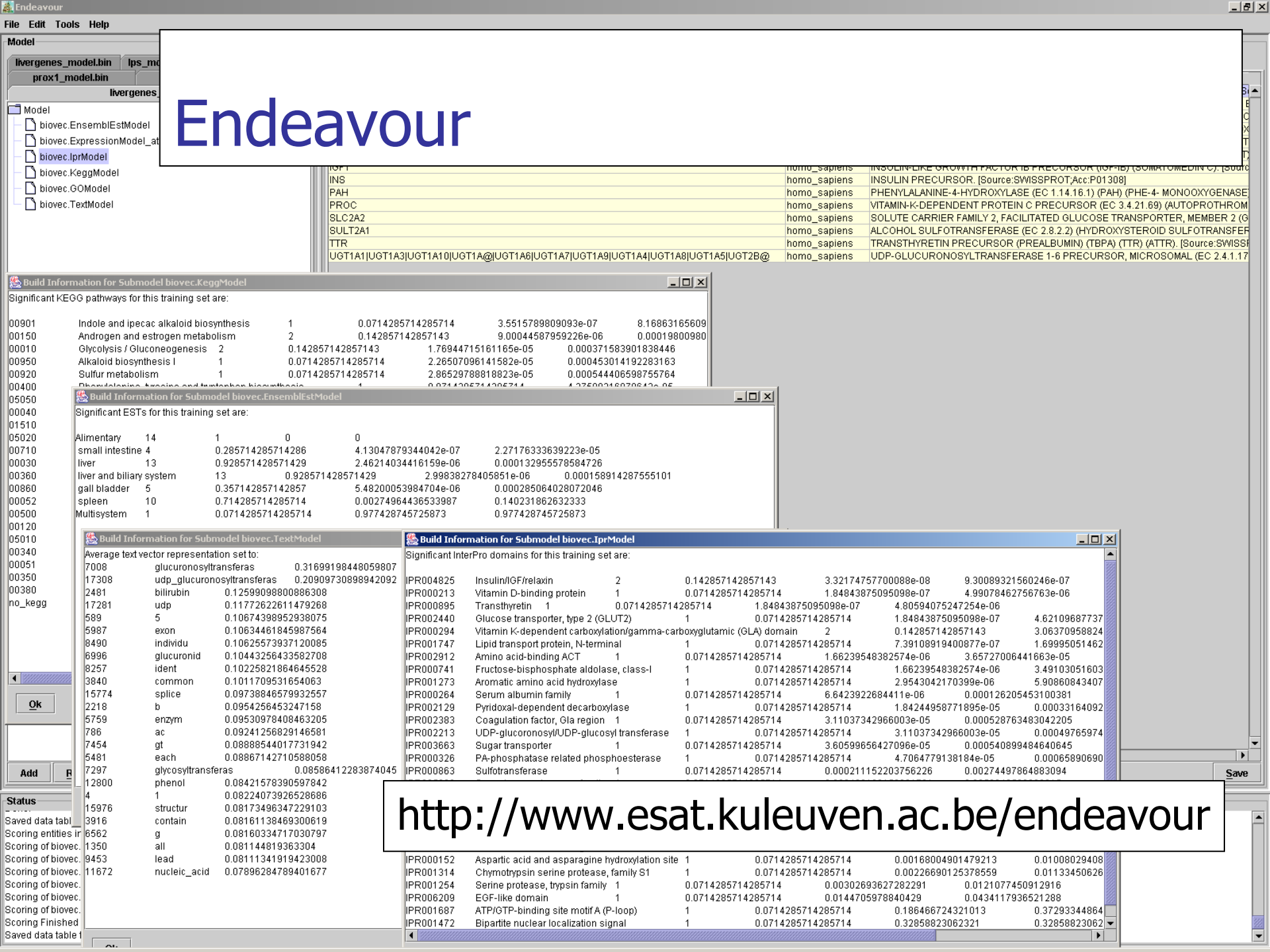
- Validation of the text model
  - Artificially high performance of text model due to explicit links between genes and diseases!
  - Roll-back experiment on textual information

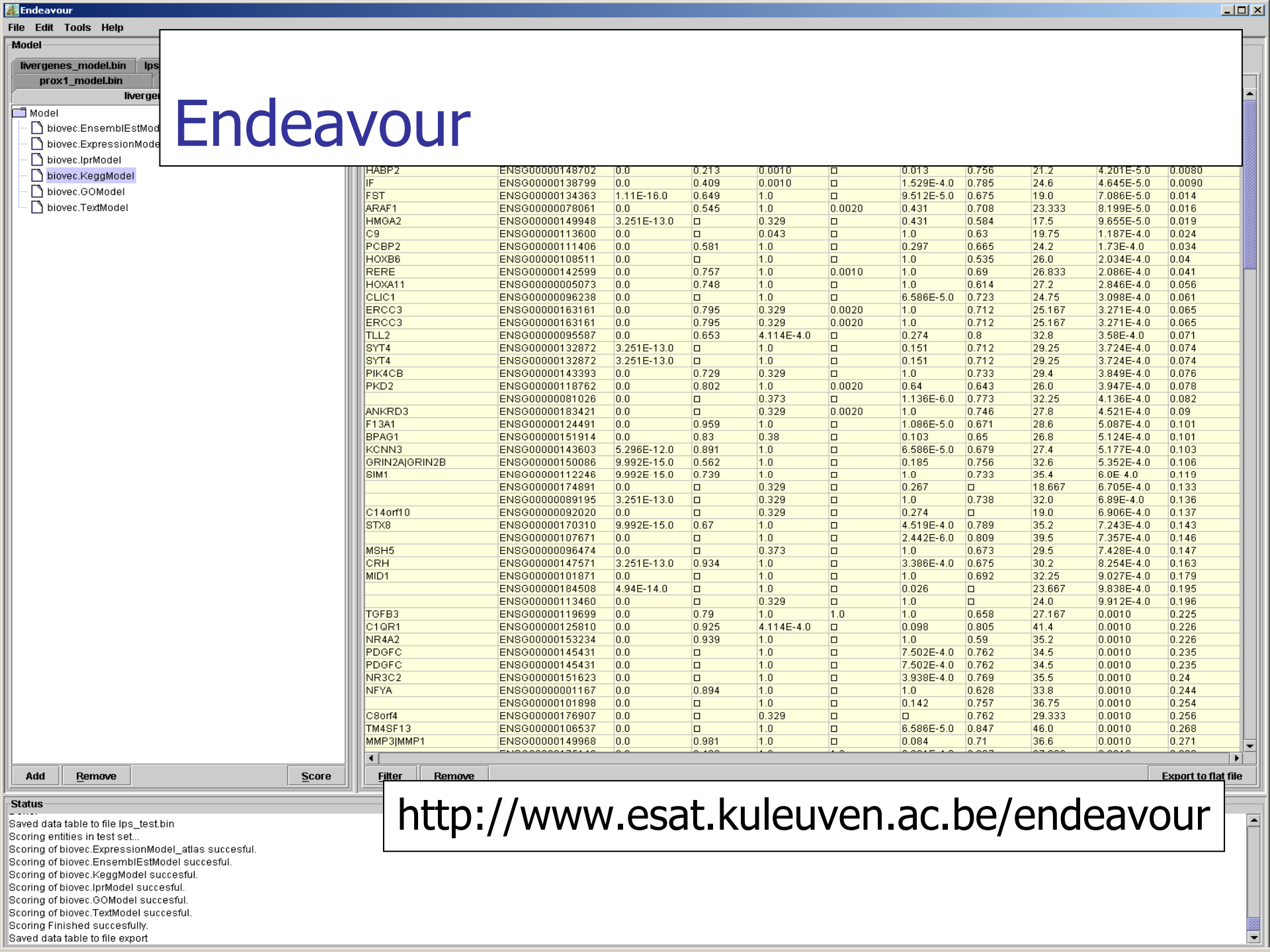
Disease	Hugo	Rolled-back text only	All	All, no text
Amyotrophic lateral sclerosis	DCTN1	97	27	23
Arrhythmias	Ca(V)1.2	3	4	4
Cardiomyopathy 1	CAV3	1	2	8
Cardiomyopathy 2	ABCC9	51	1	1
Charcot-Marie-Tooth	DNM2	100	14	12
Congenital heart disease	CRELD1	1	3	6
Cornelia de Lange	NIPBL	75	9	3
Distal hereditary motor neuropathy	BSCL2	62	15	6
Klippel-Trenaunay	VG5Q	39	3	3
Parkinson's disease	LRRK2	No text available	50	42
Average Rank		48±13	13±5	11±4



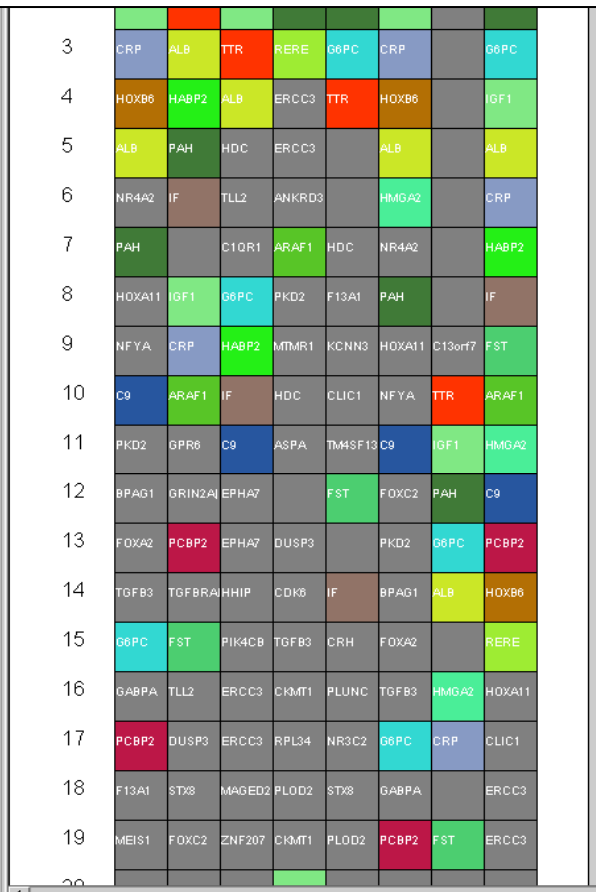
Complex disease

Disease	Gene	All	All, no Text
Atherosclerosis 1	TNFSF4	54	111
Crohn's Disease	OCTN	71	85
Parkinson's Disease	GBA	23	2
Rheumatoid Arthritis	PTPN22	11	22
Atherosclerosis 2	ALOX5AP	29	46
Alzheimer's Disease	UBQNL1	54	56
Average rank		40±10	54±17


Endeavour
File Edit Tools Help
Model
livergenes_model.bin
prox1_model.bin
Model
biovec.EnsemblEstModel
biovec.ExpressionModel_at
biovec.IprModel
biovec.KeggModel
biovec.GOModel
biovec.TextModel
INS homo_sapiens INSULIN PRECURSOR. [Source:SWISSPROT;Acc:P01308]
PAH homo_sapiens PHENYLALANINE-4-HYDROXYLASE (EC 1.14.16.1) (PAH) (PHE-4- MONOOXYGENASE
PROC homo_sapiens VITAMIN-K-DEPENDENT PROTEIN C PRECURSOR (EC 3.4.21.69) (AUTOPROTHROM
SLC2A2 homo_sapiens SOLUTE CARRIER FAMILY 2, FACILITATED GLUCOSE TRANSPORTER, MEMBER 2 (G
SULT2A1 homo_sapiens ALCOHOL SULFOTRANSFERASE (EC 2.8.2.2) (HYDROXYSTEROID SULFOTRANSFER
TTR homo_sapiens TRANSTHYRETIN PRECURSOR (PREALBUMIN) (TBPA) (TTR) (ATTR). [Source:SWISSP
UGT1A1|UGT1A3|UGT1A10|UGT1A@|UGT1A6|UGT1A7|UGT1A9|UGT1A4|UGT1A8|UGT1A5|UGT2B@ homo_sapiens UDP-GLUCURONOSYLTRANSFERASE 1-6 PRECURSOR, MICROSOMAL (EC 2.4.1.17
Build Information for Submodel biovec.KeggModel
Significant KEGG pathways for this training set are:
00901 Indole and ipecac alkaloid biosynthesis 1 0.0714285714285714 3.5515789809093e-07 8.16863165609
00150 Androgen and estrogen metabolism 2 0.142857142857143 9.00044587959226e-06 0.00019800980
00010 Glycolysis / Gluconeogenesis 2 0.142857142857143 1.76944715161165e-05 0.00037158390183846
00950 Alkaloid biosynthesis I 1 0.0714285714285714 2.26507096141582e-05 0.000453014192283163
00920 Sulfur metabolism 1 0.0714285714285714 2.86529788818823e-05 0.000544406598755764
Build Information for Submodel biovec.EnsemblEstModel
Significant ESTs for this training set are:
Alimentary 14 1 0 0
small intestine 4 0.28571428571286 4.13047879344042e-07 2.27176333639223e-05
liver 13 0.928571428571429 2.46214034416159e-06 0.000132955578584726
liver and biliary system 13 0.928571428571429 2.99838278405851e-06 0.000158914287555101
gall bladder 5 0.357142857142857 5.48200053984704e-06 0.000285064028072046
spleen 10 0.714285714285714 0.00274964436533987 0.140231862632333
Multisystem 1 0.0714285714285714 0.977428745725873 0.977428745725873
Build Information for Submodel biovec.TextModel
Average text vector representation set to:
7008 glucuronosyltransferas 0.31699198448059807
17308 udp_glucuronosyltransferas 0.20909730898942092
2481 bilirubin 0.12599098800886308
17281 udp 0.11772622611479268
589 5 0.10674398952938075
5987 exon 0.10634461845987564
8490 individu 0.10625573937120085
6996 glucuronid 0.10443256433582708
8257 ident 0.10225821864645528
3840 common 0.1011709531654063
15774 splice 0.09738846579932557
2218 b 0.0954256453247158
5759 enzym 0.09530978408463205
786 ac 0.09241256829146581
7454 gt 0.08888544017731942
5481 each 0.08867142710588058
7297 glycosyltransferas 0.08586412283874045
12800 phenol 0.08421578390597842
4 1 0.08224073926528686
15976 structur 0.08173496347229103
3916 contain 0.08161138469300619
6562 g 0.08160334717030797
1350 all 0.081144819363304
9453 lead 0.08111341919423008
11672 nucleic_acid 0.07896284789401677
Build Information for Submodel biovec.IprModel
Significant InterPro domains for this training set are:
IPR004825 Insulin/IGF/relaxin 2 0.142857142857143 3.32174757700088e-08 9.30089321560246e-07
IPR000213 Vitamin D-binding protein 1 0.0714285714285714 1.84843875095098e-07 4.99078462756763e-06
IPR000895 Transthyretin 1 0.0714285714285714 1.84843875095098e-07 4.80594075247254e-06
IPR002440 Glucose transporter, type 2 (GLUT2) 1 0.0714285714285714 1.84843875095098e-07 4.62109687737
IPR000294 Vitamin K-dependent carboxylation/gamma-carboxyglutamic (GLA) domain 2 0.142857142857143 3.06370958824
IPR001747 Lipid transport protein, N-terminal 1 0.0714285714285714 7.39108919400877e-07 1.69995051462
IPR002912 Amino acid-binding ACT 1 0.0714285714285714 1.66239548382574e-06 3.65727006441663e-05
IPR000741 Fructose-bisphosphate aldolase, class-I 1 0.0714285714285714 1.66239548382574e-06 3.49103051603
IPR001273 Aromatic amino acid hydroxylase 1 0.0714285714285714 2.9543042170399e-06 5.90860843407
IPR000264 Serum albumin family 1 0.0714285714285714 6.64239226844l1e-06 0.000126205453100381
IPR002129 Pyridoxal-dependent decarboxylase 1 0.0714285714285714 1.84244958771895e-05 0.00033164092
IPR002383 Coagulation factor, Gla region 1 0.0714285714285714 3.11037342966003e-05 0.000528763483042205
IPR002213 UDP-glucoronosyl/UDP-glucosyl transferase 1 0.0714285714285714 3.11037342966003e-05 0.00049765974
IPR003663 Sugar transporter 1 0.0714285714285714 3.60599656427096e-05 0.000540899484640645
IPR000326 PA-phosphatase related phosphoesterase 1 0.0714285714285714 4.7064779138184e-05 0.00065890690
IPR000863 Sulfotransferase 1 0.0714285714285714 0.000211152203756226 0.00274497864883094
IPR000152 Aspartic acid and asparagine hydroxylation site 1 0.0714285714285714 0.00168004901479213 0.01008029408
IPR001314 Chymotrypsin serine protease, family S1 1 0.0714285714285714 0.00226690125378559 0.01133450626
IPR001254 Serine protease, trypsin family 1 0.0714285714285714 0.00302693627282291 0.0121077450912916
IPR006209 EGF-like domain 1 0.0714285714285714 0.0144705978840429 0.0434117936521288
IPR001687 ATP/GTP-binding site motif A (P-loop) 1 0.0714285714285714 0.186466724321013 0.37293344864
IPR001472 Bipartite nuclear localization signal 1 0.0714285714285714 0.32858823062321 0.32858823062
Ok
Add
Save
Status
Saved data tabl
Scoring entities in
Scoring of biovec.
Scoring of biovec.
Scoring of biovec.
Scoring of biovec.
Scoring of biovec.
Scoring Finished
Saved data table
http://www.esat.kuleuven.ac.be/endeavour


Endeavour
File Edit Tools Help
Model
livergenes_model.bin
prox1_model.bin
Model
biovec.EnsemblEstMod
biovec.ExpressionMode
biovec.IprModel
biovec.KeggModel
biovec.GOModel
biovec.TextModel
Add
Remove
Score
HABP2 ENSG00000148702 0.0 0.213 0.0010 0.013 0.756 21.2 4.201E-5.0 0.0080
IF ENSG00000138799 0.0 0.409 0.0010 1.529E-4.0 0.785 24.6 4.645E-5.0 0.0090
FST ENSG00000134363 1.11E-16.0 0.649 1.0 9.512E-5.0 0.675 19.0 7.086E-5.0 0.014
ARAF1 ENSG00000078061 0.0 0.545 1.0 0.0020 0.431 0.708 23.333 8.199E-5.0 0.016
HMGA2 ENSG00000149948 3.251E-13.0 0.329 0.431 0.584 17.5 9.655E-5.0 0.019
C9 ENSG00000113600 0.0 0.043 1.0 0.63 19.75 1.187E-4.0 0.024
PCBP2 ENSG00000111406 0.0 0.581 1.0 0.297 0.665 24.2 1.73E-4.0 0.034
HOXB6 ENSG00000108511 0.0 1.0 1.0 0.535 26.0 2.034E-4.0 0.04
RERE ENSG00000142599 0.0 0.757 1.0 0.0010 1.0 0.69 26.833 2.086E-4.0 0.041
HOXA11 ENSG00000005073 0.0 0.748 1.0 1.0 0.614 27.2 2.846E-4.0 0.056
CLIC1 ENSG00000096238 0.0 1.0 1.0 6.586E-5.0 0.723 24.75 3.098E-4.0 0.061
ERCC3 ENSG00000163161 0.0 0.795 0.329 0.0020 1.0 0.712 25.167 3.271E-4.0 0.065
ERCC3 ENSG00000163161 0.0 0.795 0.329 0.0020 1.0 0.712 25.167 3.271E-4.0 0.065
TLL2 ENSG00000095587 0.0 0.653 4.114E-4.0 0.274 0.8 32.8 3.58E-4.0 0.071
SYT4 ENSG00000132872 3.251E-13.0 1.0 0.151 0.712 29.25 3.724E-4.0 0.074
SYT4 ENSG00000132872 3.251E-13.0 1.0 0.151 0.712 29.25 3.724E-4.0 0.074
PIK4CB ENSG00000143393 0.0 0.729 0.329 1.0 0.733 29.4 3.849E-4.0 0.076
PKD2 ENSG00000118762 0.0 0.802 1.0 0.0020 0.64 0.643 26.0 3.947E-4.0 0.078
ENSG00000081026 0.0 0.373 1.136E-6.0 0.773 32.25 4.136E-4.0 0.082
ANKRD3 ENSG00000183421 0.0 0.329 0.0020 1.0 0.746 27.8 4.521E-4.0 0.09
F13A1 ENSG00000124491 0.0 0.959 1.0 1.086E-5.0 0.671 28.6 5.087E-4.0 0.101
BPAG1 ENSG00000151914 0.0 0.83 0.38 0.103 0.65 26.8 5.124E-4.0 0.101
KCNN3 ENSG00000143603 5.296E-12.0 0.891 1.0 6.586E-5.0 0.679 27.4 5.177E-4.0 0.103
GRIN2A|GRIN2B ENSG00000150086 9.992E-15.0 0.562 1.0 0.185 0.756 32.6 5.352E-4.0 0.106
SIM1 ENSG00000112246 9.992E 15.0 0.739 1.0 1.0 0.733 35.4 6.0E 4.0 0.119
ENSG00000174891 0.0 0.329 0.267 18.667 6.705E-4.0 0.133
ENSG00000089195 3.251E-13.0 0.329 1.0 0.738 32.0 6.89E-4.0 0.136
C14orf10 ENSG00000092020 0.0 0.329 0.274 19.0 6.906E-4.0 0.137
STX8 ENSG00000170310 9.992E-15.0 0.67 1.0 4.519E-4.0 0.789 35.2 7.243E-4.0 0.143
ENSG00000107671 0.0 1.0 2.442E-6.0 0.809 39.5 7.357E-4.0 0.146
MSH5 ENSG00000096474 0.0 0.373 1.0 0.673 29.5 7.428E-4.0 0.147
CRH ENSG00000147571 3.251E-13.0 0.934 1.0 3.386E-4.0 0.675 30.2 8.254E-4.0 0.163
MID1 ENSG00000101871 0.0 1.0 1.0 0.692 32.25 9.027E-4.0 0.179
ENSG00000184508 4.94E-14.0 1.0 0.026 23.667 9.838E-4.0 0.195
ENSG00000113460 0.0 0.329 1.0 24.0 9.912E-4.0 0.196
TGFB3 ENSG00000119699 0.0 0.79 1.0 1.0 0.658 27.167 0.0010 0.225
C1QR1 ENSG00000125810 0.0 0.925 4.114E-4.0 0.098 0.805 41.4 0.0010 0.226
NR4A2 ENSG00000153234 0.0 0.939 1.0 1.0 0.59 35.2 0.0010 0.226
PDGFC ENSG00000145431 0.0 1.0 7.502E-4.0 0.762 34.5 0.0010 0.235
PDGFC ENSG00000145431 0.0 1.0 7.502E-4.0 0.762 34.5 0.0010 0.235
NR3C2 ENSG00000151623 0.0 1.0 3.938E-4.0 0.769 35.5 0.0010 0.24
NFYA ENSG00000001167 0.0 0.894 1.0 1.0 0.628 33.8 0.0010 0.244
ENSG00000101898 0.0 1.0 0.142 0.757 36.75 0.0010 0.254
C8orf4 ENSG00000176907 0.0 0.329 0.762 29.333 0.0010 0.256
TM4SF13 ENSG00000106537 0.0 1.0 6.586E-5.0 0.847 46.0 0.0010 0.268
MMP3|MMP1 ENSG00000149968 0.0 0.981 1.0 0.084 0.71 36.6 0.0010 0.271
Filter
Remove
Export to flat file
Status
Saved data table to file lps_test.bin
Scoring entities in test set...
Scoring of biovec.ExpressionModel_atlas succesful.
Scoring of biovec.EnsemblEstModel succesful.
Scoring of biovec.KeggModel succesful.
Scoring of biovec.IprModel succesful.
Scoring of biovec.GOModel succesful.
Scoring of biovec.TextModel succesful.
Scoring Finished succesfully.
Saved data table to file export
http://www.esat.kuleuven.ac.be/endeavour

Endeavour



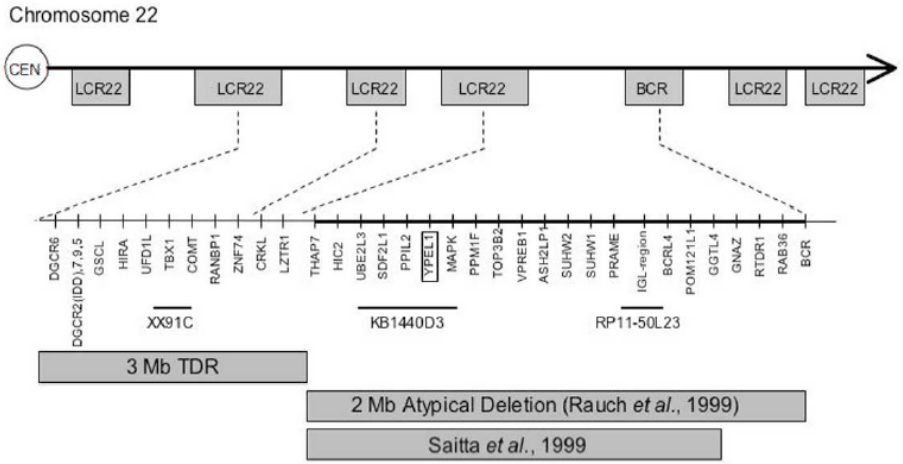
http://www.esat.kuleuven.ac.be/endeavour

Demo, manual, and exercise

- GENERAL: www.esat.kuleuven.be/endeavour
- DEMO: http://homes.esat.kuleuven.be/~bioiuser/endeavour/endeavour_demo.php
- MANUAL: http://homes.esat.kuleuven.be/~bioiuser/bioiwiki/index.php/Endeavour_Manual_Web_Server
- EXERCISE: http://homes.esat.kuleuven.be/~bioiuser/bioiwiki/index.php/Endeavour_handson

DiGeorge candidate

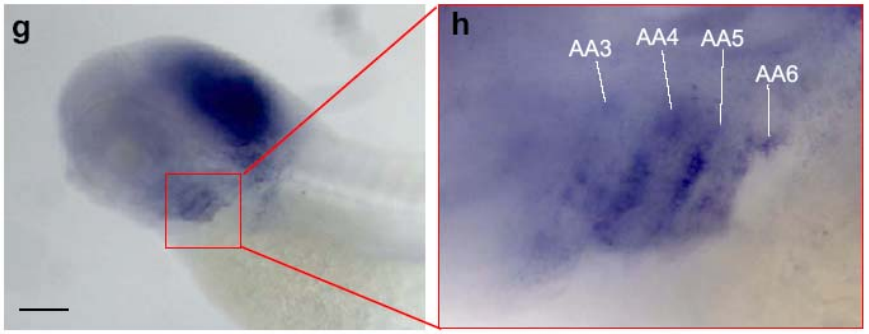
- *D. Lambrechts, S. Maity, P. Carmeliet, KUL Cardio*
- TBX1 critical gene in typical 3Mb aberration
- Atypical 2Mb deletion (58 candidates)



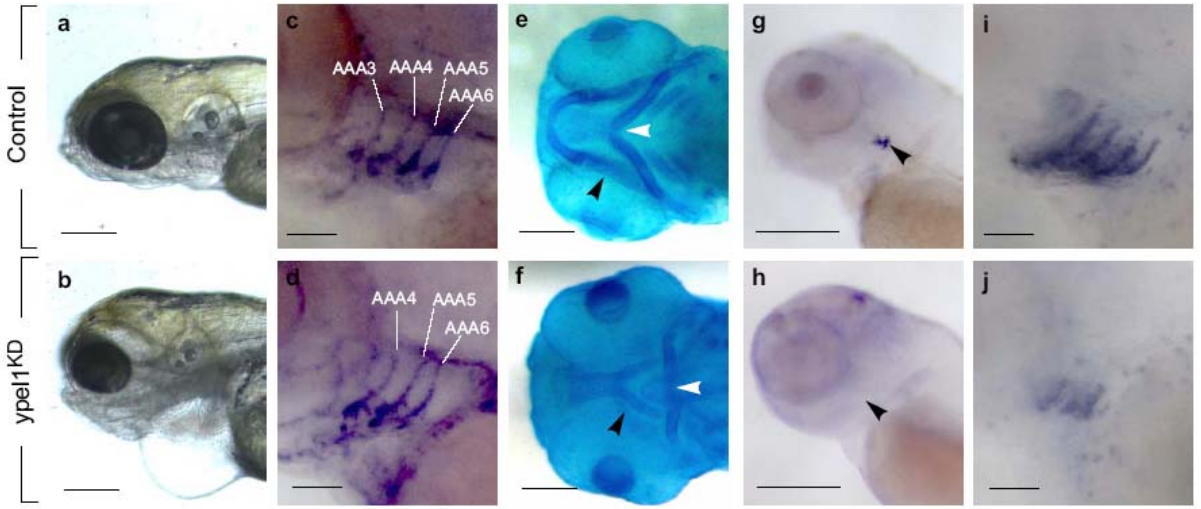
Training sets used to prioritize TBX1 or YPEL1	Rank assigned to YPEL1	Rank assigned to TBX1
DGS-related		
DGS (14)	1	1
Cardiovascular birth defects (14)	3	1
Cleft palate birth defects (9)	2	1
Neural crest genes (14)	1	2
Average rank	1.75 ± 0.48	1.25 ± 0.25

YPEL1

- YPEL1 is expressed in the pharyngeal arches during arch development



- YPEL1^{KD} zebrafish embryos exhibit typical DGS-like features



Congenital heart disease genes

- *B. Thienpont, K. Devriendt, J. Vermeesch, KUL CME*
- 60 patients without diagnosis
  - Congenital heart defect
  - & Chromosomal phenotype
    - 2nd major congenital anomaly
    - Or mental retardation/special education
    - Or > 3 minor anomalies
- Array Comparative Genomic Hybridization
  - 1 Mb resolution
- 11 anomalies detected
  - 5 deletions
  - 2 duplications
  - 3 complex rearrangements
  - 1 mosaic monosomy 7

Candidate regions

- 4 regions with known critical genes, 6 new regions, 80 candidate genes

aberration	***gene***
del(5)(q23)	?
del(5)(q35.1)	*NKX2.5*
del(5)(q35.2qter)	*NSD1*
del(14)(q22.1q23.1)	?
del(22)(q12.2)	?
dup(22)(q11)	*TBX1*
dup(19)(p13.12p13.11)	?
del(9)(q34.3qter),dup(20)(q13.33qter)	*NOTCH1, EHMT1*
del(13)(q31.1q31.3),dup(13)(q31.3q33.2),inv(13)	?
del(4)(q34.3q35.1),dup(4)(q34),inv(4)	?

Gene prioritization

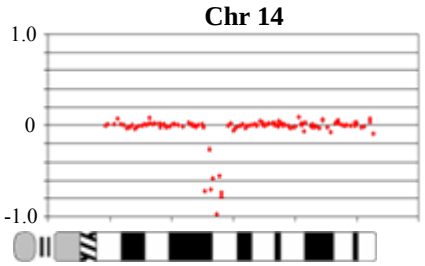
del(14)(q22.1q23.1) ?

	Expression data	KEGG pathways	Pubmed textmining	Protein domains	Cis-regulatory module	BLAST	Protein interactions	
1.	*CNIH*	*DACT1*	*BMP4*	*RTN1*	*BMP4*	*KIAA1344*	*BMP4*	*EXOC5*
2.		*DAAM1*	*PTGER2*	*DLG7*	*DAAM1*		*OTX2*	
3.	*KIAA1344*		*PTGDR*	*ARID4A*	*OTX2*	*ARID4A*	*WDHD1*	
4.	*CGRRF1*		*SOCS4*	*BMP4*	*KIAA0586*	*CDKN3*	*SOCS4*	*TIMM9*
5.	*DDHD1*	*STYX*	*DAAM1*	*PSMA3*		*SAMD4*	*DACT1*	*ERO1L*
6.	*ACTR10*	*KTN1*	*PSMC6*	*OTX2*		*STYX*	*SAMD4*	*PSMA3*
7.	*CDKN3*	*TIMM9*	*PSMA3*	*KTN1*	*SOCS4*	*FBXO34*		*BMP4*
8.	*RTN1*		*GNPNAT1*	*PSMC6*	*PSMC6*	*OTX2*	*RTN1*	*WDHD1*
9.	*FBXO34*		*TBPL2*	*WDHD1*	*WDHD1*	*PSMC6*	*KTN1*	*SOCS4*
10.		*CNIH*	*ERO1L*	*CNIH*	*KIAA1344*	*BMP4*	*FBXO34*	*KIAA1344*
11.	*PLEKHC1*		*GCH1*	*SOCS4*	*DACT1*	*KTN1*	*CDKN3*	*DACT1*
12.		*PSMA3*	*DDHD1*		*KTN1*	*PLEKHC1*	*DDHD1*	*OTX2*
13.		*PLEKHC1*	*WDHD1*	*STYX*	*ARID4A*			*DAAM1*
14.	*BMP4*	*SAMD4*	*KIAA1344*	*PLEKHC1*		*DACT1*		
15.	*GCH1*	*GMFB*	*DACT1*	*DAAM1*	*STYX*		*ERO1L*	
16.	*KTN1*	*DLG7*	*OTX2*	*FBXO34*	*SAMD4*	*GPR135*		
...		*ACTR10*		*PTGER2*	*DLG7*	*DAAM1*		*KTN1*
80.		...	...	...	...	...	...	...

BMP4
OTX2
DAAM1
WDHD1
KTN1
DACT1
ARID4A
SOCS4
SAMD4
KIAA1344
EXOC5
DLG7
PSMC6
STYX
...

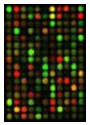
Congenital heart disorders

Congenital heart defect patient

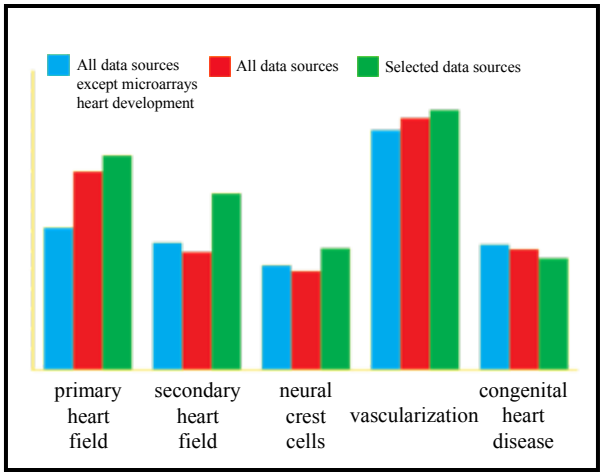


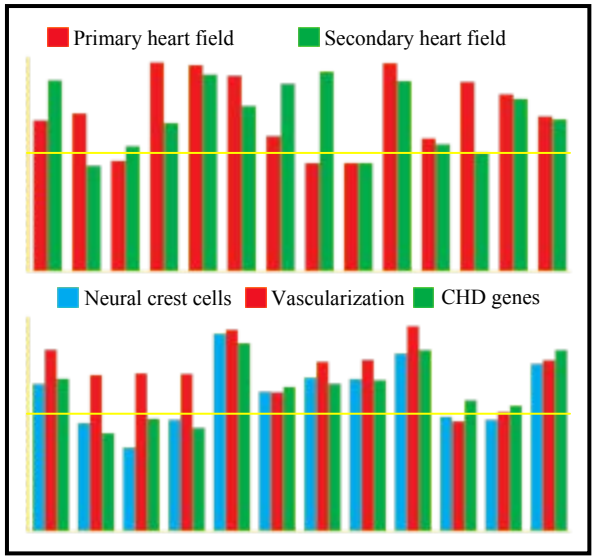
del(14q22.1-23.1)
56 candidate genes

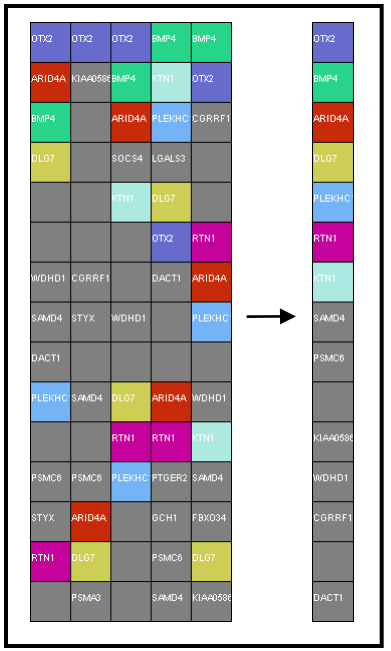
MA data embryonic heart development

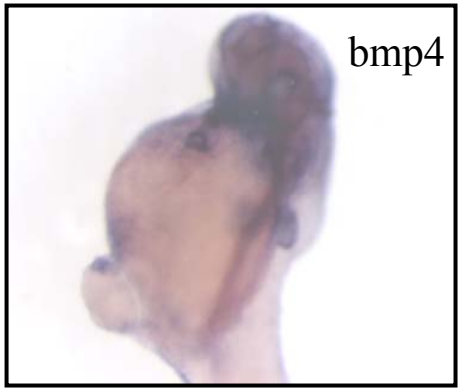


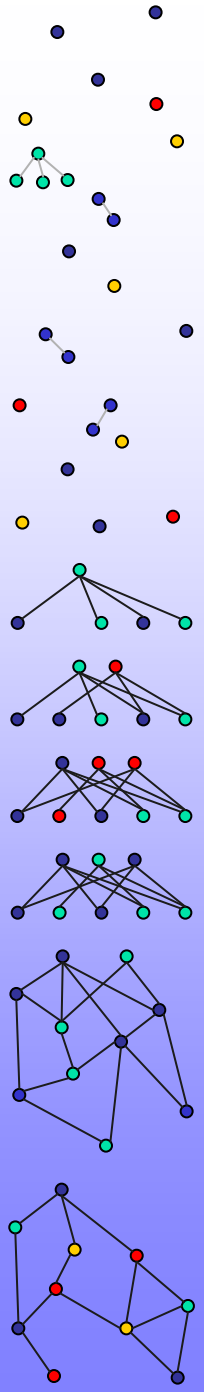
5 sets of training genes:
primary heart field
secondary heart field
neural crest cells
vascularization
congenital heart disease





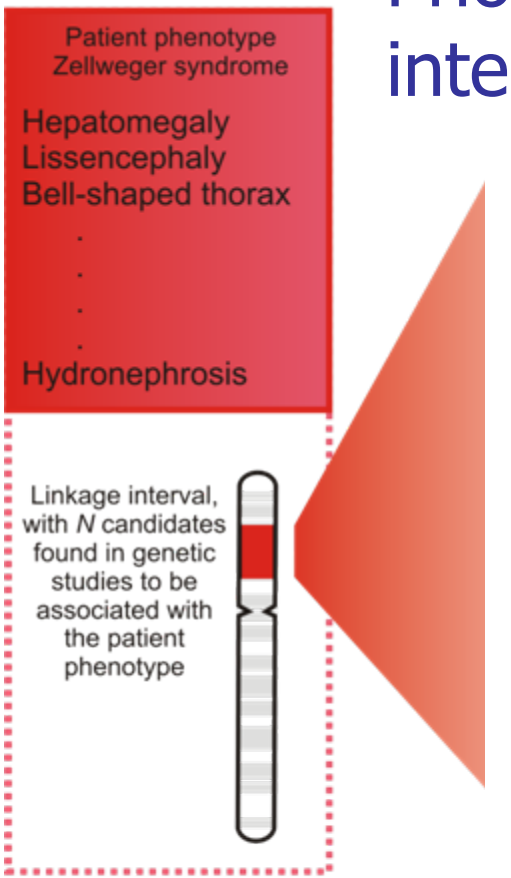






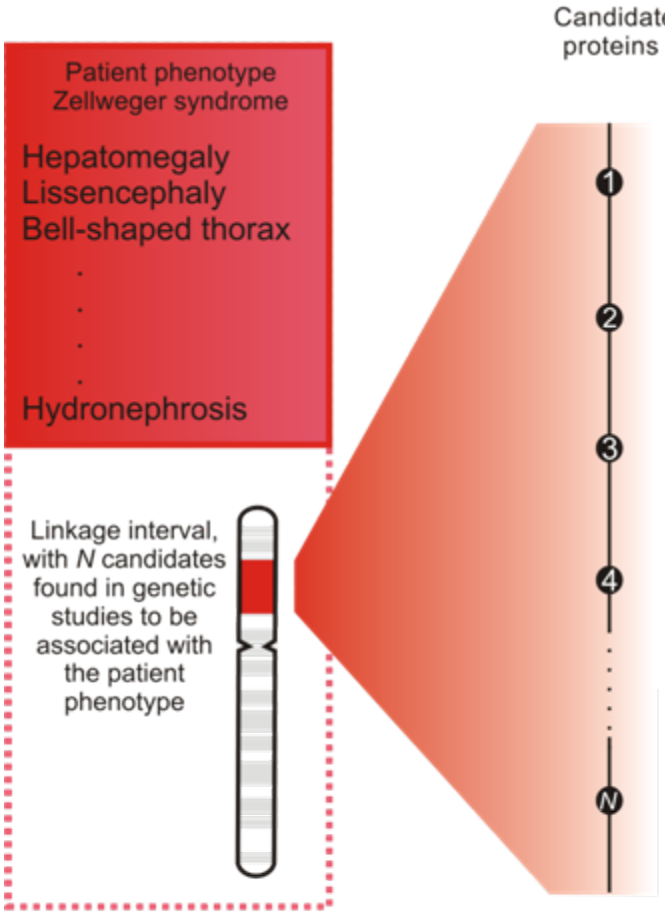
Prioritization by virtual pulldown

Prioritization by virtual protein-protein interaction pulldown and text mining

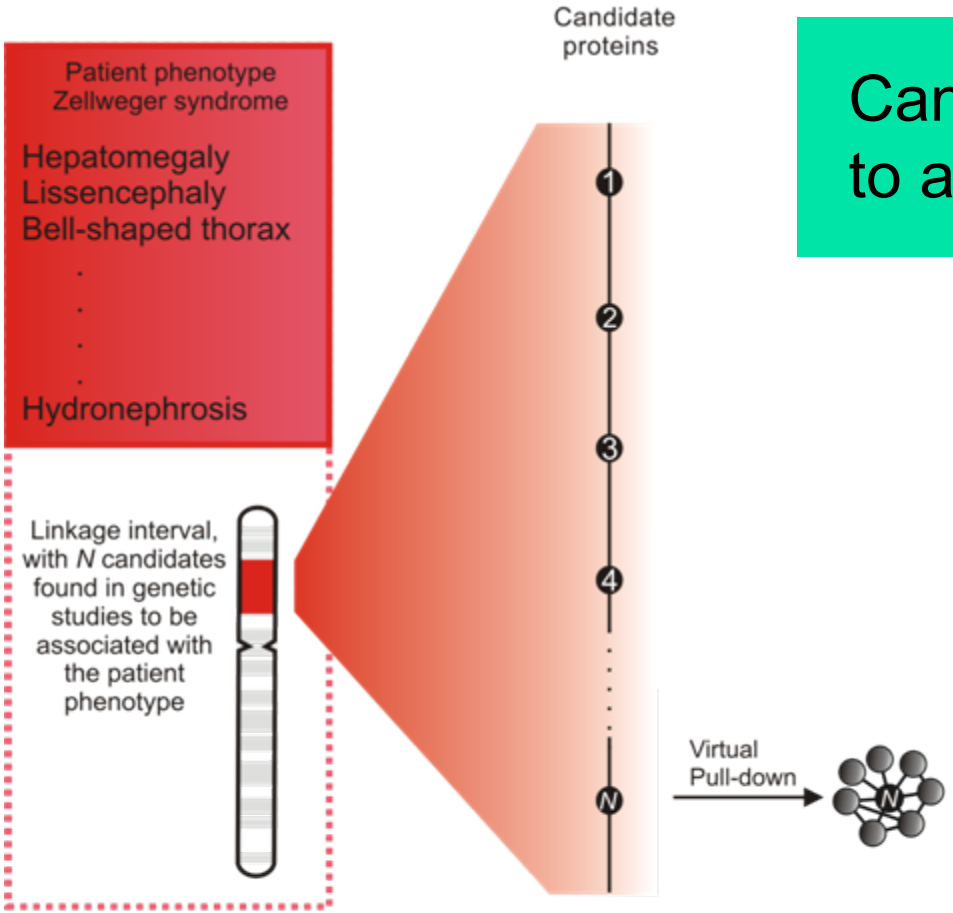


CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

- *Lage et al. Nature Biotech. March 2007*

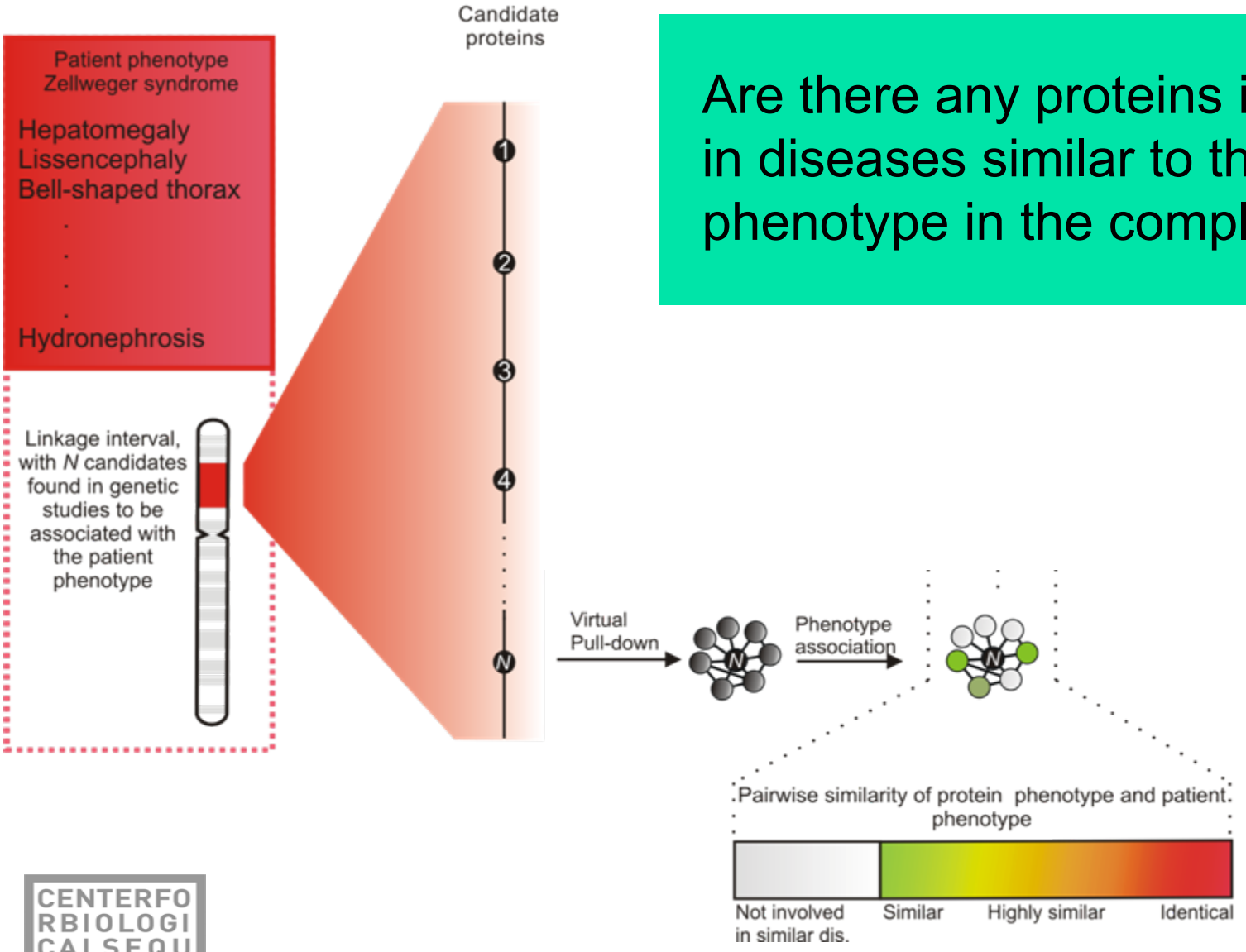


CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS **CBS**



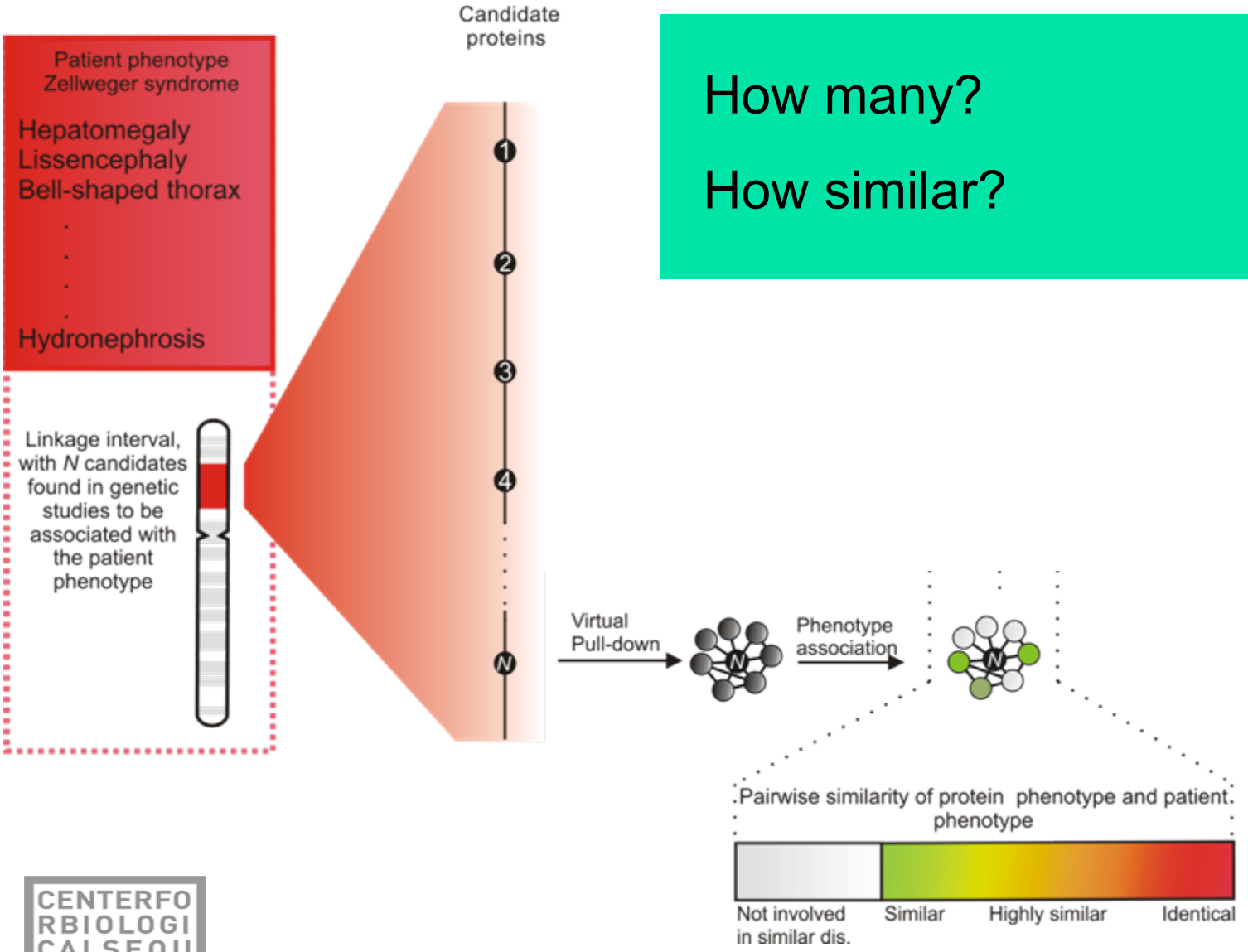
Can the candidate be assigned to a protein complex?

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS



Are there any proteins involved in diseases similar to the patient phenotype in the complex?

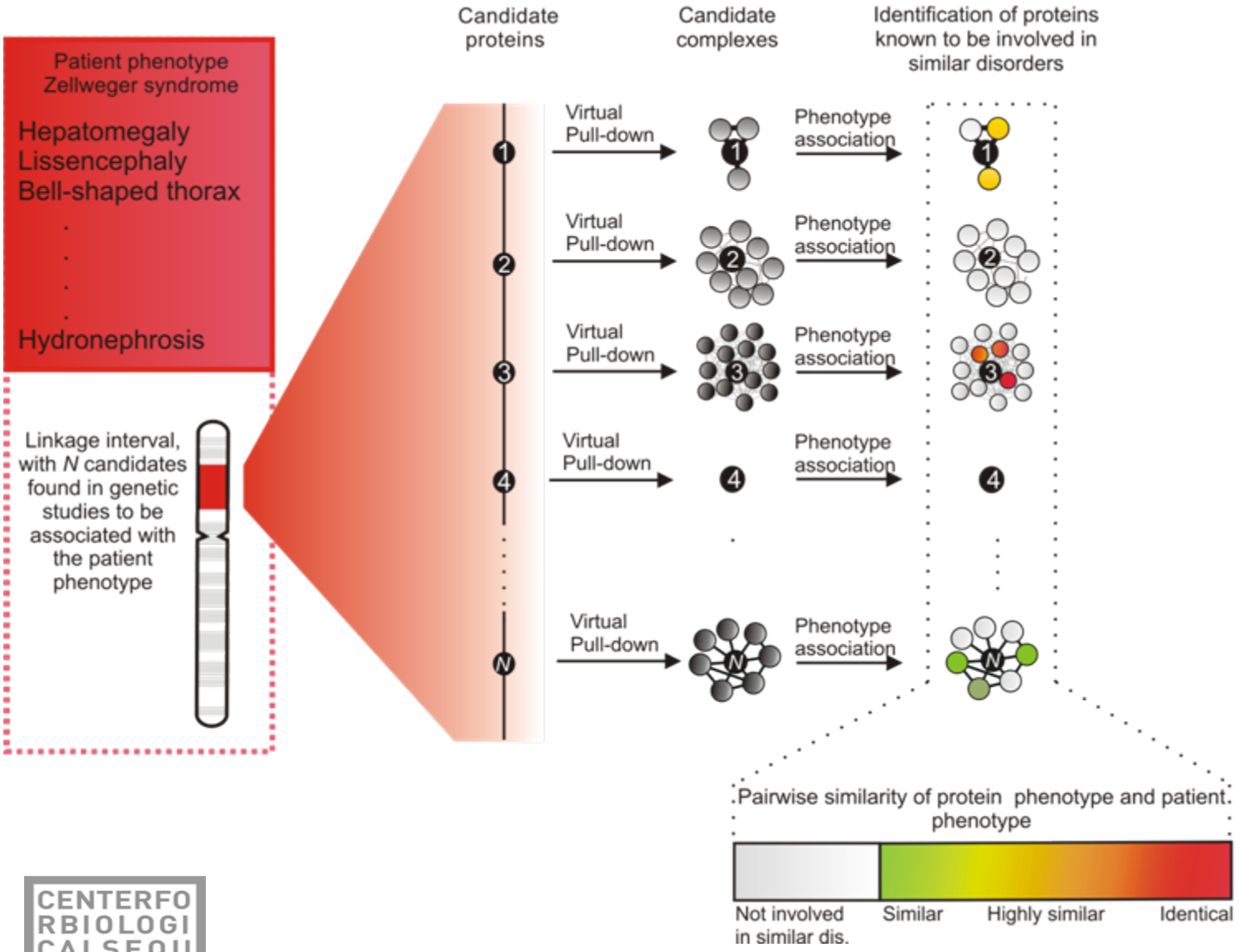
CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS



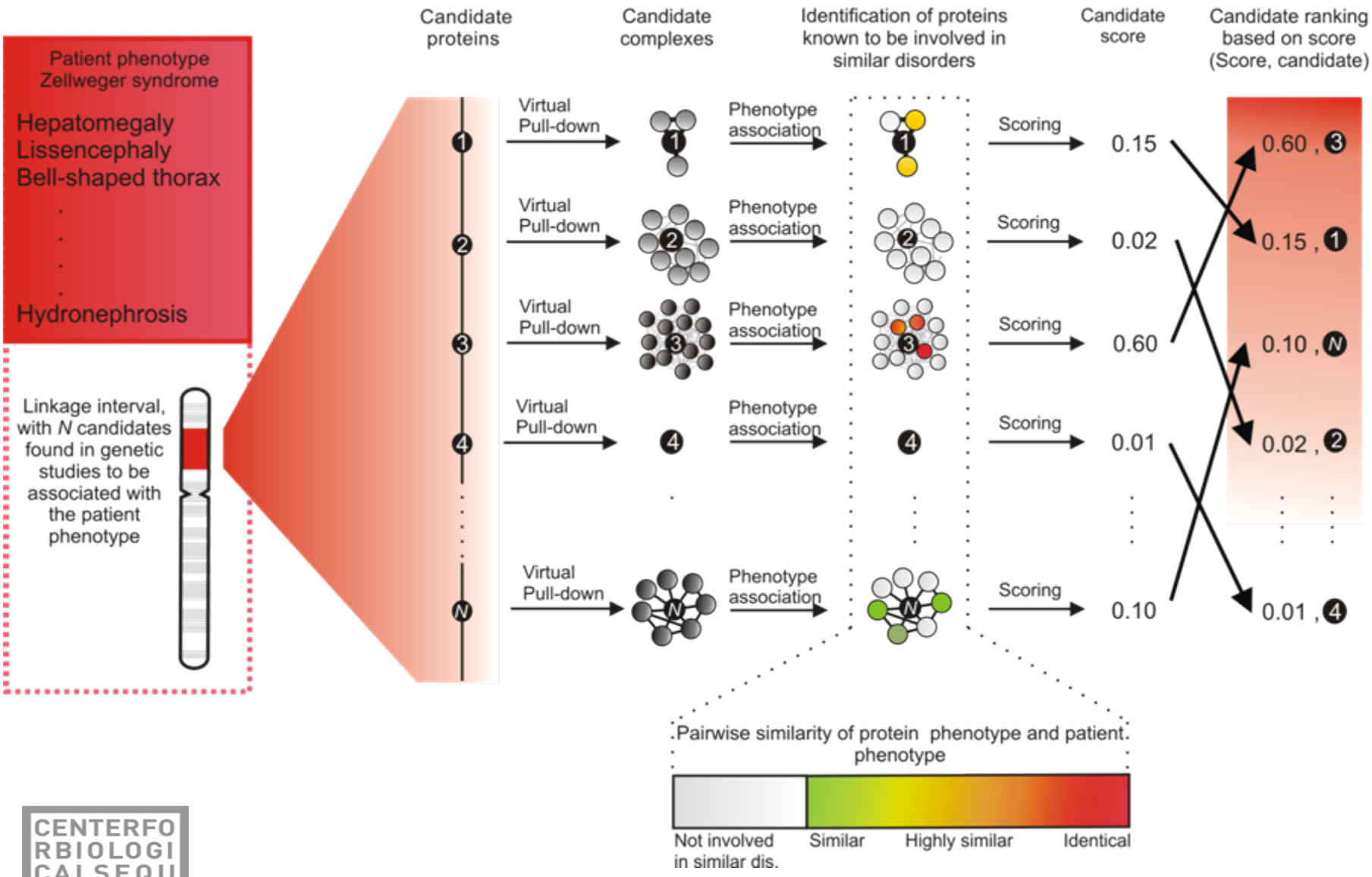
How many?

How similar?

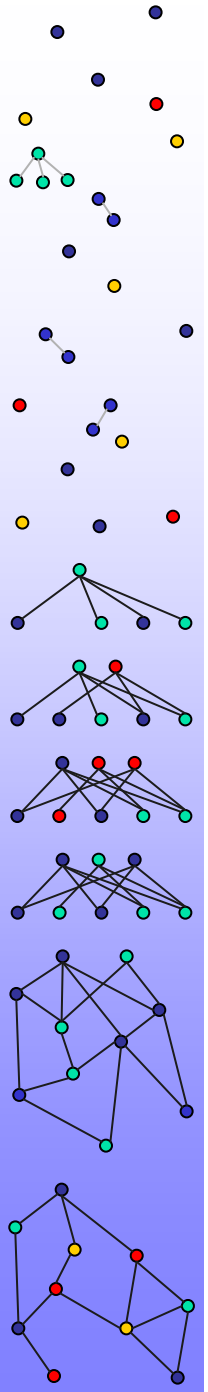
CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS



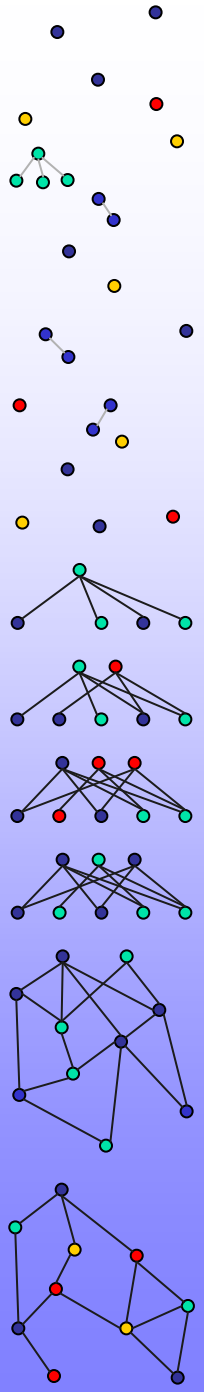
CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS



CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS



Putting it all together…



Integrating gene prioritization into daily biological work

- Gene prioritization is "interesting"...
  - Needs also to be integrated with "network" view of systems biology
- How can we bring it closer to the daily routine of wet bench?
  - Still left with a large number of candidates
  - Bioinformatics tool should not be trusted blindly
  - Need for reinterpretation and "ownership"
- "Wikis" can be used as "collaborative electronic notebooks"
  - Same technology as Wikipedia
  - Addition of database back-end for structured information
  - http://homes.esat.kuleuven.be/~rbarriot/genewiki/index.php/CHD:Home
  - http://homes.esat.kuleuven.be/~rbarriot/genewiki/index.php/CHDGene:YM70

CHD:AEPC Nomenclature

CHD
Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography

Wiki Syntax
Help
Basic formatting
Links
Images
Tables
Categories
Templates

Toolbox
What links here
Related changes
Upload file
Special pages
Printable version
Permanent link

AEPC Nomenclature [hide]

Diagnostic congenital and generic cardiac codes

01.01.00 Normal heart

01.03.10 Normal atrial arrangement (situs), AV & VA connections

01.03.00 Usual atrial arrangement (atrial situs solitus)

01.05.00 Concordant VA connections

10.12.01 Innocent murmur

Abnormalities of position and connection of heart

02.01.09 Position-orientation of heart abnormal

02.01.02 Dextrocardia: heart predominantly in R hemithorax

03.01.09 Position or morphology of thoraco-abdominal organs abnormal

01.03.06 Abnormal atrial arrangement

03.01.03 Total mirror imagery (atrial situs inversus)

03.01.04 Right isomerism (\'asplenia\')

03.01.05 Left isomerism (\'polysplenia\')

01.03.09 AV and/or VA connections abnormal

01.01.14 Double inlet ventricle

01.04.03 Double inlet RV

01.04.04 Double inlet LV

06.01.01 Tricuspid atresia

06.02.01 Mitral atresia

02.03.05 Solitary ventricle of indeterminate morphology

01.05.01 Discordant VA connections (TGA)

01.01.02 Complete transposition of great arteries (IVS)

CHD:Genes

CHD

Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography

Wiki Syntax

Help
Basic formatting
Links
Images
Tables
Categories
Templates

Toolbox

What links here
Related changes
Upload file
Special pages
Printable version
Permanent link

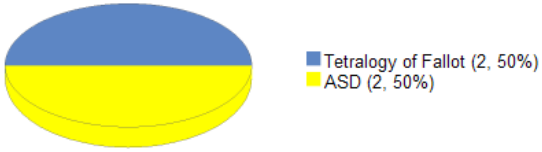
Genes currently associated to CHDs

- ATRX - 1 association(s)
- BCOR - 1 association(s)
- BRAF - 1 association(s)
- CBP/CREBBP - 2 association(s)
- CCN1/PPP3CA - 1 association(s)
- CFC1/CRYPTIC - 7 association(s)
- CHD7 - 4 association(s)
- CITED2 - 5 association(s)
- COL2A1 - 1 association(s)
- CRELD1 - 2 association(s)
- EHMT1 - 1 association(s)
- ELN - 9 association(s)
- EVC - 3 association(s)
- FBN1 - 3 association(s)
- FLNA - 1 association(s)
- FOG2/ZFPM2 - 2 association(s)
- GATA4 - 5 association(s)
- GJA1 - 1 association(s)
- GPC3 - 1 association(s)
- HAND1 - 2 association(s)
- HAND2 - 1 association(s)
- HEY2 - 3 association(s)
- HRAS - 1 association(s)
- JAG1 - 8 association(s)
- KRAS - 1 association(s)
- LBR - 1 association(s)
- MAP2K1/MEK1 - 1 association(s)
- MAP2K2/MEK2 - 1 association(s)
- MGP - 1 association(s)
- MID1 - 2 association(s)
- MYH11 - 2 association(s)
- MYH6 - 2 association(s)
- NF1 - 1 association(s)
- NKX2-5/NKX2.5 - 7 association(s)
- NOTCH1 - 2 association(s)
- NOTCH2 - 1 association(s)

CHDGene:ENSG00000136574

GATA4

Non syndromic associated CHDs overview



Synopsis

Encodes an essential transcription factor for cardiac development. Mutations in this gene have been found sporadically in families with congenital heart defects.

GATA4 is located on human chromosome 8p23.1 in a region flanked by low copy repeats (LCRs). Non-allelic homologuous recombination between these LCRs can result in deletion of 8p23.1. This imbalance is associated with congenital heart defects, microcephaly, intrauterine growth retardation, mental retardation and a characteristic behavior(Devriendt et al).

Developmental biology

Essential and dosage-dependent regulation of cardiac morphogenesis. Reduction of protein level below certain threshold (30-50%) results in reduced cardiomyocyte replication, myocardial hypoplasia, and endocardial cushion defects (Pu et al.).

External references for *GATA4*

- ensembl: ENSG00000136574
- OMIM: 600576
- search miRBase for GATA4
- Expression patterns from 4DXPress: GATA4

Known phenotypes for ***GATA4***

Non-syndromic

ASD

- **Support:** confirmed: 2 or more independent reports; >1% incidence
- **References:** PMID:15810002 (population study with screening of similar CHD patients and normal controls) PMID:15689439 (population study with screening of similar CHD patients and normal controls) PMID:12845333 (population study with screening of similar CHD patients and normal controls)
- **Inheritence:** Nonsyndromic dominant atrial septal defect type 2 (OMIM:607941).
- **Incidence:** 16 families (9/16 isolated nonsyndromic ASD) with multiple affected members and 13 unrelated sporadic individuals (9/13 isolated nonsyndromic ASD): mutations in NKX2.5 in 3 of the 29 index patients; mutations in GATA4 in 2/29 (PMID:15689439); 16 families (12/16 isolated nonsyndromic ASD): mutations in NKX2.5 in 3/16 probands and mutations in GATA4 in 2/16 probands (PMID:15810002); 1 family (16 individuals with ASD; 9/16 isolated nonsyndromic ASD; 1/16 AVSD/ASD/PS, 3/16 VSD/ASD, 3/16 ASD/PS) with GATA4 mutation (PMID:12845333).
- **miRNAs binding site:**

CHD
Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography
Wiki Syntax
Help
Basic formatting
Links
Images
Tables
Categories
Templates
Toolbox
What links here
Related changes
Upload file
Special pages
Printable version
Permanent link

Add another phenotype.

Syndromic

none

Add another phenotype.

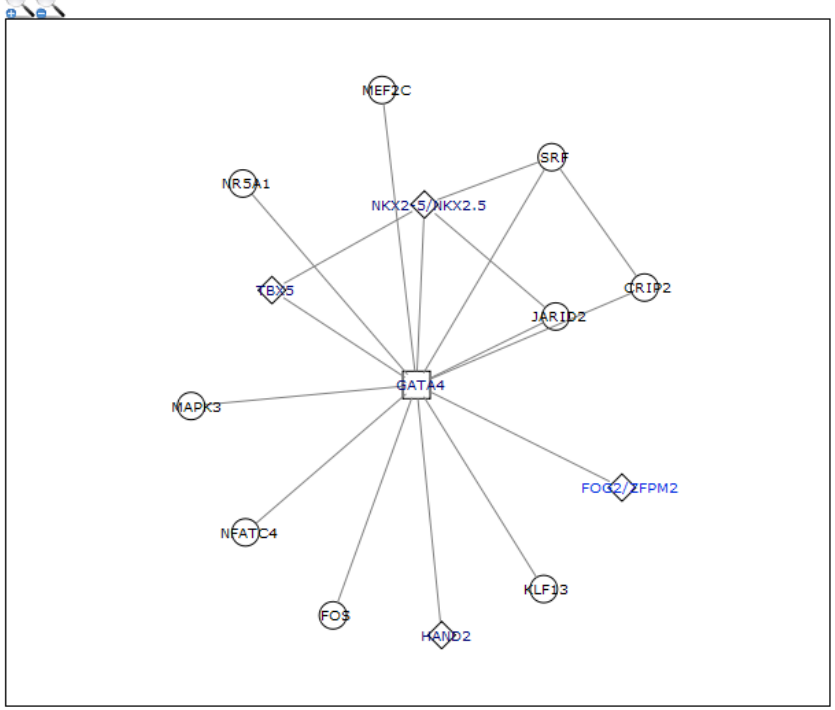
CGHGate case reports for *GATA4*

none

Decipher data for *GATA4*

Patient Report 128 **decipher:novel:del:hard** note
Patient Report 589 **decipher:novel:del:hard** note
Patient Report 879 **decipher:novel:del:hard** note
Patient Report 1351 **decipher:novel:del:hard** note
Syndrome Report 39 **decipher:known:del** note 8p23.1 deletion syndrome

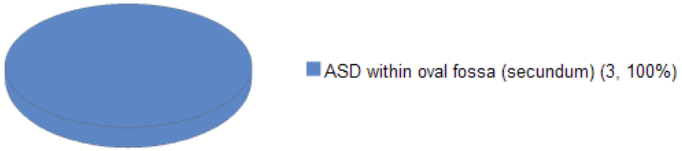
Protein interaction partners



CHDGene:ENSG00000183072

NKX2-5/NKX2.5

Non syndromic associated CHDs overview

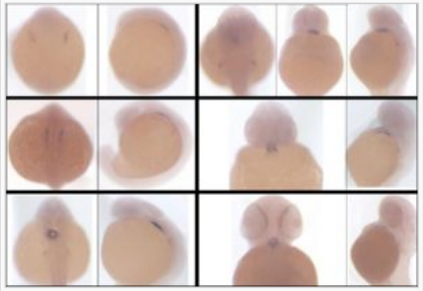


Synopsys

NKX2.5 is a homeobox-containing gene expressed in the first and second heart field during heart development. Mutations in NKX2.5 are found in patients with sporadic or familial atrial septal defects (with or without atrioventricular block) (OMIM:108900) or Tetralogy of Fallot (OMIM:187500).

developmental biology

cardiac lineage specification.



pattern of NKX2.5 mRNA expression during zebrafish development

cellular function

transcription factor

External references for *NKX2-5/NKX2.5*

- ensembl: ENSG00000183072
- OMIM: 600584
- search miRBase for NKX2-5/NKX2.5
- Expression patterns from 4DXPress: NKX2-5/NKX2.5

Known phenotypes for ***NKX2-5/NKX2.5***

Non-syndromic

ASD

- **Support:** confirmed: 2 or more independent reports; >1% incidence
- **References:** PMID:9651244 (population study with screening of similar CHD patients and normal controls) PMID:15810002 (population study with screening of similar CHD patients and normal controls) PMID:15689439 (population study with screening of similar CHD patients and normal controls) PMID:14607454 (population study with screening of similar CHD patients and normal controls)
- **Inheritence:** Dominant. Atrial septal defect +/- AV block (OMIM:108900)
- **Incidence:** 16 families (9/16 isolated nonsyndromic ASD) with multiple affected members and 13 unrelated sporadic individuals (9/13 isolated nonsyndromic ASD): mutations in NKX2.5 in 3 of the 29 index patients; mutations in GATA4 in 2/29 (PMID:15689439); 16 families (12/16 isolated nonsyndromic ASD): mutations in NKX2.5 in 3/16 probands and mutations in GATA4 in 2/16 probands (PMID:15810002); 3 families (33 affected patients with isolated nonsyndromic ASD): 3 mutations (PMID:9651244); 3 out of 71 (4%) index patients with secundum ASD (PMID:14607454)
- **miRNAs binding site:**

CHD
Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography

Wiki Syntax
Help
Basic formatting
Links
Images
Tables
Categories
Templates

Toolbox
What links here
Related changes
Upload file
Special pages
Printable version
Permanent link

Editing CHDGene:ENSG00000183072

CHD
Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography

Wiki Syntax
Help
Basic formatting
Links
Images
Tables
Categories
Templates

Toolbox
What links here
Related changes
Upload file
Special pages

```
===Synopsys===
NKX2.5 is a homeobox-containing gene expressed in the first and second heart field during heart development. Mutations in NKX2.5 are found in patients with sporadic or familial atrial septal defects (with or without atrioventricular block) ([[OMIM:108900]]) or Tetralogy of Fallot ([[OMIM:187500]]).


===developmental biology===
cardiac lineage specification.

[[Image:Nkx2.5.JPG|thumb|300px|pattern of NKX2.5 mRNA expression during zebrafish development]]


===cellular function===
transcription factor
```

Please note that all contributions to CHDWiki may be edited, altered, or removed by other contributors. If you don't want your writing to be edited mercilessly, then don't submit it here.
You are also promising us that you wrote this yourself, or copied it from a public domain or similar free resource (see Project:Copyrights for details). **DO NOT SUBMIT COPYRIGHTED WORK WITHOUT PERMISSION!**

Summary:

This is a minor edit Watch this page

Save page Show preview Show changes Cancel | Editing help (opens in new window)

CHD:Prioritize

CHD

Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography

Wiki Syntax

Help
Basic formatting
Links
Images
Tables
Categories
Templates

Toolbox

What links here
Related changes
Upload file
Special pages
Printable version
Permanent link

Models | Training genes | Candidates

- [x] Annotation - EnsemblEst
- [x] Annotation - GeneOntology
- [x] Annotation - Interpro
- [x] Annotation - Kegg
- [x] Annotation - Swissprot
- [x] Blast
- [x] Expression - SonEtAl
- [ ] Expression - SuEtAl
- [x] Interaction - String
- [x] Precalculated - Ouzounis
- [ ] Precalculated - Prospectr

Prioritize!

Copy/paste those values for an example

- training

```
ENSG00000163217 ENSG00000164107 ENSG00000081189 ENSG00000113196 ENSG00000183072 ENSG00000164532
ENSG00000169946 ENSG00000134817 ENSG00000089225 ENSG00000141052 ENSG00000171388
```

- candidates

```
ENSG00000136574
ENSG00000158555  ENSG00000149257  ENSG00000171533  ENSG00000166391
ENSG00000062282  ENSG00000198382  ENSG00000085741  ENSG00000137492
ENSG00000179240  ENSG00000158636  ENSG00000137507  ENSG00000204529
ENSG00000182704  ENSG00000078124  ENSG00000198488  ENSG00000149260
ENSG00000137474  ENSG00000178795  ENSG00000149269  ENSG00000198407
ENSG00000197650  ENSG00000178301  ENSG00000074201  ENSG00000048649
ENSG00000087884  ENSG00000149262  ENSG00000201756  ENSG00000200256
ENSG00000210530  ENSG00000199563  ENSG00000199362  ENSG00000206816
ENSG00000200853  ENSG00000210489  ENSG00000210462  ENSG00000212030
```

CHD:Prioritize

CHD

- Home
- Browse by CHD
- Browse by gene
- News
- Map
- Add gene-CHD association
- Prioritize
- Protein interactions
- Recent changes
- Access and registration
- Bibliography

Wiki Syntax

- Help
- Basic formatting
- Links
- Images
- Tables
- Categories
- Templates

Toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

Models | Training genes | Candidates

Used training genes identifiers appear hereafter

```
ENSG00000163217 ENSG00000164107
ENSG00000081189 ENSG00000113196
ENSG00000183072 ENSG00000164532
ENSG00000169946 ENSG00000134817
ENSG00000089225 ENSG00000141052
ENSG00000171388
```

Prioritize!

Copy/paste those values for an example

- training

```
ENSG00000163217 ENSG00000164107 ENSG00000081189 ENSG00000113196 ENSG00000183072 ENSG00000164532
ENSG00000169946 ENSG00000134817 ENSG00000089225 ENSG00000141052 ENSG00000171388
```

- candidates

```
ENSG00000136574
ENSG00000158555  ENSG00000149257  ENSG00000171533  ENSG00000166391
ENSG00000062282  ENSG00000198382  ENSG00000085741  ENSG00000137492
ENSG00000179240  ENSG00000158636  ENSG00000137507  ENSG00000204529
ENSG00000182704  ENSG00000078124  ENSG00000198488  ENSG00000149260
ENSG00000137474  ENSG00000178795  ENSG00000149269  ENSG00000198407
ENSG00000197650  ENSG00000178301  ENSG00000074201  ENSG00000048649
ENSG00000087884  ENSG00000149262  ENSG00000201756  ENSG00000200256
ENSG00000210530  ENSG00000199563  ENSG00000199362  ENSG00000206816
ENSG00000200853  ENSG00000210489  ENSG00000210462  ENSG00000212030
```

CHD:Prioritize

CHD
Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography

Wiki Syntax
Help
Basic formatting
Links
Images
Tables
Categories
Templates

Toolbox
What links here
Related changes
Upload file
Special pages
Printable version
Permanent link

Models | Training genes | Candidates

Candidate genes to prioritize

```
ENSG00000136574
ENSG00000158555  ENSG00000149257
ENSG00000171533  ENSG00000166391
ENSG00000062282  ENSG00000198382
ENSG00000085741  ENSG00000137492
ENSG00000179240  ENSG00000158636
ENSG00000137507  ENSG00000204529
ENSG00000182704  ENSG00000078124
ENSG00000198488  ENSG00000149260
ENSG00000137474  ENSG00000178795
ENSG00000149269  ENSG00000198407

ENSG00000197650  ENSG00000178301
ENSG00000074201  ENSG00000048649
ENSG00000087884  ENSG00000149262
ENSG00000201756  ENSG00000200256
ENSG00000210530  ENSG00000199563
ENSG00000199362  ENSG00000206816
ENSG00000200853  ENSG00000210489
ENSG00000210462  ENSG00000212030
```

Prioritize!

Copy/paste those values for an example

- training

```
ENSG00000163217 ENSG00000164107 ENSG00000081189 ENSG00000113196 ENSG00000183072 ENSG00000164532
ENSG00000169946 ENSG00000134817 ENSG00000089225 ENSG00000141052 ENSG00000171388
```

- candidates

```
ENSG00000136574
ENSG00000158555  ENSG00000149257  ENSG00000171533  ENSG00000166391
ENSG00000062282  ENSG00000198382  ENSG00000085741  ENSG00000137492
ENSG00000179240  ENSG00000158636  ENSG00000137507  ENSG00000204529
ENSG00000182704  ENSG00000078124  ENSG00000198488  ENSG00000149260
ENSG00000137474  ENSG00000178795  ENSG00000149269  ENSG00000198407
ENSG00000197650  ENSG00000178301  ENSG00000074201  ENSG00000048649
ENSG00000087884  ENSG00000149262  ENSG00000201756  ENSG00000200256
ENSG00000210530  ENSG00000199563  ENSG00000199362  ENSG00000206816
ENSG00000200853  ENSG00000210489  ENSG00000210462  ENSG00000212030
```

CHD:Prioritize

Models | Training genes | Candidates | Results | Sprint plot | Network | XML

Global prioritization	Annotation - EnsemblEst	Annotation - GeneOntology	Annotation - Interpro	Annotation - Kegg	Annotation - Swissprot	Blast	Expression - SonEtAl	Interaction - String	Pr
GATA4	WNT11	GATA4	GATA4	PAK1	GATA4	PAK1	CAPN5	TSKU	
UVRAG	PAK1	RSF1	RSF1	WNT11	RSF1	C11orf30	UVRAG	GATA4	
C11orf30	CAPN5	C11orf30	C11orf30	B3GNT6	C11orf30	ENSG00000179240	PHCA	UVRAG	
PAK1	AQP11	MYO7A	MYO7A	PHCA	PRKRIR	UVRAG	SERPINH1	MOGAT2	
RSF1	UVRAG	PRKRIR	PRKRIR	GATA4	INTS4	AQP11	GDPD5	WNT11	
WNT11	TSKU	PAK1	PAK1	RSF1	CLNS1A	GATA4	GATA4	ENSG00000198407	
ENSG00000179240	ENSG00000179240	WNT11	WNT11	C11orf30	WNT11	RSF1	CLNS1A	C11orf30	
MYO7A	RSF1	ENSG00000179240	INTS4	MYO7A	MYO7A	PRKRIR	PAK1	PAK1	
TSKU	GATA4	INTS4	CLNS1A	PRKRIR	PAK1	INTS4	WNT11	PRKRIR	
CAPN5	GDPD5	CLNS1A	TSKU	INTS4	B3GNT6	CLNS1A	MAP6	DGAT2	
CLNS1A	PRKRIR	UVRAG	LRRC32	CLNS1A	PHCA	WNT11	LRRC32	MYO7A	
PRKRIR	B3GNT6	TSKU	B3GNT6	TSKU	TSKU	MYO7A	TSKU	CLNS1A	
AQP11	DGAT2	LRRC32	MOGAT2	LRRC32	LRRC32	B3GNT6	RSF1	PHCA	
SERPINH1	C11orf30	B3GNT6	SERPINH1	MOGAT2	MOGAT2	PHCA	PRKRIR	LRRC32	
MOGAT2	PHCA	MOGAT2	CAPN5	SERPINH1	SERPINH1	TSKU	C11orf30	CAPN5	
ENSG00000198407	MYO7A	SERPINH1	DGAT2	CAPN5	CAPN5	LRRC32	DGAT2	SERPINH1	
GDPD5	GDPD4	CAPN5	PHCA	DGAT2	DGAT2	MOGAT2	INTS4	GDPD5	
MAP6	C11orf67	DGAT2	GDPD5	GDPD5	GDPD5	SERPINH1	ENSG00000179240	MAP6	
INTS4	INTS4	PHCA	GDPD4	GDPD4	GDPD4	CAPN5	AQP11	RSF1	
C11orf67	MOGAT2	GDPD5	AQP11	AQP11	AQP11	DGAT2	MYO7A	INTS4	ENS
PHCA	CLNS1A	GDPD4	C11orf67	C11orf67	C11orf67	GDPD5	B3GNT6	ENSG00000179240	
LRRC32	ENSG00000198407	AQP11	ENSG00000198407	ENSG00000198407	UVRAG	GDPD4	MOGAT2	AQP11	
B3GNT6	MAP6	C11orf67	MAP6	MAP6	ENSG00000198407	C11orf67	GDPD4	B3GNT6	ENS
GDPD4	LRRC32	ENSG00000198407	ENSG00000179240	ENSG00000179240	MAP6	ENSG00000198407	C11orf67	GDPD4	
DGAT2	SERPINH1	MAP6	UVRAG	UVRAG	ENSG00000179240	MAP6	ENSG00000198407	C11orf67	
ENSG00000199563	ENSG00000199563	ENSG00000199563	ENSG00000199563	ENSG00000199563	ENSG00000199563	ENSG00000199563	ENSG00000199563	ENSG00000199563	ENS
GUCY2E	GUCY2E	GUCY2E	GUCY2E	GUCY2E	GUCY2E	GUCY2E	GUCY2E	GUCY2E	
ENSG00000201756	ENSG00000201756	ENSG00000201756	ENSG00000201756	ENSG00000201756	ENSG00000201756	ENSG00000201756	ENSG00000201756	ENSG00000201756	ENS
ENSG00000197650	ENSG00000197650	ENSG00000197650	ENSG00000197650	ENSG00000197650	ENSG00000197650	ENSG00000197650	ENSG00000197650	ENSG00000197650	ENS
ENSG00000212030	ENSG00000212030	ENSG00000212030	ENSG00000212030	ENSG00000212030	ENSG00000212030	ENSG00000212030	ENSG00000212030	ENSG00000212030	ENS

CHD
Home
Browse by CHD
Browse by gene
News
Map
Add gene-CHD association
Prioritize
Protein interactions
Recent changes
Access and registration
Bibliography

Wiki Syntax
Help
Basic formatting
Links
Images
Tables
Categories
Templates

Toolbox
What links here
Related changes
Upload file
Special pages
Printable version
Permanent link

CHD:Protein Interactions

CHD

- Home
- Browse by CHD
- Browse by gene
- News
- Map
- Add gene-CHD association
- Prioritize
- Protein interactions
- Recent changes
- Access and registration
- Bibliography

Wiki Syntax

- Help
- Basic formatting
- Links
- Images
- Tables
- Categories
- Templates

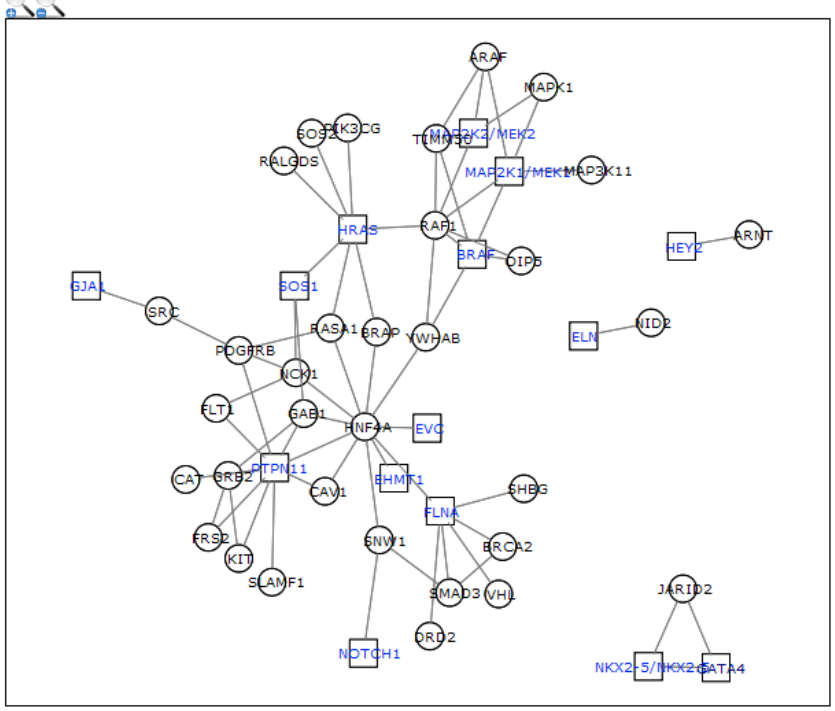
Toolbox

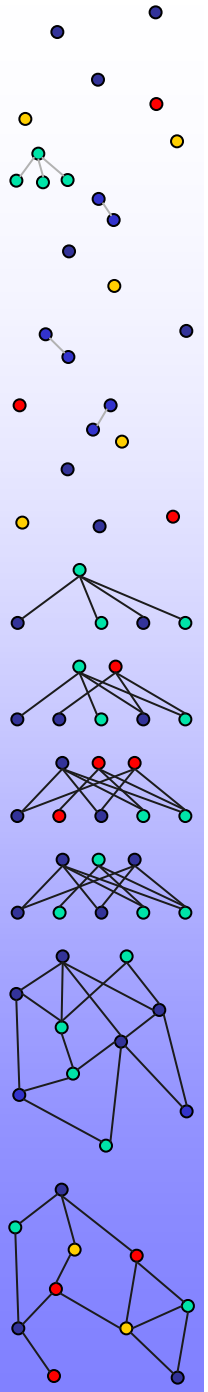
- What links here
- Upload file
- Special pages

Source bind

Select all associated genes having a minimum support of likely: 2 or more patients (with CHD and a mutation in the candidate gene)

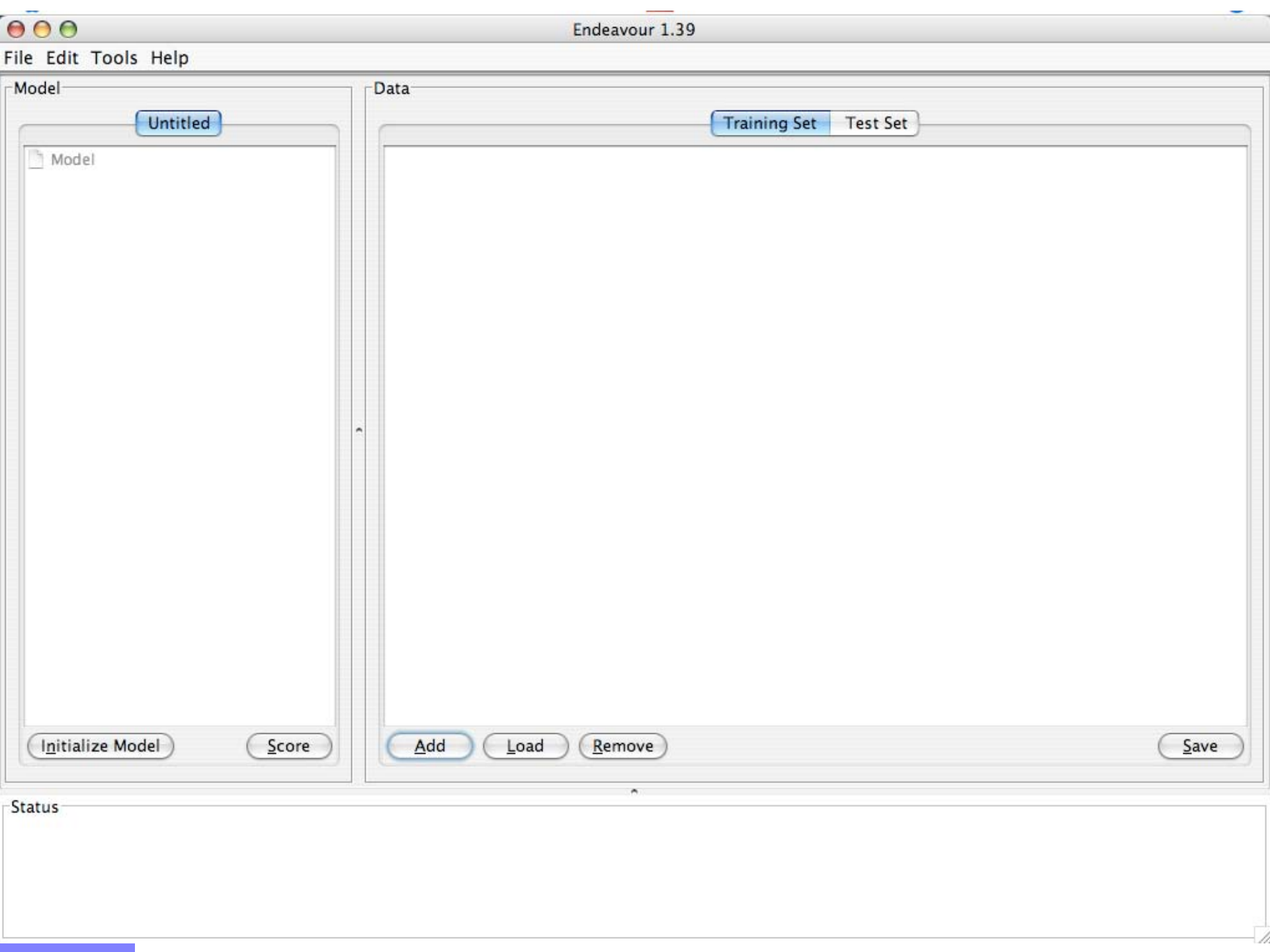
Show

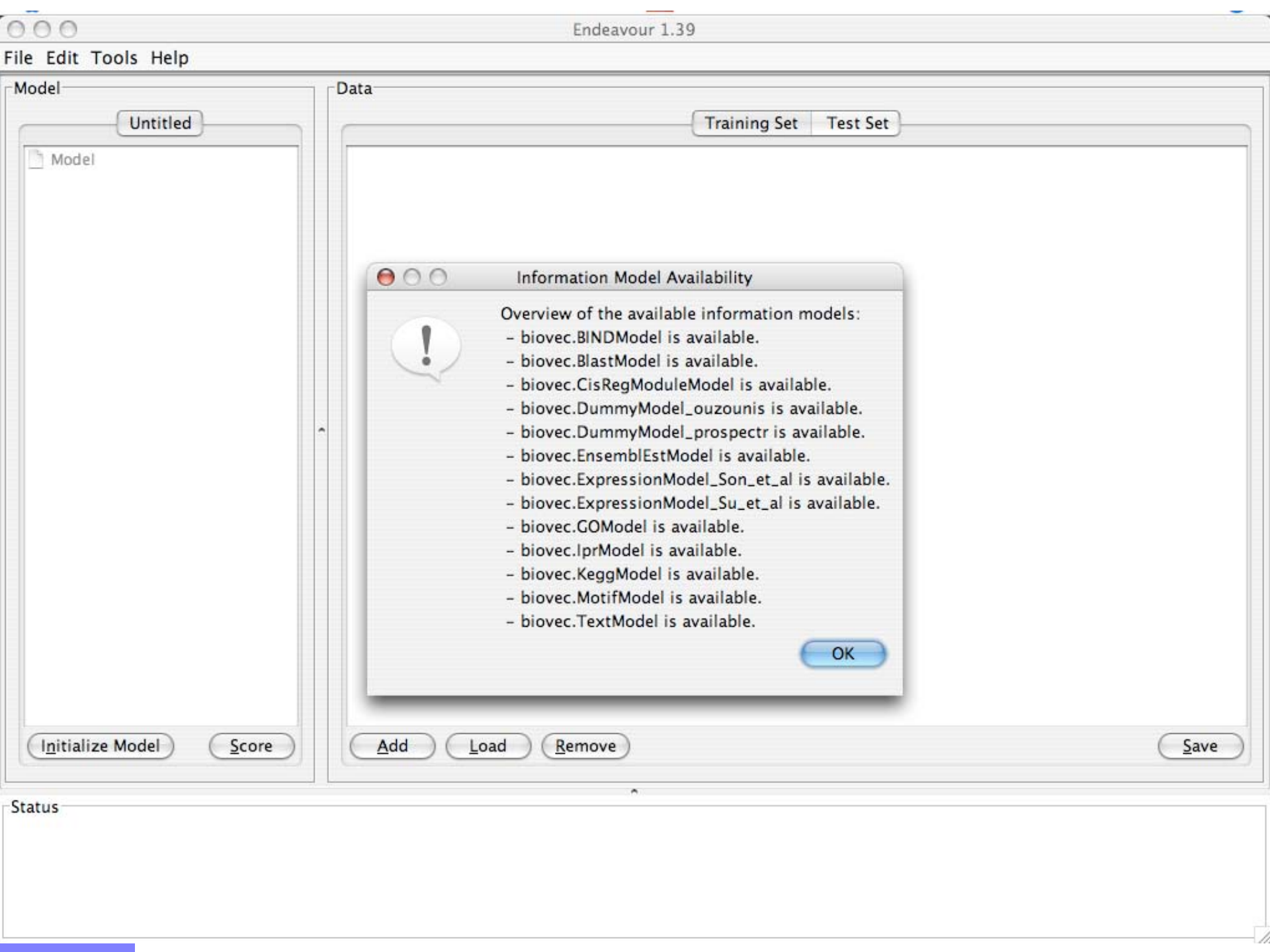


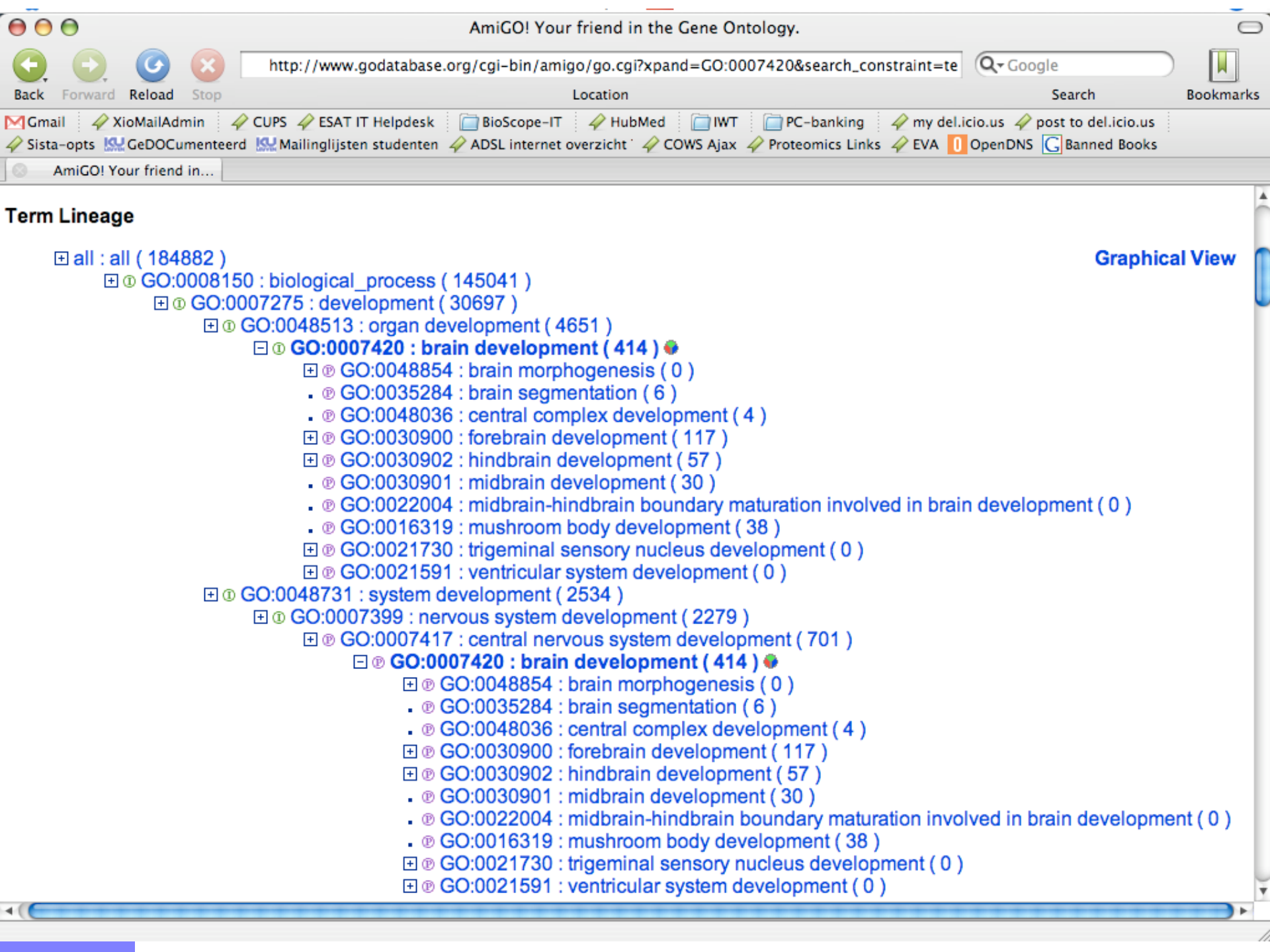


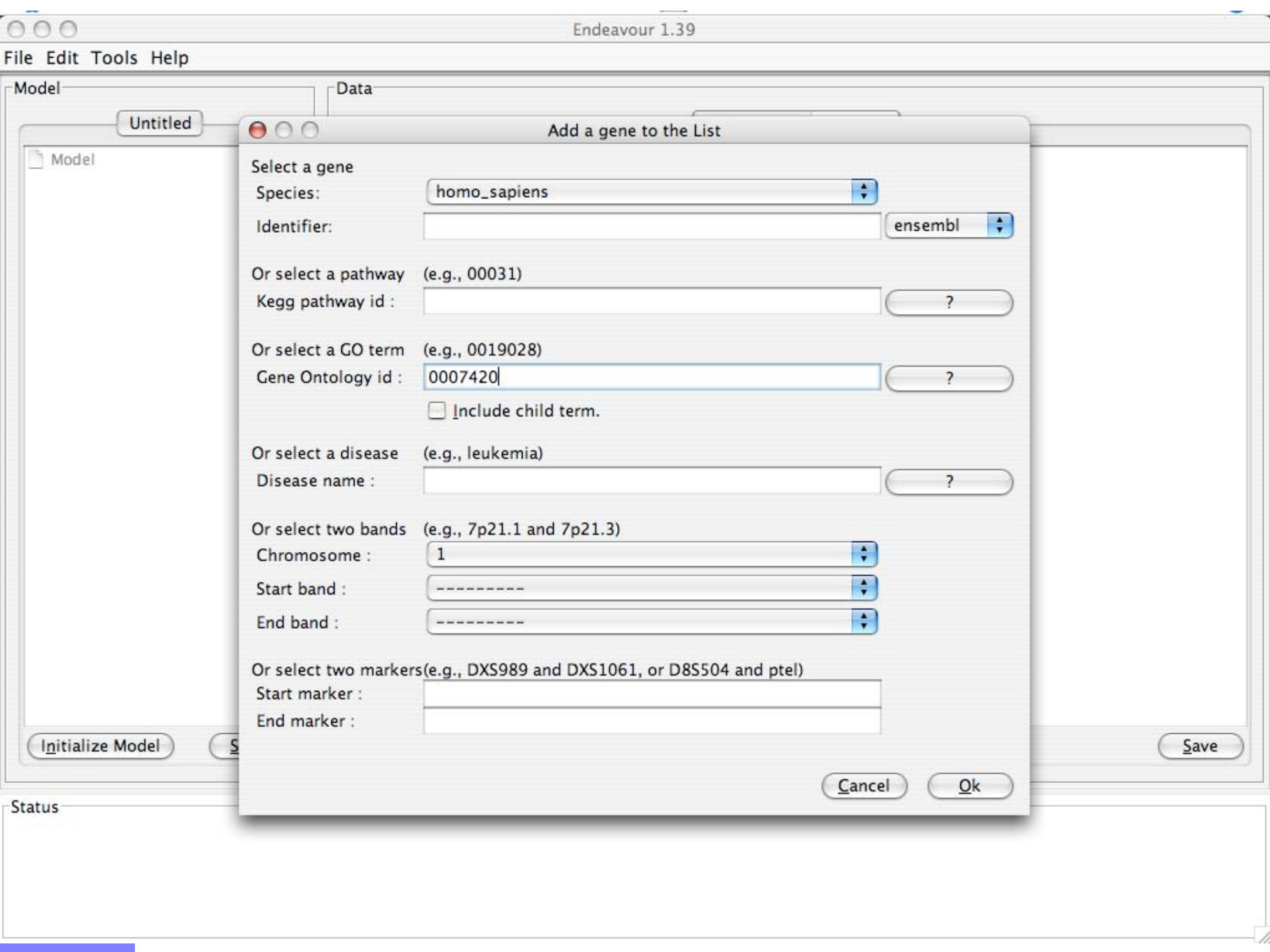
Offline demo

- Chediak-Higashi syndrome (OMIM:214500)
  - Psychomotor retardation
- Syndrome mapped to 1q42-qter
  - Caused by mutation in LYST gene
- Gene prioritization
  - Candidates from 1q42-qter (353 candidates)
  - Training genes: Gene Ontology category
    - Brain development GO:0007420 (60 genes)
  - LYST gene ranks 8/353


Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
Initialize Model
Score
Data
Training Set
Test Set
Add
Load
Remove
Save
Status

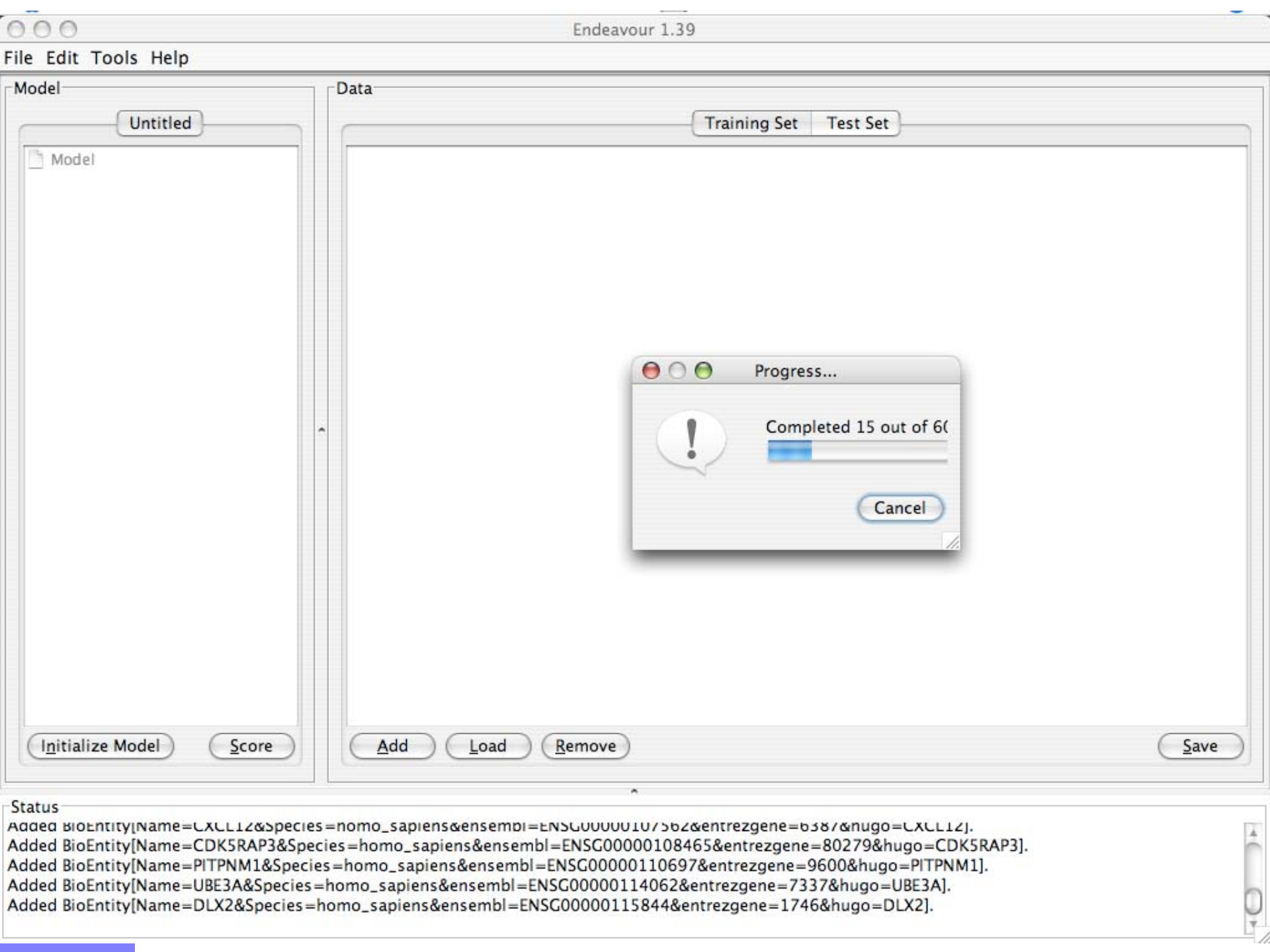

Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
Initialize Model
Score
Data
Training Set
Test Set
Information Model Availability
Overview of the available information models:
- biovec.BINDModel is available.
- biovec.BlastModel is available.
- biovec.CisRegModuleModel is available.
- biovec.DummyModel_ouzounis is available.
- biovec.DummyModel_prospectr is available.
- biovec.EnsemblEstModel is available.
- biovec.ExpressionModel_Son_et_al is available.
- biovec.ExpressionModel_Su_et_al is available.
- biovec.GOModel is available.
- biovec.IprModel is available.
- biovec.KeggModel is available.
- biovec.MotifModel is available.
- biovec.TextModel is available.
OK
Add
Load
Remove
Save
Status

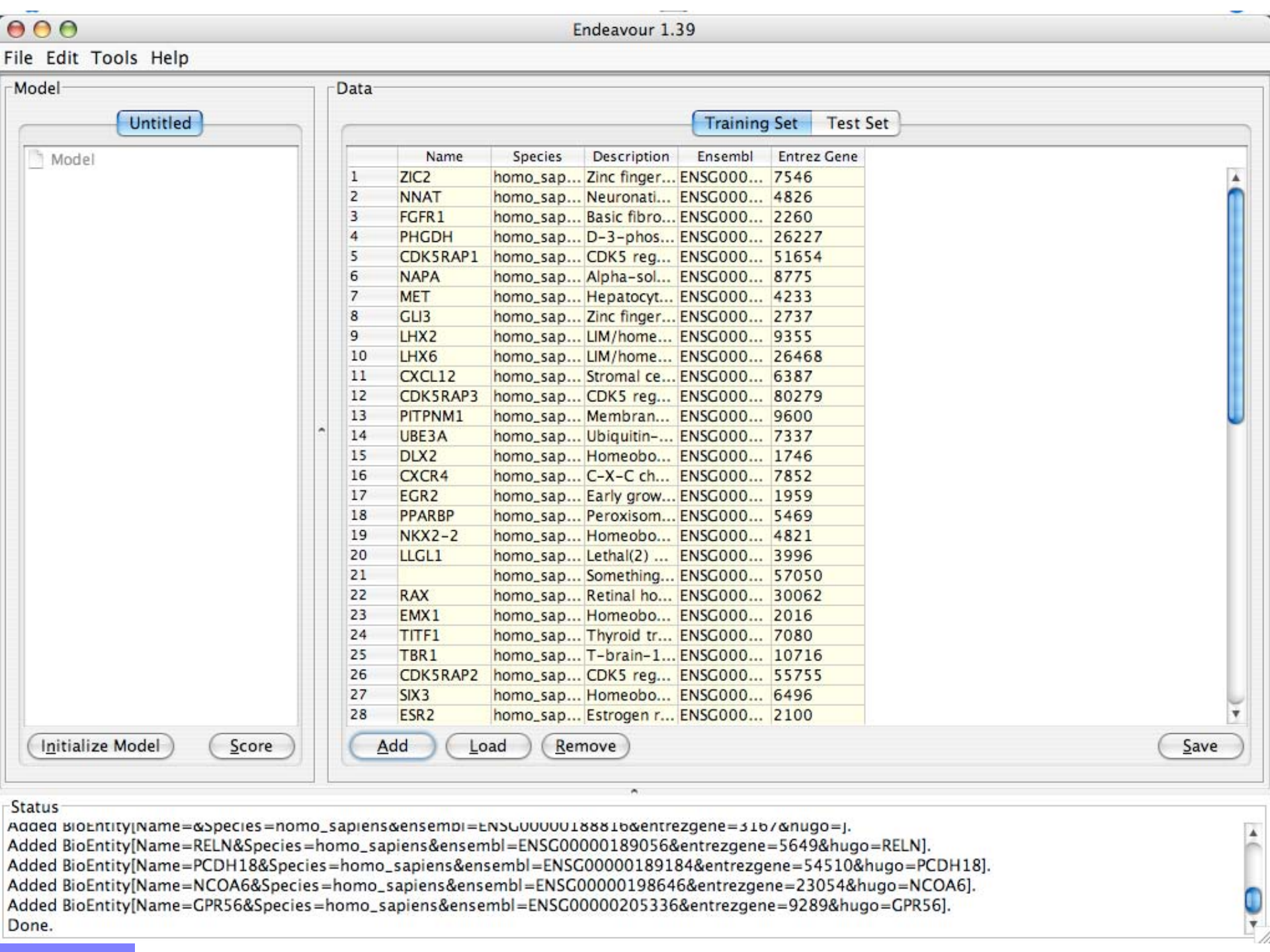

AmiGO! Your friend in the Gene Ontology.
Back
Forward
Reload
Stop
http://www.godatabase.org/cgi-bin/amigo/go.cgi?xpand=GO:0007420&search_constraint=te
Location
Google
Search
Bookmarks
Gmail
XioMailAdmin
CUPS
ESAT IT Helpdesk
BioScope-IT
HubMed
IWT
PC-banking
my del.icio.us
post to del.icio.us
Sista-opts
GeDOCumenteerd
Mailinglijsten studenten
ADSL internet overzicht
COWS Ajax
Proteomics Links
EVA
OpenDNS
Banned Books
AmiGO! Your friend in...
Term Lineage
Graphical View
all : all (184882)
GO:0008150 : biological_process (145041)
GO:0007275 : development (30697)
GO:0048513 : organ development (4651)
GO:0007420 : brain development (414)
GO:0048854 : brain morphogenesis (0)
GO:0035284 : brain segmentation (6)
GO:0048036 : central complex development (4)
GO:0030900 : forebrain development (117)
GO:0030902 : hindbrain development (57)
GO:0030901 : midbrain development (30)
GO:0022004 : midbrain-hindbrain boundary maturation involved in brain development (0)
GO:0016319 : mushroom body development (38)
GO:0021730 : trigeminal sensory nucleus development (0)
GO:0021591 : ventricular system development (0)
GO:0048731 : system development (2534)
GO:0007399 : nervous system development (2279)
GO:0007417 : central nervous system development (701)
GO:0007420 : brain development (414)
GO:0048854 : brain morphogenesis (0)
GO:0035284 : brain segmentation (6)
GO:0048036 : central complex development (4)
GO:0030900 : forebrain development (117)
GO:0030902 : hindbrain development (57)
GO:0030901 : midbrain development (30)
GO:0022004 : midbrain-hindbrain boundary maturation involved in brain development (0)
GO:0016319 : mushroom body development (38)
GO:0021730 : trigeminal sensory nucleus development (0)
GO:0021591 : ventricular system development (0)

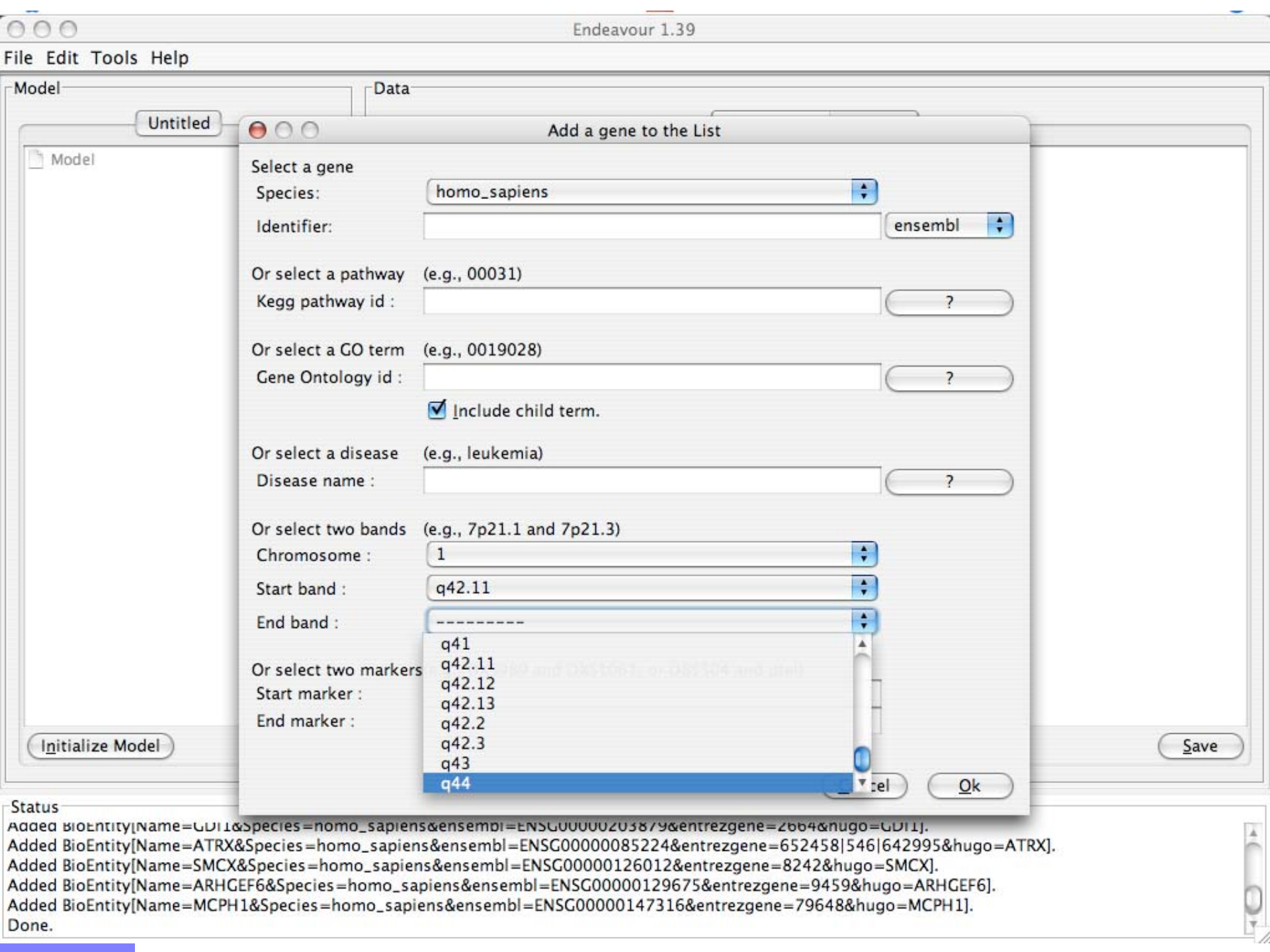

Endeavour 1.39
File Edit Tools Help
Model
Data
Untitled
Model
Initialize Model
Save
Status
Add a gene to the List
Select a gene
Species:
homo_sapiens
Identifier:
ensembl
Or select a pathway
(e.g., 00031)
Kegg pathway id :
?
Or select a GO term
(e.g., 0019028)
Gene Ontology id :
0007420
?
Include child term.
Or select a disease
(e.g., leukemia)
Disease name :
?
Or select two bands
(e.g., 7p21.1 and 7p21.3)
Chromosome :
1
Start band :

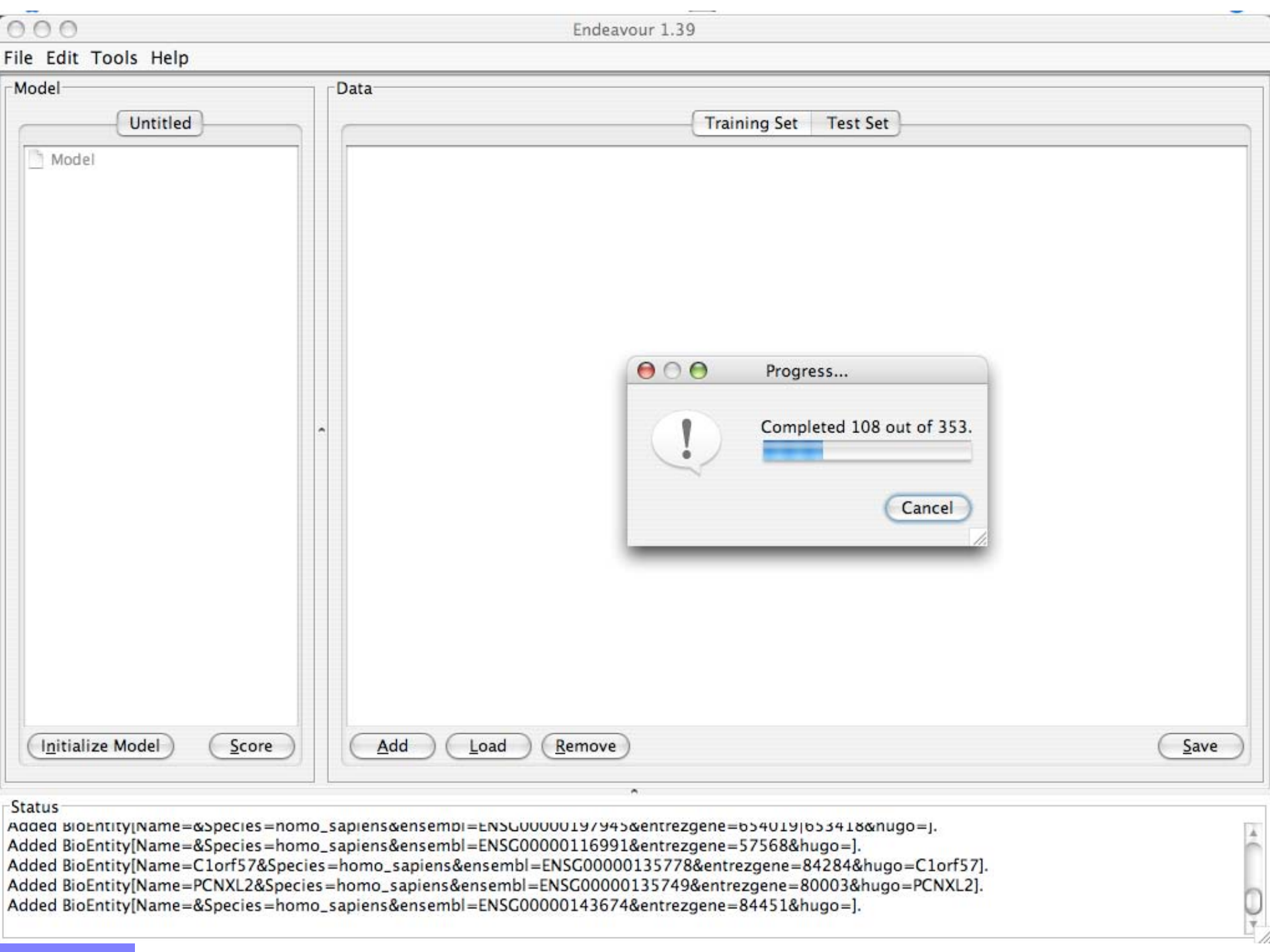
End band :

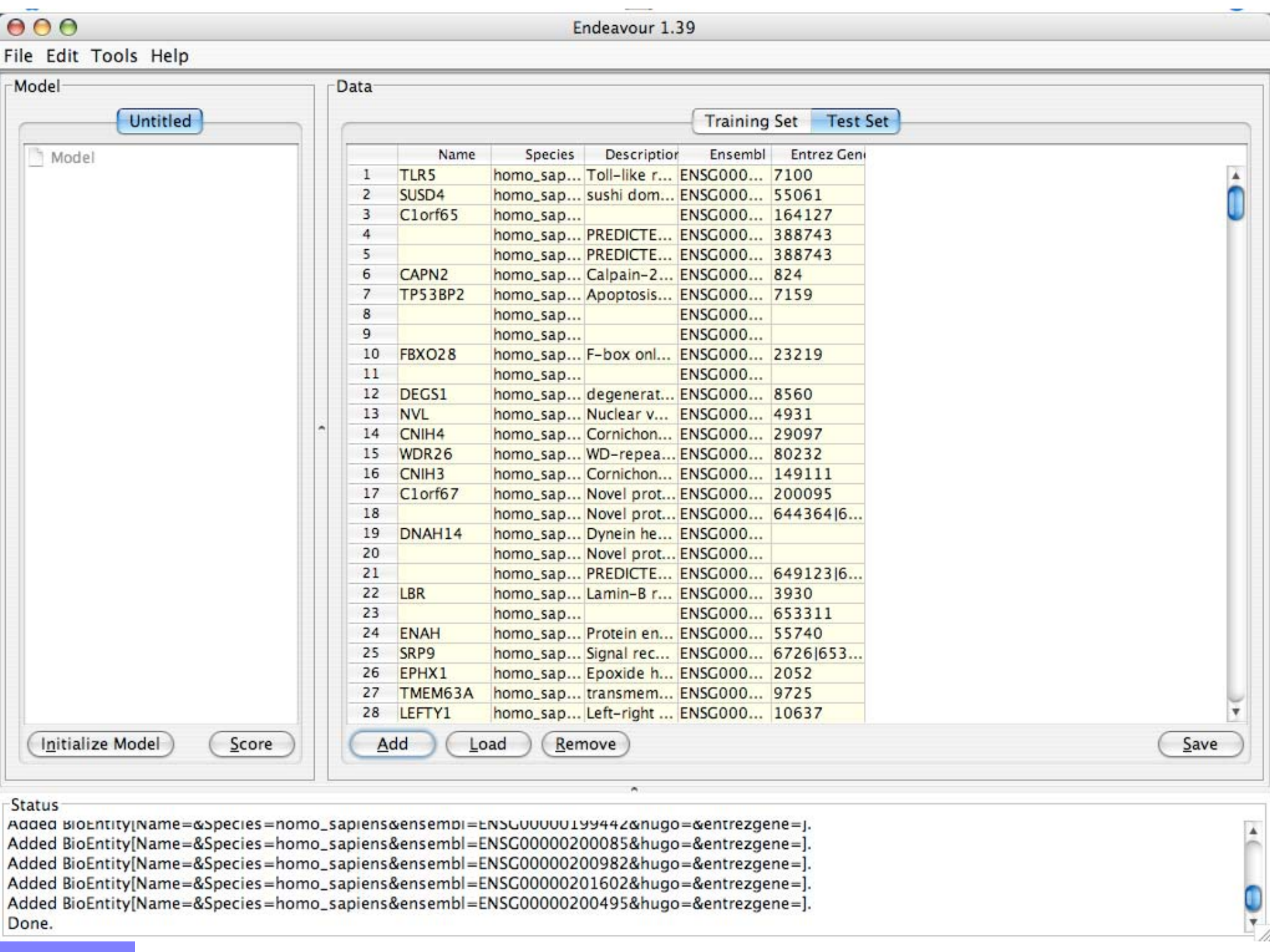
Or select two markers
(e.g., DXS989 and DXS1061, or D8S504 and ptel)
Start marker :
End marker :
Cancel
Ok

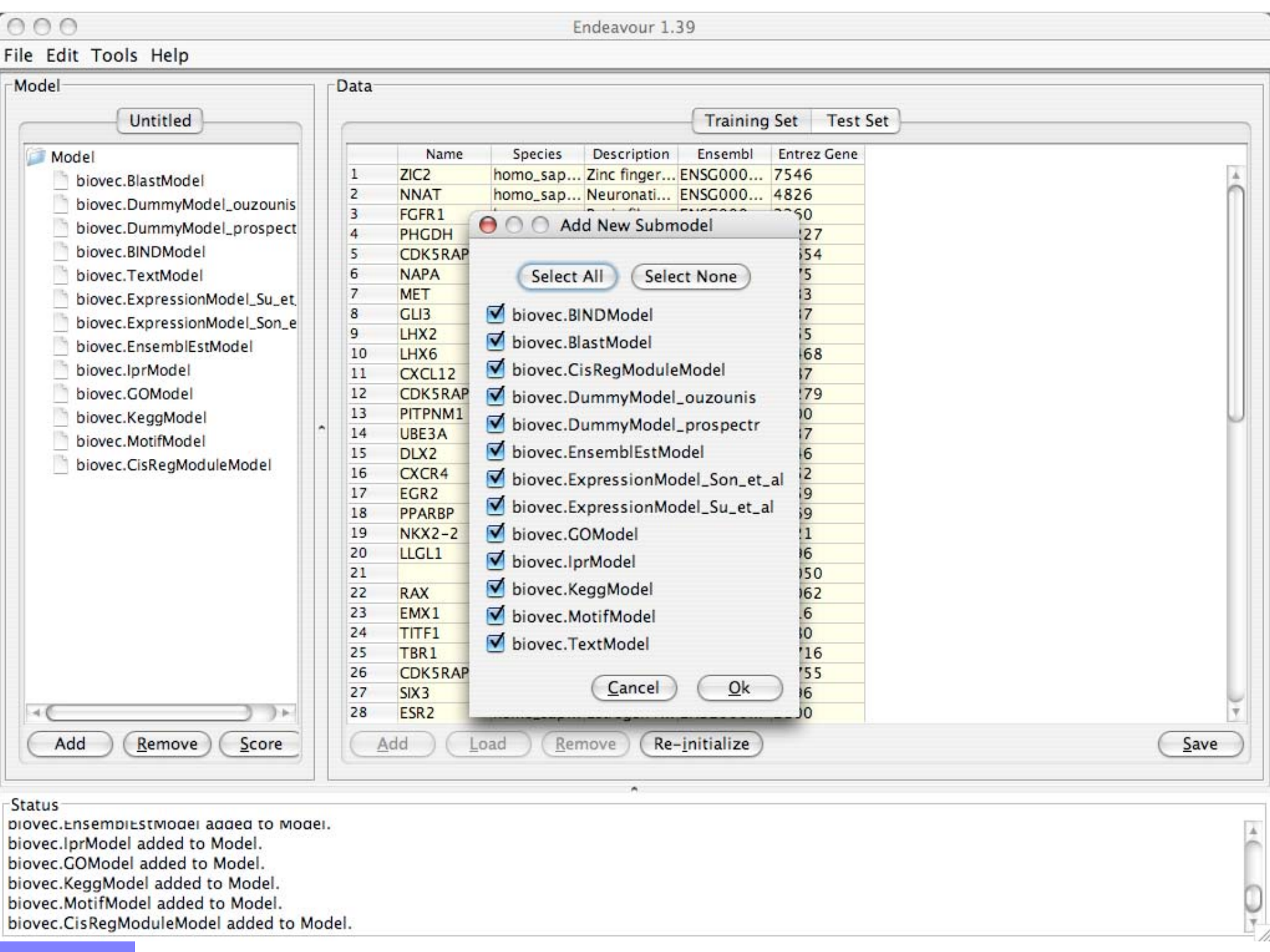

Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
Initialize Model
Score
Data
Training Set
Test Set
Add
Load
Remove
Save
Progress...
Completed 15 out of 6(
Cancel
Status
Added BioEntity[Name=CXCL12&Species=homo_sapiens&ensembl=ENSG00000107562&entrezgene=6387&hugo=CXCL12].
Added BioEntity[Name=CDK5RAP3&Species=homo_sapiens&ensembl=ENSG00000108465&entrezgene=80279&hugo=CDK5RAP3].
Added BioEntity[Name=PITPNM1&Species=homo_sapiens&ensembl=ENSG00000110697&entrezgene=9600&hugo=PITPNM1].
Added BioEntity[Name=UBE3A&Species=homo_sapiens&ensembl=ENSG00000114062&entrezgene=7337&hugo=UBE3A].
Added BioEntity[Name=DLX2&Species=homo_sapiens&ensembl=ENSG00000115844&entrezgene=1746&hugo=DLX2].

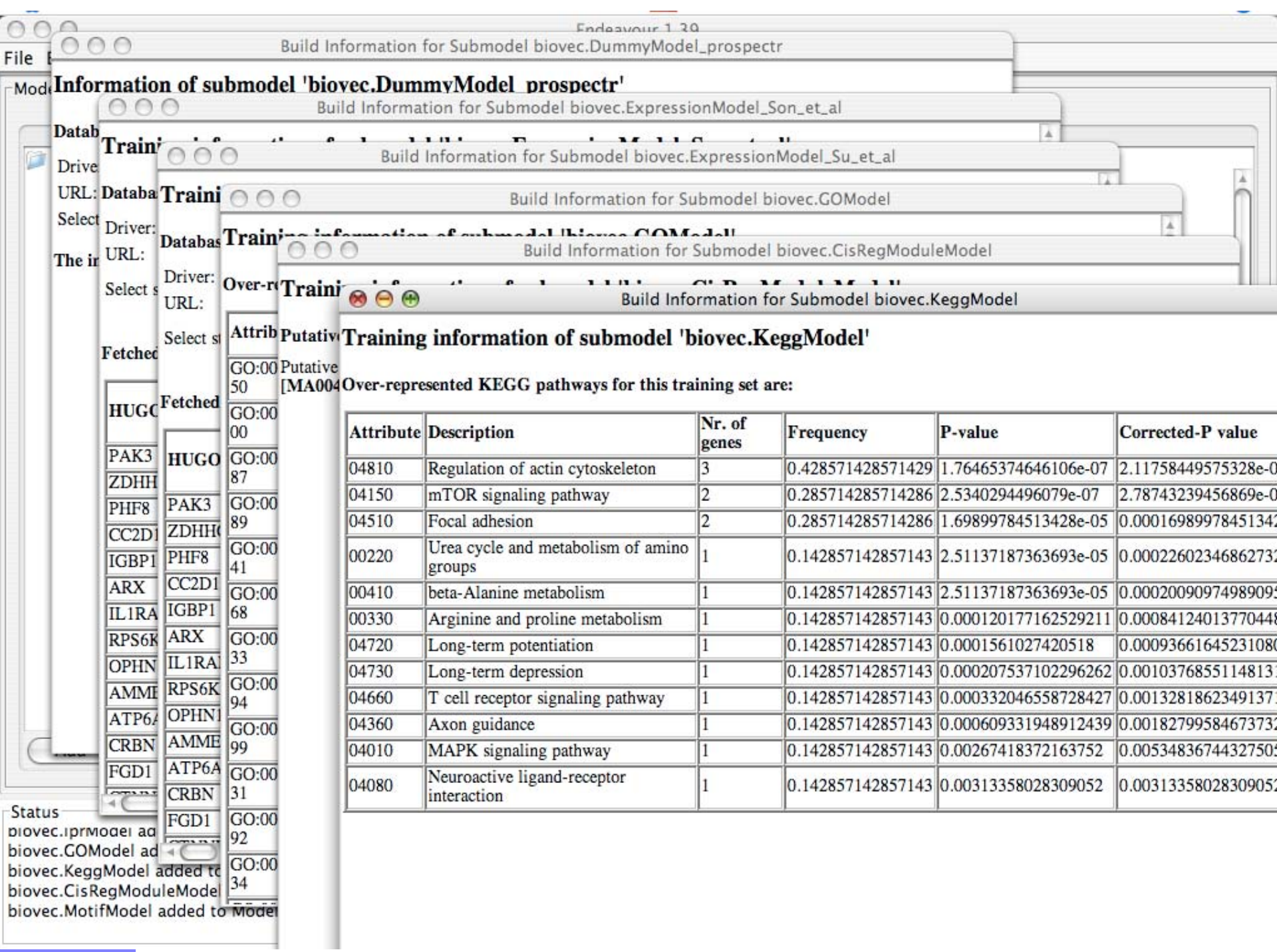

Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
Initialize Model
Score
Data
Training Set
Test Set
Name Species Description Ensembl Entrez Gene
1 ZIC2 homo_sap... Zinc finger... ENSG000... 7546
2 NNAT homo_sap... Neuronati... ENSG000... 4826
3 FGFR1 homo_sap... Basic fibro... ENSG000... 2260
4 PHGDH homo_sap... D-3-phos... ENSG000... 26227
5 CDK5RAP1 homo_sap... CDK5 reg... ENSG000... 51654
6 NAPA homo_sap... Alpha-sol... ENSG000... 8775
7 MET homo_sap... Hepatocyt... ENSG000... 4233
8 GLI3 homo_sap... Zinc finger... ENSG000... 2737
9 LHX2 homo_sap... LIM/home... ENSG000... 9355
10 LHX6 homo_sap... LIM/home... ENSG000... 26468
11 CXCL12 homo_sap... Stromal ce... ENSG000... 6387
12 CDK5RAP3 homo_sap... CDK5 reg... ENSG000... 80279
13 PITPNM1 homo_sap... Membran... ENSG000... 9600
14 UBE3A homo_sap... Ubiquitin-... ENSG000... 7337
15 DLX2 homo_sap... Homeobo... ENSG000... 1746
16 CXCR4 homo_sap... C-X-C ch... ENSG000... 7852
17 EGR2 homo_sap... Early grow... ENSG000... 1959
18 PPARBP homo_sap... Peroxisom... ENSG000... 5469
19 NKX2-2 homo_sap... Homeobo... ENSG000... 4821
20 LLGL1 homo_sap... Lethal(2) ... ENSG000... 3996
21 homo_sap... Something... ENSG000... 57050
22 RAX homo_sap... Retinal ho... ENSG000... 30062
23 EMX1 homo_sap... Homeobo... ENSG000... 2016
24 TITF1 homo_sap... Thyroid tr... ENSG000... 7080
25 TBR1 homo_sap... T-brain-1... ENSG000... 10716
26 CDK5RAP2 homo_sap... CDK5 reg... ENSG000... 55755
27 SIX3 homo_sap... Homeobo... ENSG000... 6496
28 ESR2 homo_sap... Estrogen r... ENSG000... 2100
Add
Load
Remove
Save
Status
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000188816&entrezgene=3187&hugo=].
Added BioEntity[Name=RELN&Species=homo_sapiens&ensembl=ENSG00000189056&entrezgene=5649&hugo=RELN].
Added BioEntity[Name=PCDH18&Species=homo_sapiens&ensembl=ENSG00000189184&entrezgene=54510&hugo=PCDH18].
Added BioEntity[Name=NCOA6&Species=homo_sapiens&ensembl=ENSG00000198646&entrezgene=23054&hugo=NCOA6].
Added BioEntity[Name=GPR56&Species=homo_sapiens&ensembl=ENSG00000205336&entrezgene=9289&hugo=GPR56].
Done.

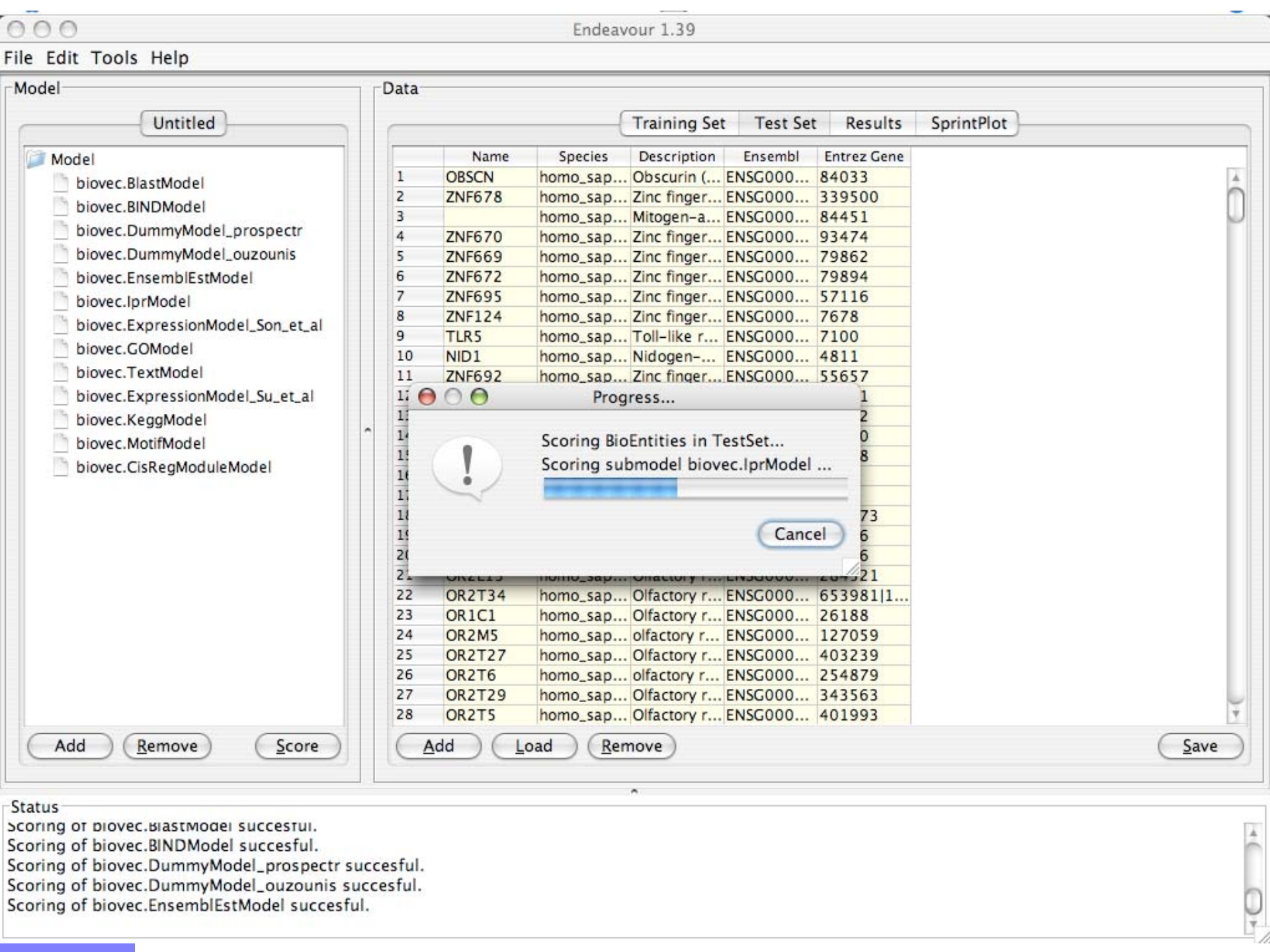

Endeavour 1.39
File Edit Tools Help
Model
Data
Untitled
Model
Add a gene to the List
Select a gene
Species:
homo_sapiens
Identifier:
ensembl
Or select a pathway
(e.g., 00031)
Kegg pathway id :
?
Or select a GO term
(e.g., 0019028)
Gene Ontology id :
?
Include child term.
Or select a disease
(e.g., leukemia)
Disease name :
?
Or select two bands
(e.g., 7p21.1 and 7p21.3)
Chromosome :
1
Start band :
q42.11
End band :
----
q41
q42.11
q42.12
q42.13
q42.2
q42.3
q43
q44
Or select two markers
Start marker :
End marker :
Ok
Initialize Model
Save
Status
Added BioEntity[Name=GDI1&Species=homo_sapiens&ensembl=ENSG00000203879&entrezgene=2664&hugo=GDI1].
Added BioEntity[Name=ATRX&Species=homo_sapiens&ensembl=ENSG00000085224&entrezgene=652458|546|642995&hugo=ATRX].
Added BioEntity[Name=SMCX&Species=homo_sapiens&ensembl=ENSG00000126012&entrezgene=8242&hugo=SMCX].
Added BioEntity[Name=ARHGEF6&Species=homo_sapiens&ensembl=ENSG00000129675&entrezgene=9459&hugo=ARHGEF6].
Added BioEntity[Name=MCPH1&Species=homo_sapiens&ensembl=ENSG00000147316&entrezgene=79648&hugo=MCPH1].
Done.


Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
Initialize Model
Score
Data
Training Set
Test Set
Add
Load
Remove
Save
Progress...
Completed 108 out of 353.
Cancel
Status
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000197945&entrezgene=654019|653418&hugo=].
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000116991&entrezgene=57568&hugo=].
Added BioEntity[Name=C1orf57&Species=homo_sapiens&ensembl=ENSG00000135778&entrezgene=84284&hugo=C1orf57].
Added BioEntity[Name=PCNXL2&Species=homo_sapiens&ensembl=ENSG00000135749&entrezgene=80003&hugo=PCNXL2].
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000143674&entrezgene=84451&hugo=].


Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
Initialize Model
Score
Data
Training Set
Test Set
Name Species Description Ensembl Entrez Gene
1 TLR5 homo_sap... Toll-like r... ENSG000... 7100
2 SUSD4 homo_sap... sushi dom... ENSG000... 55061
3 C1orf65 homo_sap... ENSG000... 164127
4 homo_sap... PREDICTE... ENSG000... 388743
5 homo_sap... PREDICTE... ENSG000... 388743
6 CAPN2 homo_sap... Calpain-2... ENSG000... 824
7 TP53BP2 homo_sap... Apoptosis... ENSG000... 7159
8 homo_sap... ENSG000...
9 homo_sap... ENSG000...
10 FBXO28 homo_sap... F-box onl... ENSG000... 23219
11 homo_sap... ENSG000...
12 DEGS1 homo_sap... degenerat... ENSG000... 8560
13 NVL homo_sap... Nuclear v... ENSG000... 4931
14 CNIH4 homo_sap... Cornichon... ENSG000... 29097
15 WDR26 homo_sap... WD-repea... ENSG000... 80232
16 CNIH3 homo_sap... Cornichon... ENSG000... 149111
17 C1orf67 homo_sap... Novel prot... ENSG000... 200095
18 homo_sap... Novel prot... ENSG000... 644364|6...
19 DNAH14 homo_sap... Dynein he... ENSG000...
20 homo_sap... Novel prot... ENSG000...
21 homo_sap... PREDICTE... ENSG000... 649123|6...
22 LBR homo_sap... Lamin-B r... ENSG000... 3930
23 homo_sap... ENSG000... 653311
24 ENAH homo_sap... Protein en... ENSG000... 55740
25 SRP9 homo_sap... Signal rec... ENSG000... 6726|653...
26 EPHX1 homo_sap... Epoxide h... ENSG000... 2052
27 TMEM63A homo_sap... transmem... ENSG000... 9725
28 LEFTY1 homo_sap... Left-right ... ENSG000... 10637
Add
Load
Remove
Save
Status
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000199442&hugo=&entrezgene=].
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000200085&hugo=&entrezgene=].
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000200982&hugo=&entrezgene=].
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000201602&hugo=&entrezgene=].
Added BioEntity[Name=&Species=homo_sapiens&ensembl=ENSG00000200495&hugo=&entrezgene=].
Done.


Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
biovec.BlastModel
biovec.DummyModel_ouzounis
biovec.DummyModel_prospect
biovec.BINDModel
biovec.TextModel
biovec.ExpressionModel_Su_et
biovec.ExpressionModel_Son_e
biovec.EnsemblEstModel
biovec.IprModel
biovec.GOModel
biovec.KeggModel
biovec.MotifModel
biovec.CisRegModuleModel
Add
Remove
Score
Data
Training Set
Test Set
Name
Species
Description
Ensembl
Entrez Gene
1 ZIC2 homo_sap... Zinc finger... ENSG000... 7546
2 NNAT homo_sap... Neuronati... ENSG000... 4826
3 FGFR1
4 PHGDH
5 CDK5RAP
6 NAPA
7 MET
8 GLI3
9 LHX2
10 LHX6
11 CXCL12
12 CDK5RAP
13 PITPNM1
14 UBE3A
15 DLX2
16 CXCR4
17 EGR2
18 PPARBP
19 NKX2-2
20 LLGL1
21
22 RAX
23 EMX1
24 TITF1
25 TBR1
26 CDK5RAP
27 SIX3
28 ESR2
Add New Submodel
Select All
Select None
biovec.BINDModel
biovec.BlastModel
biovec.CisRegModuleModel
biovec.DummyModel_ouzounis
biovec.DummyModel_prospectr
biovec.EnsemblEstModel
biovec.ExpressionModel_Son_et_al
biovec.ExpressionModel_Su_et_al
biovec.GOModel
biovec.IprModel
biovec.KeggModel
biovec.MotifModel
biovec.TextModel
Cancel
Ok
Add
Load
Remove
Re-initialize
Save
Status
biovec.EnsemblEstModel added to Model.
biovec.IprModel added to Model.
biovec.GOModel added to Model.
biovec.KeggModel added to Model.
biovec.MotifModel added to Model.
biovec.CisRegModuleModel added to Model.


Build Information for Submodel biovec.DummyModel_prospectr
Information of submodel 'biovec.DummyModel_prospectr'
Build Information for Submodel biovec.ExpressionModel_Son_et_al
Build Information for Submodel biovec.ExpressionModel_Su_et_al
Build Information for Submodel biovec.GOModel
Build Information for Submodel biovec.CisRegModuleModel
Build Information for Submodel biovec.KeggModel
Training information of submodel 'biovec.KeggModel'
Over-represented KEGG pathways for this training set are:
Attribute
Description
Nr. of genes
Frequency
P-value
Corrected-P value
04810
Regulation of actin cytoskeleton
3
0.428571428571429
1.76465374646106e-07
2.11758449575328e-0
04150
mTOR signaling pathway
2
0.285714285714286
2.5340294496079e-07
2.78743239456869e-0
04510
Focal adhesion
2
0.285714285714286
1.69899784513428e-05
0.0001698997845134
00220
Urea cycle and metabolism of amino groups
1
0.142857142857143
2.51137187363693e-05
0.0002260234686273
00410
beta-Alanine metabolism
1
0.142857142857143
2.51137187363693e-05
0.0002009097498909
00330
Arginine and proline metabolism
1
0.142857142857143
0.000120177162529211
0.0008412401377044
04720
Long-term potentiation
1
0.142857142857143
0.0001561027420518
0.0009366164523108
04730
Long-term depression
1
0.142857142857143
0.000207537102296262
0.0010376855114813
04660
T cell receptor signaling pathway
1
0.142857142857143
0.000332046558728427
0.0013281862349137
04360
Axon guidance
1
0.142857142857143
0.000609331948912439
0.0018279958467373
04010
MAPK signaling pathway
1
0.142857142857143
0.00267418372163752
0.0053483674432750
04080
Neuroactive ligand-receptor interaction
1
0.142857142857143
0.00313358028309052
0.0031335802830905
Status
biovec.GOModel added
biovec.KeggModel added
biovec.CisRegModuleModel
biovec.MotifModel added


Endeavour 1.39
File Edit Tools Help
Model
Untitled
Model
biovec.BlastModel
biovec.BINDModel
biovec.DummyModel_prospectr
biovec.DummyModel_ouzounis
biovec.EnsemblEstModel
biovec.IprModel
biovec.ExpressionModel_Son_et_al
biovec.GOModel
biovec.TextModel
biovec.ExpressionModel_Su_et_al
biovec.KeggModel
biovec.MotifModel
biovec.CisRegModuleModel
Add
Remove
Score
Data
Training Set
Test Set
Results
SprintPlot
Name
Species
Description
Ensembl
Entrez Gene
1 OBSCN homo_sap... Obscurin (... ENSG000... 84033
2 ZNF678 homo_sap... Zinc finger... ENSG000... 339500
3 homo_sap... Mitogen-a... ENSG000... 84451
4 ZNF670 homo_sap... Zinc finger... ENSG000... 93474
5 ZNF669 homo_sap... Zinc finger... ENSG000... 79862
6 ZNF672 homo_sap... Zinc finger... ENSG000... 79894
7 ZNF695 homo_sap... Zinc finger... ENSG000... 57116
8 ZNF124 homo_sap... Zinc finger... ENSG000... 7678
9 TLR5 homo_sap... Toll-like r... ENSG000... 7100
10 NID1 homo_sap... Nidogen-... ENSG000... 4811
11 ZNF692 homo_sap... Zinc finger... ENSG000... 55657
Progress...
Scoring BioEntities in TestSet...
Scoring submodel biovec.IprModel ...
Cancel
22 OR2T34 homo_sap... Olfactory r... ENSG000... 653981|1...
23 OR1C1 homo_sap... Olfactory r... ENSG000... 26188
24 OR2M5 homo_sap... olfactory r... ENSG000... 127059
25 OR2T27 homo_sap... Olfactory r... ENSG000... 403239
26 OR2T6 homo_sap... olfactory r... ENSG000... 254879
27 OR2T29 homo_sap... Olfactory r... ENSG000... 343563
28 OR2T5 homo_sap... Olfactory r... ENSG000... 401993
Add
Load
Remove
Save
Status
Scoring of biovec.BlastModel succesful.
Scoring of biovec.BINDModel succesful.
Scoring of biovec.DummyModel_prospectr succesful.
Scoring of biovec.DummyModel_ouzounis succesful.
Scoring of biovec.EnsemblEstModel succesful.

Endeavour 1.39

File Edit Tools Help

Model

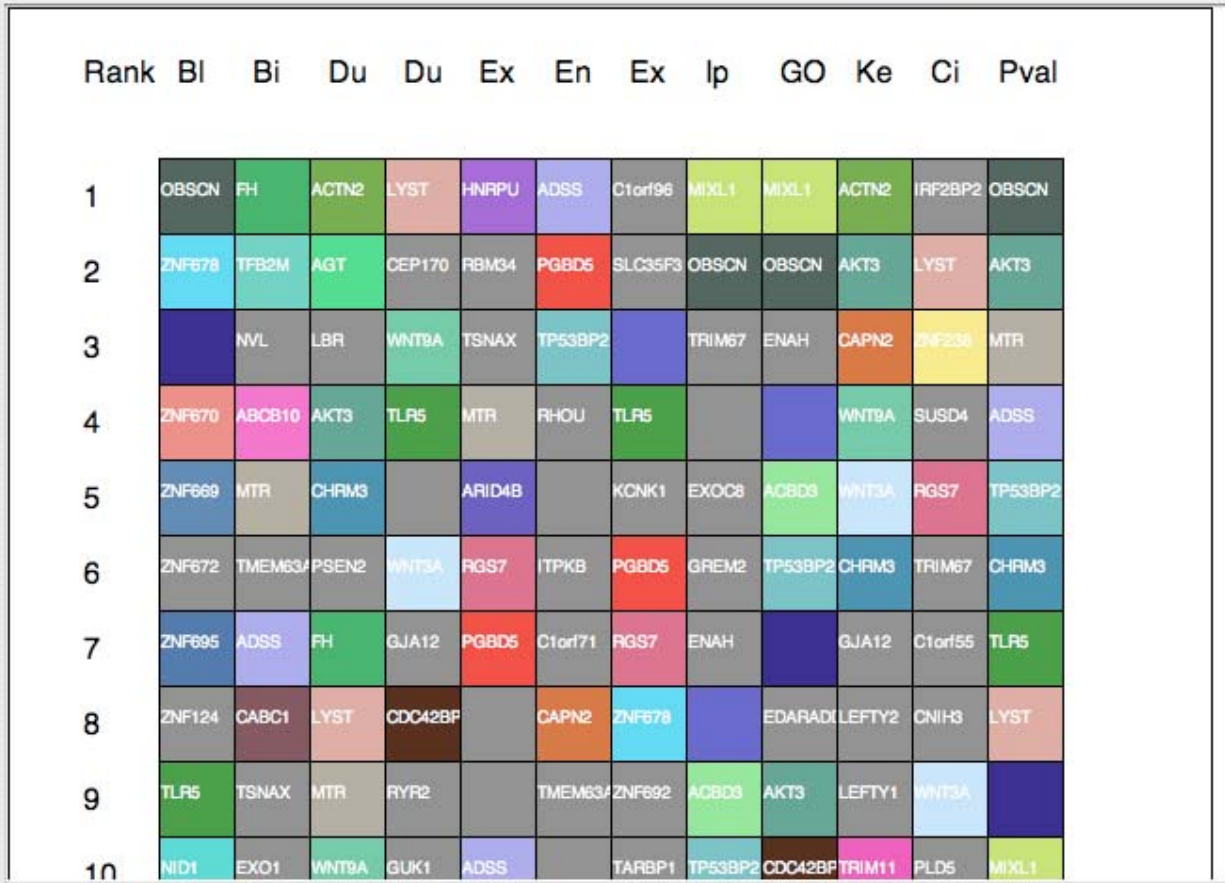
Untitled

- Model
  - biovec.BlastModel
  - biovec.BINDModel
  - biovec.DummyModel_ouzounis
  - biovec.DummyModel_prospect
  - biovec.ExpressionModel_Su_et
  - biovec.EnsemblEstModel
  - biovec.ExpressionModel_Son_e
  - biovec.IprModel
  - biovec.GOModel
  - biovec.KeggModel
  - biovec.CisRegModuleModel

Add Remove Score

Data

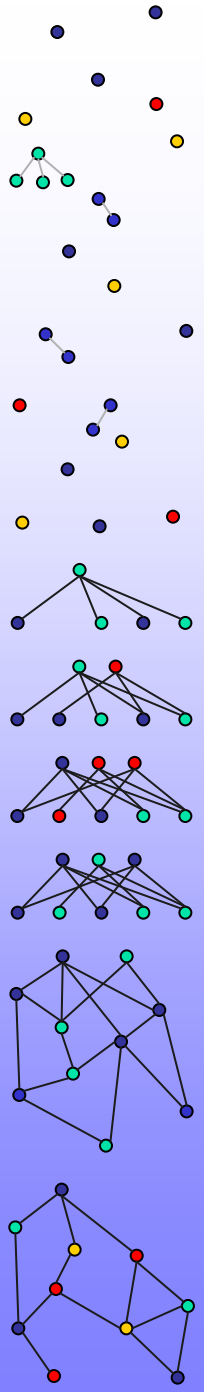
Training Set | Test Set | Results | SprintPlot



Refresh Save figure

Status

Scoring of biovec.ExpressionModel_Son_et_al succesful.
Scoring of biovec.IprModel succesful.
Scoring of biovec.GOModel succesful.
Scoring of biovec.KeggModel succesful.
Scoring of biovec.CisRegModuleModel succesful.
Scoring Finished succesfully.



Conclusion

- Prioritization of candidate genes
  - Central problem in molecular biology
- Prioritization with order statistics
  - Large-scale crossvalidation
  - Endeavour
  - DiGeorge syndrome candidate
- Quick-and-dirty prioritization of diabetes genes

KATHOLIEKE UNIVERSITEIT
LEUVEN

SymBioSys

BIOMAGNET
Bioinformatics and Modelling: from Genomes to Networks

K.U.L. ESAT-SCD: L. Tranchevent, R. Barriot, Y. Shi, B. Coessens, S. Van Vooren, D. Nitsch, S. Leach
U. Bristol: T. De Bie
K.U.L. CME-UZ: J. Vermeesch, K. Devriendt, B. Thienpont, F. Hannes
K.U.L. VIB3: D. Lambrechts, S. Maity, P. Carmeliet
K.U.L. VIB4: S. Aerts, B. Hassan, P. Van Loo, P. Marynen

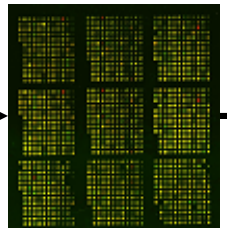
Array CGH: from diagnosis to gene discovery

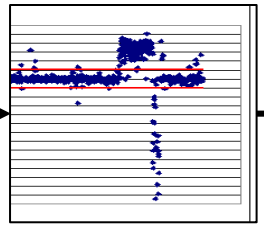
Patients with congenital & acquired disorders

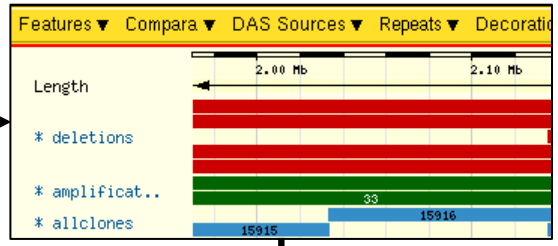
CGH microarrays Molecular karyotyping

Statistical analysis

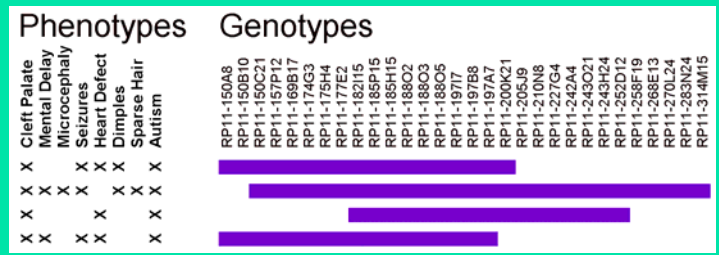
Location of chromosomal imbalances







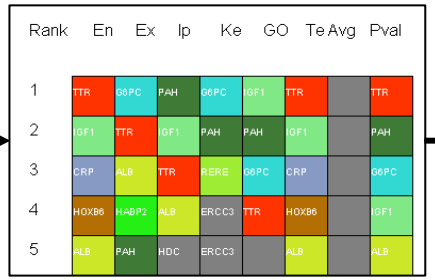
Databasing



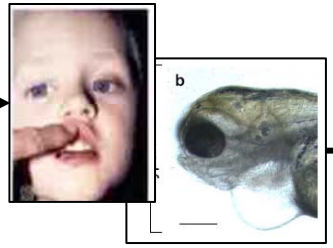
- Map chromosomal abnormalities
- Improved diagnosis

Discover new disease causing genes and explain their function

Prioritized candidate genes



Validation



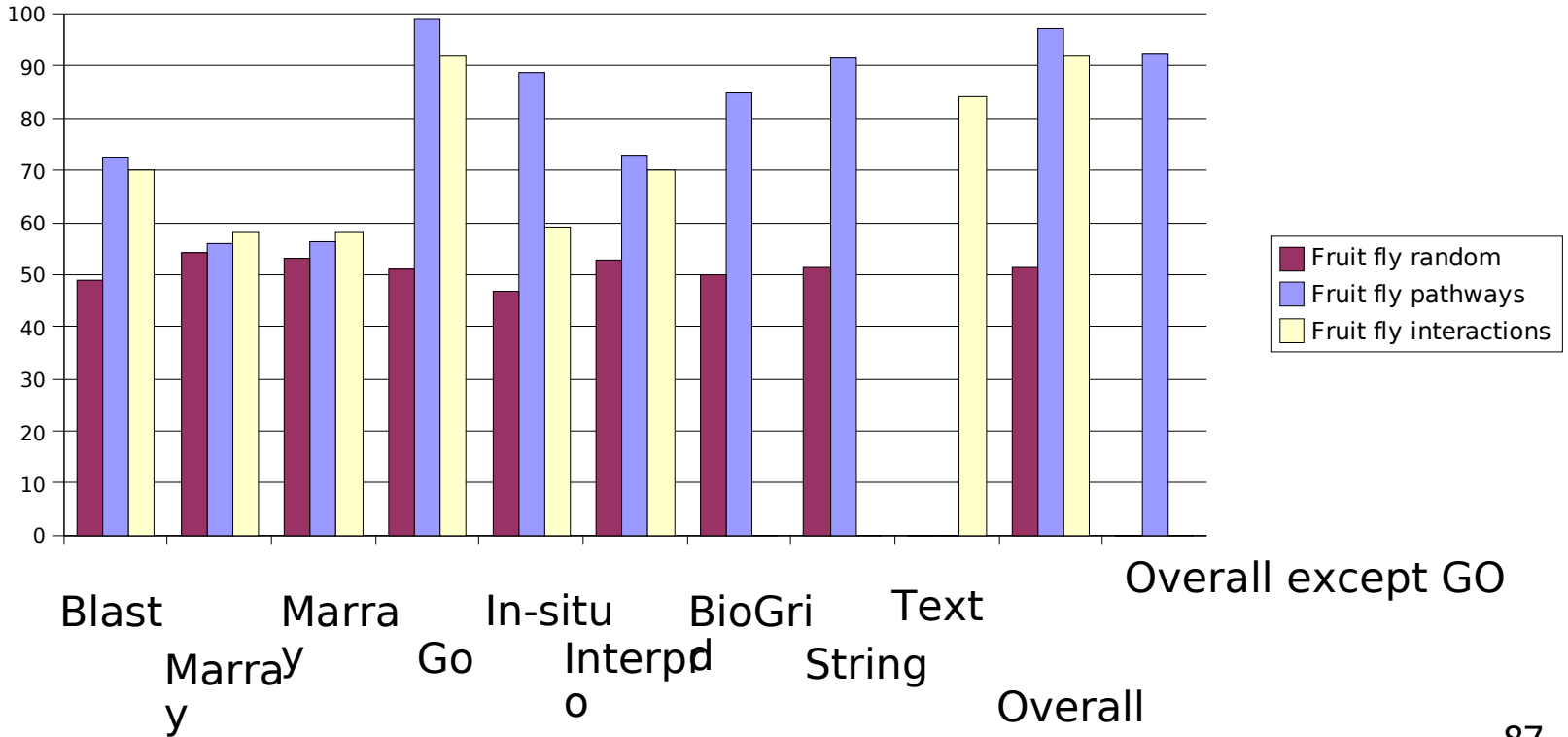
Gene prioritization in animal models (fly)

- *S. Aerts, B. Hassan, KUL DME Neurobiology*
- New data sources
  - In-situ data from the BDGP
  - String data
  - BioGrid data

- Also available
  - Gene ontology
  - Interpro domains
  - Text mining data
  - Blast alignments
  - Microarray data

Validation

- 10 pathway sets and 46 interactions sets
- Use of the leave-one-out cross-validation again
- Comparison with randomized performance



Text mining

Patient Description	**hospitalised because of growth retardation and a chromosomal disorder was suspected. A partial interstitial deletion of the long arm of chromosome 2 was confirmed by routine cytogenetic investigation. He developed epileptic seizures soon after this diagnosis. At 11 months of age length was 69 cm (3rd centile), weight 6.8 kg (<3rd centile) and OFC 44 cm (3rd centile). Craniofacial dysmorphic features were present such as flat occiput, thin white hairs, downward slanting palpebral fissures, strabismus, bilaterally epicanthic folds, ptosis of the right eyelashes, prominent nasal bridge, long philtrum, thin lips, small and high palate and dysplastic ears. The penis was small and there was a sacral dimple. There was tendency for opistotonus. Surgical intervention for ptosis of the eyelids was done at the age of 2 years. At the age of 8 years 7 months he was re-examined. Anamnesis showed that he suffered from frequent colds and otitis media. He was still tube-feeded. There were no behavioural problems. Length was 114 cm (6cm <3rd centile), weight 19,5 kg (3rd to 10th centile) and OFC 48,8 cm (1cm <3rd centile). He had microcephaly, small face, downward slanting palpebral fissures, beaked nose, short philtrum and small but high palate. There was hypotonia and he was severely mentally retarded. At the age of 11 years and 11 months length was 137 cm (3rd to 10th centile), weight 27 kg (3rd to 10th centile) and OFC of 50 cm (1cm <3rd centile). He still had his first teeth and an X-ray of the oral cavity showed absent adult teeth. His voice was rather specific and resembled "Donald Duck" speech. He had contractures of the knees.**

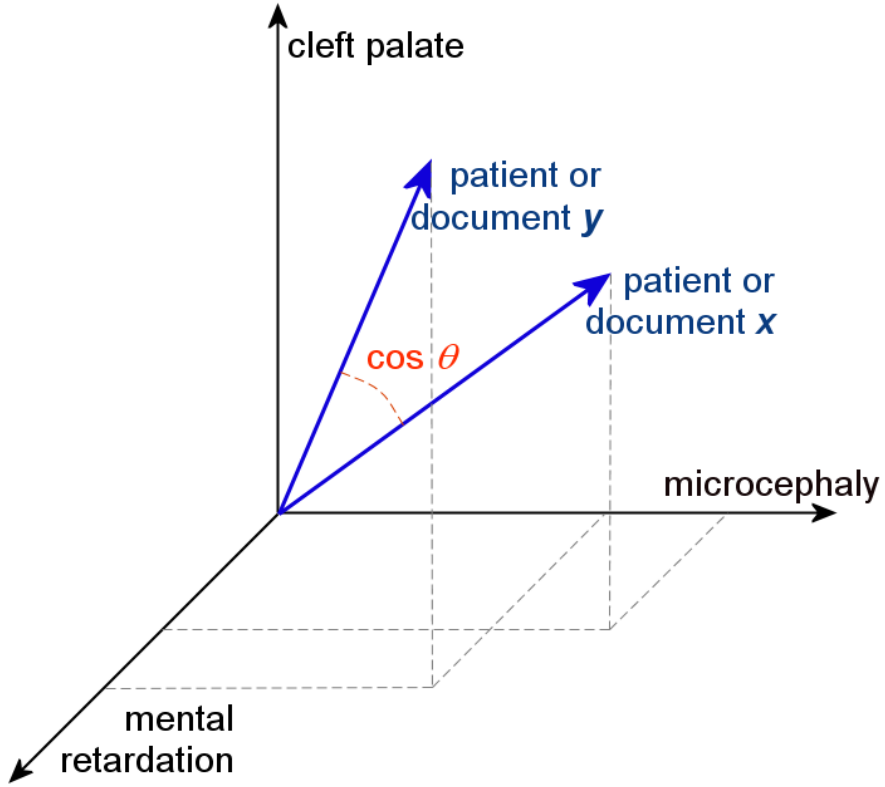
Dysmorphology Text Profile

Below, a text profile is shown for this case report. Relevant taxons from the LDDB ontology are associated to the patient description and keywords.

word_id	value	bar
200300_Male genitalia, general abnormalities	0.302	
080500_Palpebral fissures, general abnormalities	0.272	
030105_Microcephaly	0.272	
090107_Convex/beaked profile of nose	0.26	
060106_Dysplastic ears	0.257	
180109_Inguinal hernia	0.256	
250500_Knee, general abnormalities	0.25	
320122_Seizures/abnormal EEG	0.249	
080602_Epicanthic folds	0.239	
130100_Teeth, general abnormalities	0.235	
110303_Short philtrum	0.217	
320112_Hypotonia	0.204	
120403_High palate	0.176	
080200_Eyelashes, general abnormalities	0.172	
030401_Flat occiput	0.169	
110301_Long philtrum	0.162	
140100_Voice, general abnormalities	0.155	
160204_Sacral dimple/sinus	0.152	
100113_Small face	0.149	
090302_High/prominent nasal bridge	0.119	

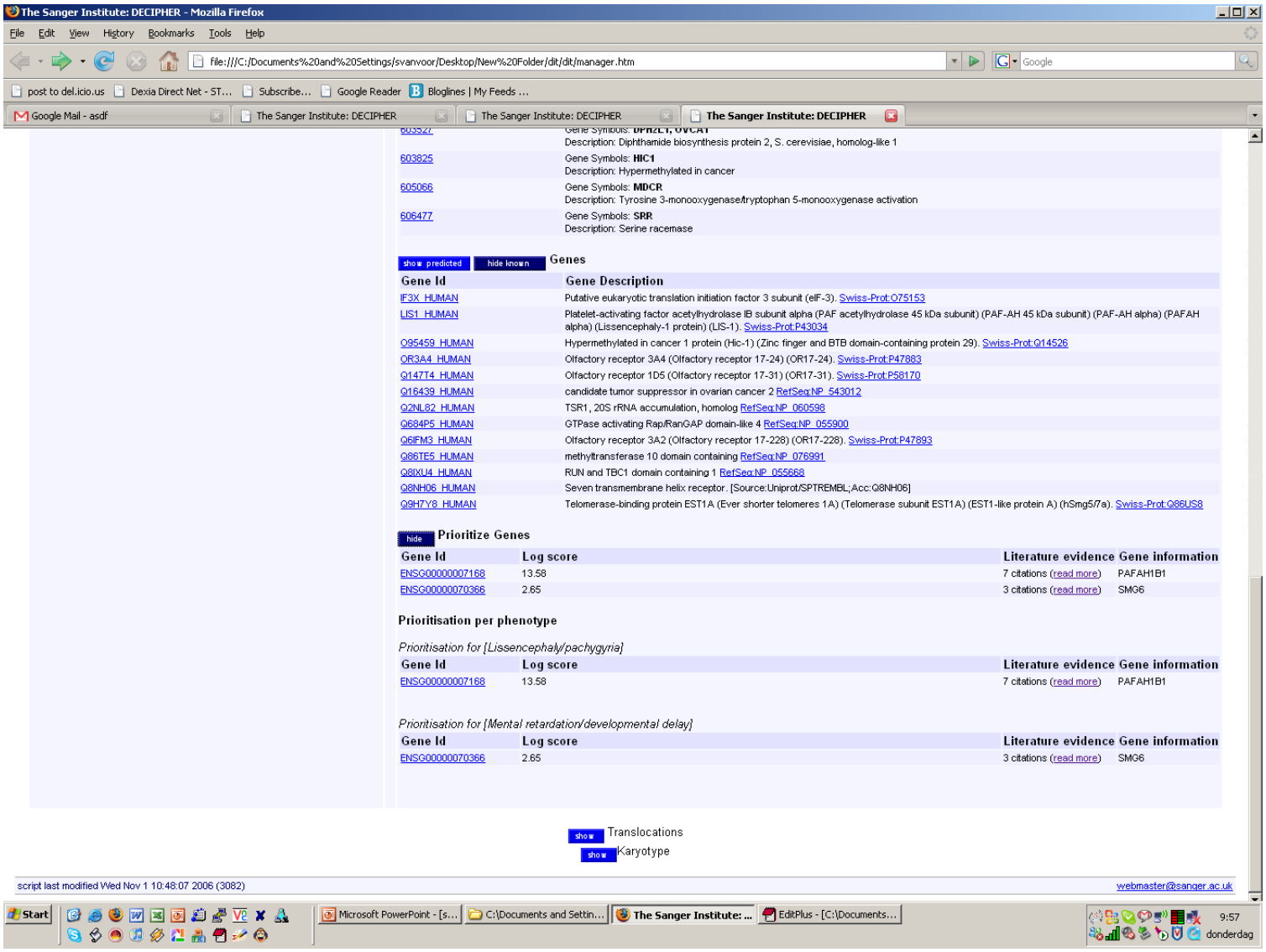
Text mining

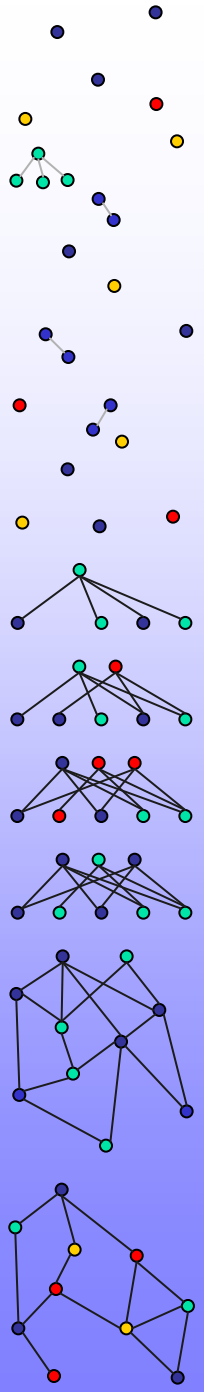
word_id	w	bar
glycin	58%	
heme	32%	
mitochondri	26%	
cytochrom	25%	
mitochondr...	25%	
inner	24%	
group	23%	
cleavag	20%	
system	19%	
cytochrom_c	18%	
lyase	17%	
coval	17%	
shuttl	15%	
link	13%	
synthas	12%	
membran	11%	
degrad	8%	
metabol	7%	
subunit	5%	



$$sim_{vs}(q, x_i) = Q \cdot X_i = \frac{\sum_{j=1}^{m} v_j \cdot w_{ij}}{\sqrt{\sum_{j=1}^{m} (v_j)^2 \cdot \sum_{j=1}^{m} (w_{ij})^2}}$$

Text mining


The Sanger Institute: DECIPHER - Mozilla Firefox
603825 Gene Symbols: HIC1 Description: Hypermethylated in cancer
605066 Gene Symbols: MDCR Description: Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation
606477 Gene Symbols: SRR Description: Serine racemase
show predicted
hide known
Genes
Gene Id
Gene Description
IF3X_HUMAN Putative eukaryotic translation initiation factor 3 subunit (eIF-3). Swiss-Prot:O75153
LIS1_HUMAN Platelet-activating factor acetylhydrolase IB subunit alpha (PAF acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha) (PAFAH alpha) (Lissencephaly-1 protein) (LIS-1). Swiss-Prot:P43034
Q95459_HUMAN Hypermethylated in cancer 1 protein (Hic-1) (Zinc finger and BTB domain-containing protein 29). Swiss-Prot:Q14526
OR3A4_HUMAN Olfactory receptor 3A4 (Olfactory receptor 17-24) (OR17-24). Swiss-Prot:P47883
Q147T4_HUMAN Olfactory receptor 1D5 (Olfactory receptor 17-31) (OR17-31). Swiss-Prot:P58170
Q16439_HUMAN candidate tumor suppressor in ovarian cancer 2 RefSeq:NP_543012
Q2NL82_HUMAN TSR1, 20S rRNA accumulation, homolog RefSeq:NP_060598
Q684P5_HUMAN GTPase activating Rap/RanGAP domain-like 4 RefSeq:NP_055900
Q6IFM3_HUMAN Olfactory receptor 3A2 (Olfactory receptor 17-228) (OR17-228). Swiss-Prot:P47893
Q86TE5_HUMAN methyltransferase 10 domain containing RefSeq:NP_076991
Q8IXU4_HUMAN RUN and TBC1 domain containing 1 RefSeq:NP_055668
Q8NH06_HUMAN Seven transmembrane helix receptor. [Source:Uniprot/SPTREMBL;Acc:Q8NH06]
Q9H7Y8_HUMAN Telomerase-binding protein EST1A (Ever shorter telomeres 1A) (Telomerase subunit EST1A) (EST1-like protein A) (hSmg5/7a). Swiss-Prot:Q86US8
hide
Prioritize Genes
Gene Id
Log score
Literature evidence
Gene information
ENSG00000007168 13.58 7 citations (read more) PAFAH1B1
ENSG00000070366 2.65 3 citations (read more) SMG6
Prioritisation per phenotype
Prioritisation for [Lissencephaly/pachygyria]
Prioritisation for [Mental retardation/developmental delay]
show Translocations
show Karyotype
script last modified Wed Nov 1 10:48:07 2006 (3082)
webmaster@sanger.ac.uk



Array CGH: from diagnosis to gene discovery

1. Processing of array CGH data
2. Databasing and mining of patient descriptions
3. Genotype-phenotype correlation
4. Candidate gene prioritization
5. Experimental validation of candidate genes

Genotype-phenotype correlation

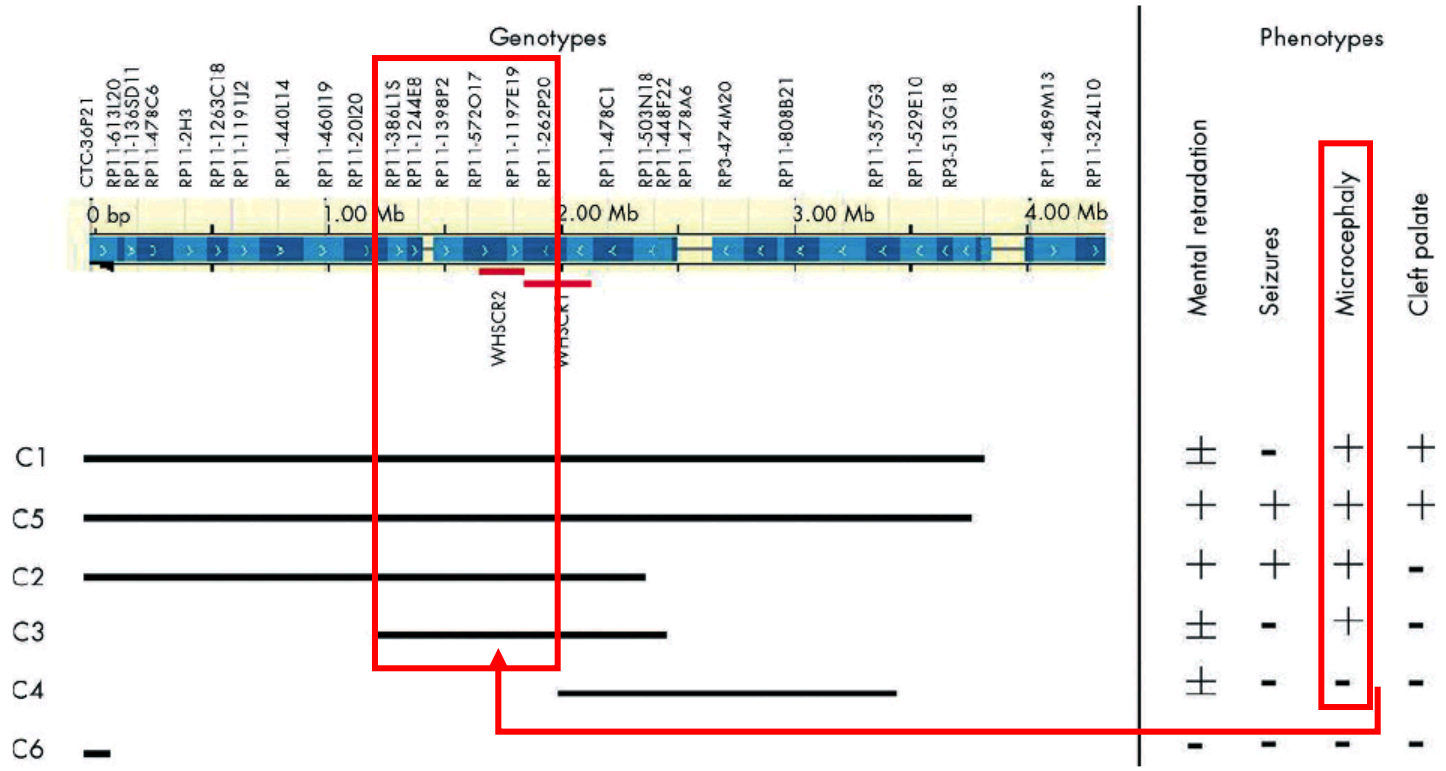


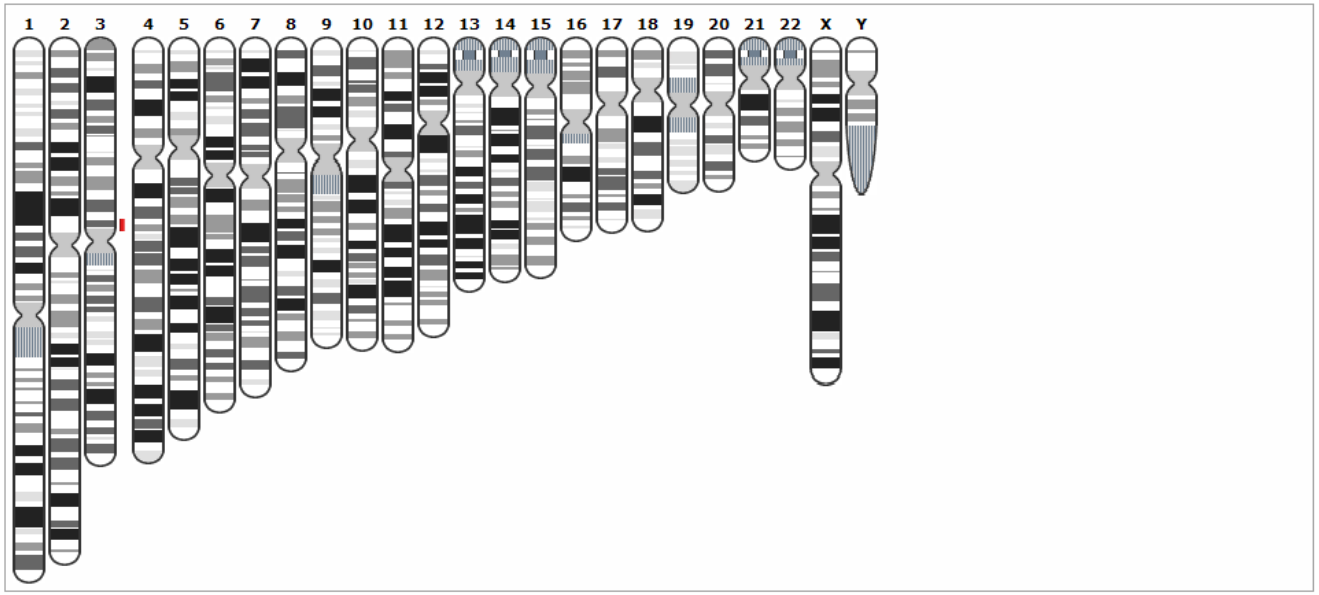
Figure 1 Genotypes and phenotypes of the patients analysed in this study. The top part shows the clones represented on the array from the telomeric 4 Mb together with the DNA contig representation of Ensembl (01/2004). Clones in italics are not represented in the Golden Path sequence. The Wolf-Hirschhorn critical regions WHSCR1 and WHSCR2 are indicated with the lines under the Ensembl contig representation. The bottom shows a summary of the genotypes of all the patients analysed in this study. The lines indicate the sizes of the 4p deletions. On the right, the main phenotypic features discussed in the text are presented.

My Reports

Project	ID	CME ID	UZ ID	External ID	Comment	BAC/PAC Aberrations	Agilent Aberrations	Clinical Data	OMD LNDB	Keywords	Full Text	Photo	Consent Data	Consent Images	In DECIPHER	Action
cmeleuven	305057361710303V293					2	0		0	0						
cmeleuven	304282561910524M245					2	0		0	0						
cmeleuven	3032755		640620M084			1	0		0	0						
cmeleuven	302224256660330V222					1	0		0	0						
cmeleuven	301243077					1	0		0	0						
cmeleuven	300342199950509M125			VGDm	16 dup16p13.3	1	0	✓	11	0			✓			
cmeleuven	299200920070202M107			DRj	10 dup10q24.2	1	0	✓	14	0			✓			
cmeleuven	29832141					1	0		0	0						
cmeleuven	297011264					1	0		0	0						
cmeleuven	296362873061120G009					1	0		0	0						
cmeleuven	295075021860814M076					1	0		0	0						
cmeleuven	294275355940513V144					1	0		0	0						
cmeleuven	293355824					1	0		0	0						
cmeleuven	292251620010920G052			Rl	4/18 del4q35.1 dup18q22.3	2	0	✓	6	0			✓			
cmeleuven	291340640041202G-Foreign			Cl	15/21 del15q26.3 del21q21.2	1	0	✓	9	0			✓			
cmeleuven	290293754040125B013			Ll	9 del9p24.3p22.2 dup9p22.2p21.3	2	0	✓	15	0			✓			
cmeleuven	288054023791031V094			Fs	15 del15q13.3	1	0	✓	0	0			✓			
cmeleuven	287186198920911V193				6 del6	1	0	✓	0	0			✓			
cmeleuven	286296287900227M261			Ck	7 del7q31.1	1	0	✓	0	0			✓			
cmeleuven	285351673	Foreign		Ae	2 del2q21.2q22.3	1	0	✓	11	0			✓			
cmeleuven	284123583911010V002			Gk	1 del1p36.33	1	0	✓	5	0			✓			
cmeleuven	281999999					0	0		0	0						
cmeleuven	280356414	Bulgaria		Gv	9 del 9p22 Foreign	1	0		0	0						
cmeleuven	279354904	Bulgaria		Ci	9 del 9p22 Foreign	1	0		0	0						
cmeleuven	278341907060303V065			Kj	9 dup 9p21.3	1	0	✓	4	0			✓			
cmeleuven	277345774850511M-Nl			VDm	18 del18q23	1	0	✓	0	0			✓			
cmeleuven	276300777810205M			Ds	6 del6q14.1	1	0	✓	0	0			✓			
cmeleuven	275325158800503M144			Pw	12 dup12p13.2mos	1	0	✓	8	0			✓			
cmeleuven	274294498040206B049			Db	17 del17p12	1	0	✓	6	1			✓			
cmeleuven	273036488791122M021			Gd	18 del18p11.22	1	0	✓	5	0			✓			
cmeleuven	272137955761115M			Ls	mar 11 dup11q21	1	0	✓	11	0			✓			
cmeleuven	270102693871220M048			VBe	15 dup15q12	1	0	✓	2	0			✓			
cmeleuven	269097412900316V048			VKt	16 del16q22.2	1	0	✓	12	0			✓			
cmeleuven	268353219	fetus		Mr	t(6;12)del6q27 dup12pter12p11.1	2	0	✓	0	1			✓			
cmeleuven	267256008011011B052			Rs	8 del8p23	1	0	✓	2	0			✓			
cmeleuven	266353449	fetus		Td	13 del13q31.1q33.1	1	0	✓	5	0			✓			

CGHGate v3.1 beta

Cytogenetic Report



Cytogenetic Report

General | Karyogram | Clinical Data | Full Text Description | Free Text Keywords | Chromosomal Aberrations | OMD Taxons | Prioritized Genes | Pictures | Printable Report

Add Sanger Clones - Add Agilent Reporters

General overview of Sanger Clone chromosomal aberrations

chromosome	type
3	deletion

Detailed information on chromosomal aberrations:

Chromosome: 3
Type: deletion
Fold: null
Start: 84462053
Stop: 84640085
Proximal Clone: RP11-447J13
Affected Clones: RP11-474M18 |
Distal Clone: RP11-382L10
Method of Confirmation: fish

No genes in this region.

No **Agilent** chromosomal aberrations have been reported yet.

Cytogenetic Report

General | Karyogram | Clinical Data | Full Text Description | Free Text Keywords | Chromosomal Aberrations | OMD Taxons | Prioritized Genes | Pictures | Printable Report

Add - Remove

11.02.05 - Thick lower lip (association: +)
09.01.04 - Large nose (association: +)
13.01.14 - Small teeth (association: +)
08.02.04 - Long/prominent eyelashes (association: +)
10.01.04 - Coarse facial features (association: +)
32.32.14 - Dandy-Walker malformation (association: +)
32.32.15 - Pons/medulla/basal ganglia, abnormal (association: +)
32.31.08 - Cerebellar vermis hypoplasia/aplasia (association: +)
04.04.01 - Generalized hirsutism (association: +)
34.03.07 - Nevi or lentigines (association: +)
14.01.02 - Hoarse voice (association: +)
29.01.12 - Recurrent infections (association: +)

https://tomcatbackup.esat.kuleuven.be/cghgate_cme/edit/enteromddynamic.jsp

CGHGate v3.1 beta

Search: Case Report Go

New | My Reports | Export | Edit | Karyogram | Add DAS | Ensembl | Help | Admin Panel | Log Out

Logged in as cme_admin. **Case Report 52 is active for editing.**

Search Ontology Keywords

CGHGate will look for *any* of the words you enter.

Continue

All Ontology Kewords

Please select relevant key phrases from this list. These will be used to annotate your submission to the CGHGate database.

The phenotype is clearly present.
The phenotype is only marginally present.
The phenotype is clearly not present, although it reasonably could be.
There is no information. This is the default setting, so this option needn't be marked.

- **01.00.00.....BUILD**
- **02.00.00.....STATURE**
- **04.00.00.....HAIR**
- **05.00.00.....FOREHEAD**
- **06.00.00.....EARS**
- **07.00.00.....EYES, GLOBES**
  - *07.01.00.....Eyes, general abnormalities (including spacing)*
  - *07.02.00.....Anterior chamber, general abnormalities*
  - *07.03.00.....Conjunctiva, general abnormalities*
  - *07.04.00.....Cornea, general abnormalities*
  - *07.05.00.....Globes, general abnormalities*
  - *07.06.00.....Iris, general abnormalities*
    - 07.06.01.....Aniridia
    - 07.06.02.....Brushfield spots
    - 07.06.03.....Coloboma of iris
    - 07.06.04.....Heterochromia of iris
    - 07.06.05.....Pigmentary abnormalities of iris
    - 07.06.06.....Iris atrophy/dysplasia
    - 07.06.07.....Depigmentation of iris
  - *07.07.00.....Lens, general abnormalities*
  - *07.08.00.....Macula, general abnormalities*
  - *07.09.00.....Optic disc and nerve, general abnormalities*
  - *07.10.00.....Pupil, general abnormalities*
  - *07.11.00.....Retina, general abnormalities*
  - *07.12.00.....Sclera, general abnormalities*
  - *07.13.00.....Vision, general abnormalities*
  - *07.14.00.....Vitreous, general abnormalities*
- **08.00.00.....EYES, ASSOCIATED STRUCTURES**
- **09.00.00.....NOSE**
- **10.00.00.....FACE**
- **11.00.00.....MOUTH**

Neurology: [normal]
Behaviour: [normal]
Psychiatric: [normal]

Intelligence description:
Intelligence measures: IQ: [] DQ: [] test used: [] age: [y m]
Intelligence description: [mild]

Full Text Description

Edit

father has deletion

Free Text Keywords

Edit

This patient does not yet have keywords.

LNDB Taxons

Add - Remove

11.02.05 - Thick lower lip (association: +)
09.01.04 - Large nose (association: +)
13.01.14 - Small teeth (association: +)
08.02.04 - Long/prominent eyelashes (association: +)
10.01.04 - Coarse facial features (association: +)
32.32.14 - Dandy-Walker malformation (association: +)
32.32.15 - Pons/medulla/basal ganglia, abnormal (association: +)
32.31.08 - Cerebellar vermis hypoplasia/aplasia (association: +)
04.04.01 - Generalized hirsutism (association: +)
34.03.07 - Nevi or lentigines (association: +)
14.01.02 - Hoarse voice (association: +)
29.01.12 - Recurrent infections (association: +)

Chromosomal Aberrations

[deletion] 3:83768022-86268550 (Confirmation: fish)
Distal Unaffected reporter: RP11-382L10
Affected reporters: RP11-474M18 |
Proximal Unaffected reporter: RP11-447J13

OMIM genes in this region

Ensembl ID	Type	Location	OMIM	Description

Cases with Similar Aberrations

Case Report: 186 - 128343 - 921202V033 - DBb (Created by: cme_admin on 2007-07-26 17:50:24.0)
[duplication] 3:83768022-86268550 (Confirmation: qpcr)
Comment: 3 dup3p12.1

Case Report: 40 - 253458 - 020107B017 - VDCI (Created by: cme_admin on 2007-07-26 17:50:24.0)
[duplication] 3:80611850-82799401 (Confirmation: Not Specified)
Comment: 3 dup3p12.2

Cases with Similar Phenotypes

No Case Reports with similar phenotypes found.

Congenital heart disease genes

- *B. Thienpont, K. Devriendt, J. Vermeesch, KUL CME*
- 60 patients without diagnosis
  - Congenital heart defect
  - & Chromosomal phenotype
    - 2nd major congenital anomaly
    - Or mental retardation/special education
    - Or > 3 minor anomalies
- Array Comparative Genomic Hybridization
  - 1 Mb resolution
- 11 anomalies detected
  - 5 deletions
  - 2 duplications
  - 3 complex rearrangements
  - 1 mosaic monosomy 7

Candidate regions

- 4 regions with known critical genes, 6 new regions, 80 candidate genes

aberration	***gene***
del(5)(q23)	?
del(5)(q35.1)	*NKX2.5*
del(5)(q35.2qter)	*NSD1*
del(14)(q22.1q23.1)	?
del(22)(q12.2)	?
dup(22)(q11)	*TBX1*
dup(19)(p13.12p13.11)	?
del(9)(q34.3qter),dup(20)(q13.33qter)	*NOTCH1, EHMT1*
del(13)(q31.1q31.3),dup(13)(q31.3q33.2),inv(13)	?
del(4)(q34.3q35.1),dup(4)(q34),inv(4)	?

Gene prioritization

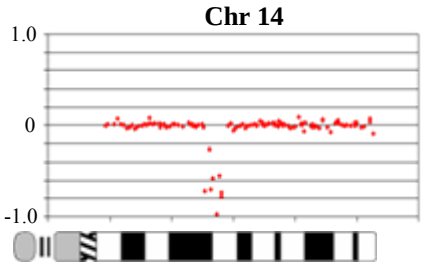
del(14)(q22.1q23.1) ?

	Expression data	KEGG pathways	Pubmed textmining	Protein domains	Cis-regulatory module	BLAST	Protein interactions	
1.	*CNIH*	*DACT1*	*BMP4*	*RTN1*	*BMP4*	*KIAA1344*	*BMP4*	*EXOC5*
2.		*DAAM1*	*PTGER2*	*DLG7*	*DAAM1*		*OTX2*	
3.	*KIAA1344*		*PTGDR*	*ARID4A*	*OTX2*	*ARID4A*	*WDHD1*	
4.	*CGRRF1*		*SOCS4*	*BMP4*	*KIAA0586*	*CDKN3*	*SOCS4*	*TIMM9*
5.	*DDHD1*	*STYX*	*DAAM1*	*PSMA3*		*SAMD4*	*DACT1*	*ERO1L*
6.	*ACTR10*	*KTN1*	*PSMC6*	*OTX2*		*STYX*	*SAMD4*	*PSMA3*
7.	*CDKN3*	*TIMM9*	*PSMA3*	*KTN1*	*SOCS4*	*FBXO34*		*BMP4*
8.	*RTN1*		*GNPNAT1*	*PSMC6*	*PSMC6*	*OTX2*	*RTN1*	*WDHD1*
9.	*FBXO34*		*TBPL2*	*WDHD1*	*WDHD1*	*PSMC6*	*KTN1*	*SOCS4*
10.		*CNIH*	*ERO1L*	*CNIH*	*KIAA1344*	*BMP4*	*FBXO34*	*KIAA1344*
11.	*PLEKHC1*		*GCH1*	*SOCS4*	*DACT1*	*KTN1*	*CDKN3*	*DACT1*
12.		*PSMA3*	*DDHD1*		*KTN1*	*PLEKHC1*	*DDHD1*	*OTX2*
13.		*PLEKHC1*	*WDHD1*	*STYX*	*ARID4A*			*DAAM1*
14.	*BMP4*	*SAMD4*	*KIAA1344*	*PLEKHC1*		*DACT1*		
15.	*GCH1*	*GMFB*	*DACT1*	*DAAM1*	*STYX*		*ERO1L*	
16.	*KTN1*	*DLG7*	*OTX2*	*FBXO34*	*SAMD4*	*GPR135*		
...		*ACTR10*		*PTGER2*	*DLG7*	*DAAM1*		*KTN1*
80.		...	...	...	...	...	...	...

BMP4
OTX2
DAAM1
WDHD1
KTN1
DACT1
ARID4A
SOCS4
SAMD4
KIAA1344
EXOC5
DLG7
PSMC6
STYX
...

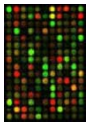
Congenital heart disorders

Congenital heart defect patient

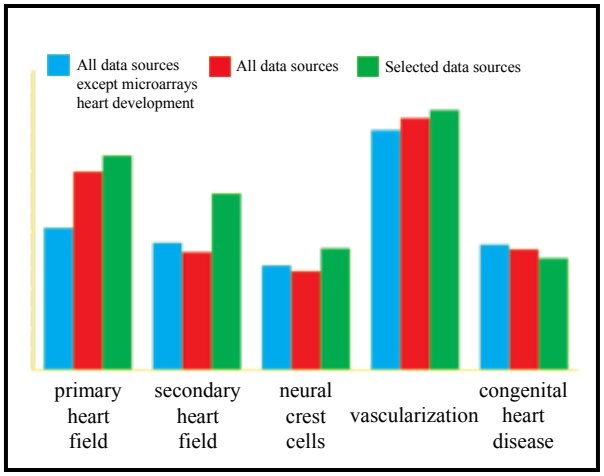


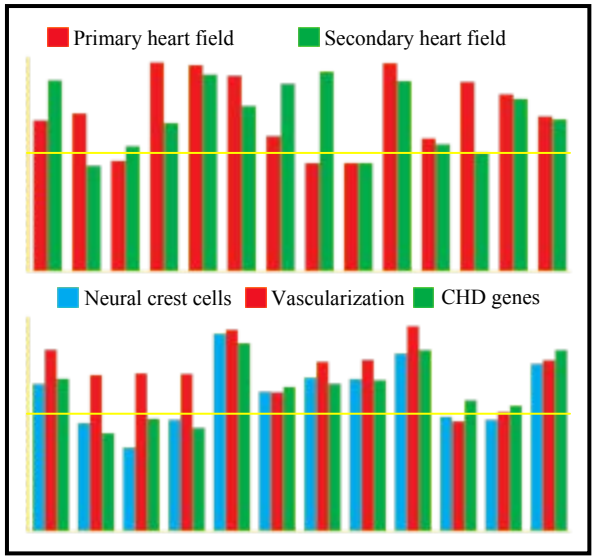
del(14q22.1-23.1)
56 candidate genes

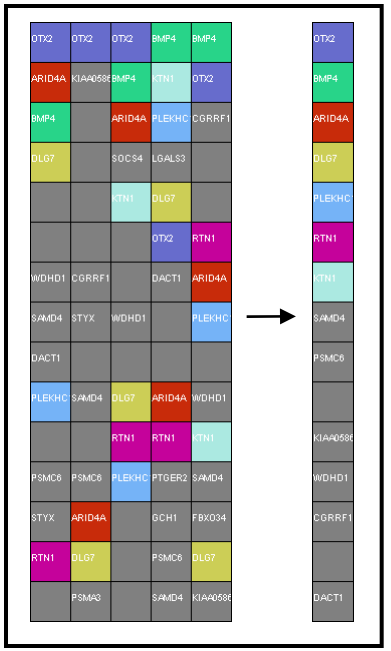
MA data embryonic heart development

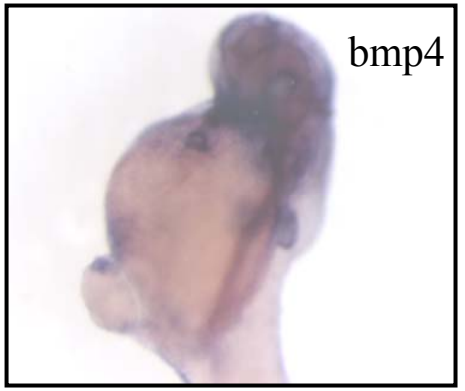


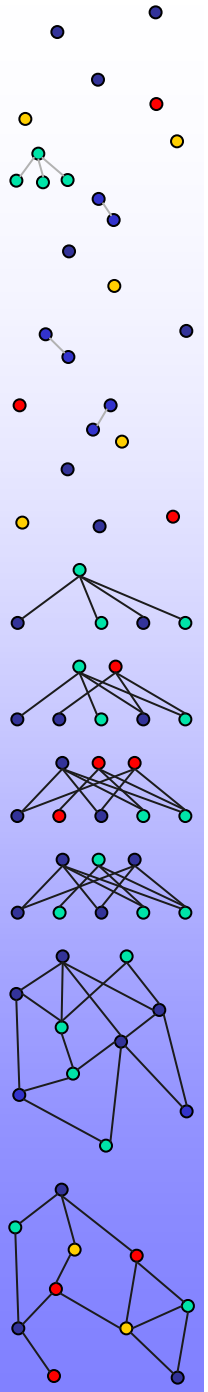
5 sets of training genes:
primary heart field
secondary heart field
neural crest cells
vascularization
congenital heart disease





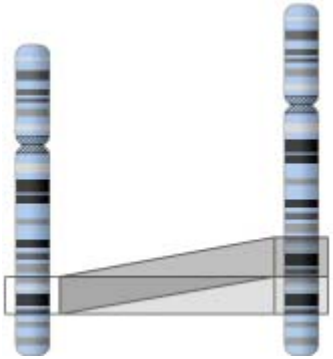






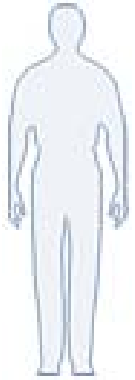
Prioritization by text mining

Prioritization by text mining



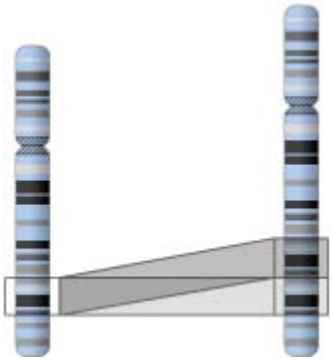
ABLIM1
ACSL5
ADD3
ADRA2A
ADRB1
CASP7
CSPG6
DCLRE1A
DUSP5
GFRA1
GPAM
GSTO1
HABP2
HSPA12A
MXI1
NHLRC2
NRAP
PDCD4
PNLIP
PNLIPRP1
RBM20
SHOC2
SLK
SMNDC1
SORCS1
TCF7L2
TDRD1
TECTB
TRUB1
VTI1A
VWA2
XPNPEP1
ZDHHC6

Microcephaly
Micrognathia
Low-set ears
Microphthalmia
Downslanting palpebral fissures
Hypertelorism
Long philtrum
Cleft lip
Short neck
Pectus excavatum
Syndactyly
Heart defects
Cryptorchidism
Mental retardation



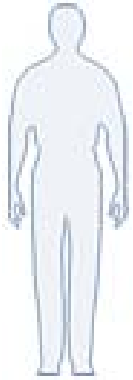
- *Steven Van Vooren in collaboration with Sanger Institute, Molecular Cytogenetics (N. Carter, H. Firth) and EBI text-mining group (D. Rebholz)*

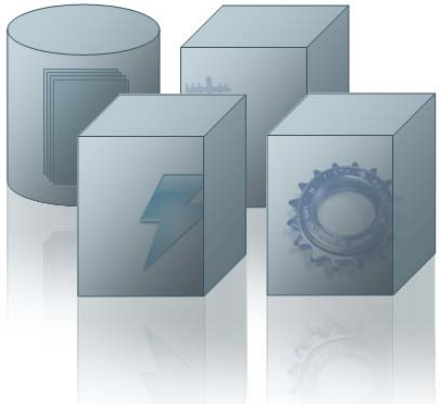
Prioritization by text mining



ABLIM1
ACSL5
ADD3
ADRA2A
ADRB1
CASP7
CSPG6
DCLRE1A
DUSP5
GFRA1
GPAM
GSTO1
HABP2
HSPA12A
MXI1
NHLRC2
NRAP
PDCD4
PNLIP
PNLIPRP1
RBM20
SHOC2
SLK
SMNDC1
SORCS1
TCF7L2
TDRD1
TECTB
TRUB1
VTI1A
VWA2
XPNPEP1
ZDHHC6

Microcephaly
Micrognathia
Low-set ears
Microphthalmia
Downslanting palpebral fissures
Hypertelorism
Long philtrum
Cleft lip
Short neck
Pectus excavatum
Syndactyly
Heart defects
Cryptorchidism
Mental retardation





1: Am J Med Genet A. 2004 Jun 1;127(2):197-200.

Mild Wolf-Hirschhorn phenotype and partial GH deficiency in a patient with a 4p terminal deletion.

Titomanlio L, Romano A, Conti A, Genesio R, Salerno M, De Brasi D, Nitsch L, Del Giudice E.

Department of Pediatrics, Child Neuropsychiatry Unit, Federico II University, Via S. Pansini 5, 80131 Naples, Italy.

Wolf-Hirschhorn syndrome (WHS) is caused by a variably-sized deletion of chromosome 4 involving band 4p16 whose typical craniofacial features are "Greek warrior helmet appearance" of the nose, microcephaly, and prominent glabella. Almost all patients show mental retardation and pre- and post-natal growth delay. Patient was born at term, after a pregnancy characterized by intra-uterine growth retardation (IUGR). Delivery was uneventful. Developmental delay was evident since the first months of life. At 2 years, he developed generalized tonic-clonic seizures. Because of short stature, low growth velocity and delayed bone age, at 4 years he underwent growth hormone (GH) evaluation. Peak GH after two provocative tests revealed a partial GH deficiency. Clinical observation at 7 years disclosed a distinctive facial appearance, with microcephaly, prominent eyes, and beaked nose. Brain MRI showed left temporal mesial sclerosis. GTG banded karyotype was normal. Because of mental retardation, subtelomeric fluorescence in situ hybridization (FISH) analysis was performed, disclosing a relatively large deletion involving 4p16.2 --> pter (about 4.5 Mb), in the proband, not present in the parents. The smallest deletion detected in a WHS patient thus far includes two candidate genes, WHSC1 and WHSC2. Interestingly, that patient did not show shortness of stature, and that could be due to the haploinsufficiency of other genes localized in the flanking regions. Contribution of GH alterations and possible GH therapy should be further considered in WHS patients.

PMID: 15108211 [PubMed - indexed for MEDLINE]

Related Links
- Early diagnosis of Wolf-Hirschhorn syndrome triggered by a life-threatening event: congenital diaphragmatic hernia. [Am J Med Genet A. 2004]
- The new Wolf-Hirschhorn syndrome critical region (WHSCR-2): a description of a second case. [Am J Med Genet A. 2005]
- Wolf-Hirschhorn syndrome with posterior intraorbital coloboma cyst: an unusual case. [Brain Dev. 2004]
- "Tandem" duplication of 4p16.1p16.3 chromosome region associated with 4p16.3pter molecular deletion resulting in Wolf-Hirschhorn [Am J Med Genet. 1999]
- The 4P-syndrome. Case description and literature review. [Minerva Pediatr. 2001]

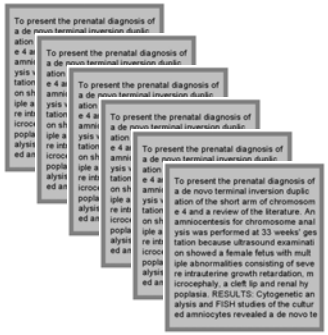
See all Related Articles...

Display AbstractPlus Show 20 Sort by Send to

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

Gene to concept association

ENSG00000000001
ENSG00000000002
...
WHSC1 **ENSG00000109685**
...
ENSG00000024999
ENSG00000025000

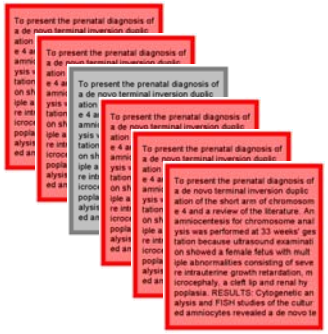


Microcephaly

PubMed

Gene to concept association

ENSG00000000001
ENSG00000000002
...
WHSC1 ENSG00000109685
...
ENSG00000024999
ENSG00000025000



Microcephaly

overrepresented in document set for WHSC1 gene

PubMed

Downstream Flank	22736020 (RP11-99L13)
Mean Ratio	-0.4
Origin of Altered Region	Unknown
Confirmation	unavailable
Interval	2765559
View in genomic context	e! cytoview

HUGO Gene Names	HUGO Gene Descriptions
TUBGCP5 (bp:20384869-20425331)	Description: tubulin, gamma complex associated protein 5. Aliases: GCP5, KIAA1899 [Ensembl:TUBGCP5] [OMIM:608147]
CYFIP1 (bp:20444104-20555043)	Description: cytoplasmic FMR1 interacting protein 1. Aliases: KIAA0068, P140SRA-1, SHYC [Ensembl:CYFIP1] [OMIM:606322]
NIPA2 (bp:20556790-20585849)	Description: non imprinted in Prader-Willi/Angelman syndrome 2. Aliases: [Ensembl:NIPA2] [OMIM:608146]
NIPA1 (bp:20594720-20637877)	Description: non imprinted in Prader-Willi/Angelman syndrome 1. Aliases: MGC35570 [Ensembl:NIPA1] [OMIM:608145]
MKRN3 (bp:21361547-21364653)	Description: makorin, ring finger protein, 3. Aliases: RNF63 [Ensembl:MKRN3] [GeneImprint:MKRN3] [OMIM:603856]
NDN (bp:21482492-21483457)	Description: necdin homolog (mouse). Aliases: HsT16328 [Ensembl:NDN] [GeneImprint:NDN] [OMIM:602117]

hide Prioritise genes by patient phenotype

Prioritisation by complete phenotype

Gene Id	Log score	traits involved	Literature evidence	Gene information
ENSG00000182636	75.03	6	(251 citations)	Necdin. [Source:Uniprot/SWISSPROT;Acc:Q99608]
ENSG00000170113	41.57	6	(14 citations)	Non-imprinted in Prader-Willi/Angelman syndrome region protein 1. [Source:Uniprot/SWISSPROT;Acc:Q7RTP0]
ENSG00000140157	40.85	6	(4 citations)	Non-imprinted in Prader-Willi/Angelman syndrome region protein 2. [Source:Uniprot/SWISSPROT;Acc:Q8N8Q9]
ENSG00000153575	39.78	6	(109 citations)	Gamma-tubulin complex component 5 (GCP-5). [Source:Uniprot/SWISSPROT;Acc:Q96RT8]
ENSG00000179455	29.49	4	(14 citations)	Makorin-3 (Zinc finger protein 127) (RING finger protein 63). [Source:Uniprot/SWISSPROT;Acc:Q13064]
ENSG00000068793	11.61	3	(4 citations)	cytoplasmic FMR1 interacting protein 1 isoform a [Source:RefSeq_peptide;Acc:NP_055423]

hide Prioritise Genes per phenotype trait

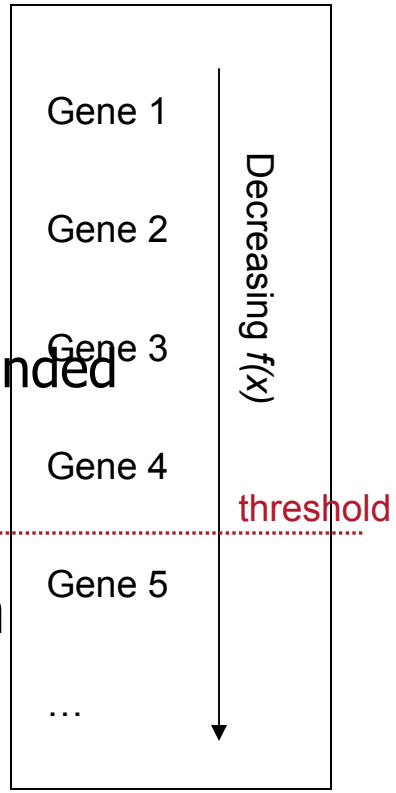
Prioritisation per individual phenotype

Autism/autistic behaviour

Gene Id	Log score	traits involved	Literature evidence	Gene information
ENSG00000182636	11.36	1	(3 citations)	Necdin. [Source:Uniprot/SWISSPROT;Acc:Q99608]
ENSG00000140157	5.64	1	(1 citation)	Non-imprinted in Prader-Willi/Angelman syndrome region protein 2. [Source:Uniprot/SWISSPROT;Acc:Q8N8Q9]
ENSG00000153575	5.42	1	(1 citation)	Gamma-tubulin complex component 5 (GCP-5). [Source:Uniprot/SWISSPROT;Acc:Q96RT8]
ENSG00000170113	4.39	1	(1 citation)	Non-imprinted in Prader-Willi/Angelman syndrome region protein 1.

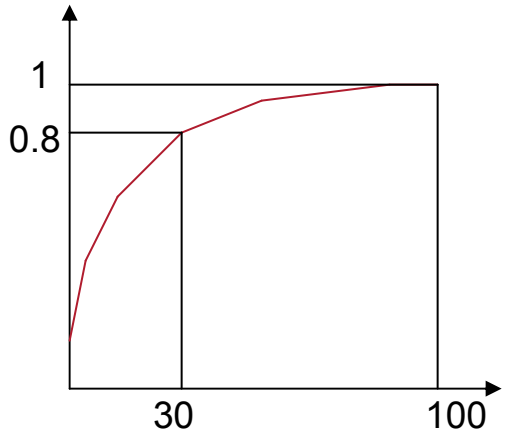
Statistical guarantees

- Theoretical guarantees:
  - Given a certain threshold on *f(x)*
  - Total number of genes *x* above it is upper bounded (positives)
  - Number disease genes *x* below it is upper bounded (false negatives)
- Often impractically loose
- Nevertheless: further backup of approach



Experimental results

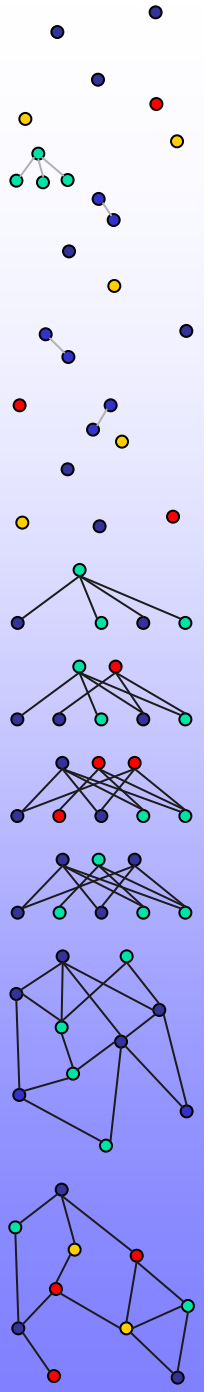
- For each disease:
  - 'Hide' one of the disease genes among 99 non-disease genes
  - Train based on remaining known disease genes
  - Compute rank of true disease gene (<100, >0)
- Do this for each disease gene and each disease
- Plot summary ROC curve



Performance measure:

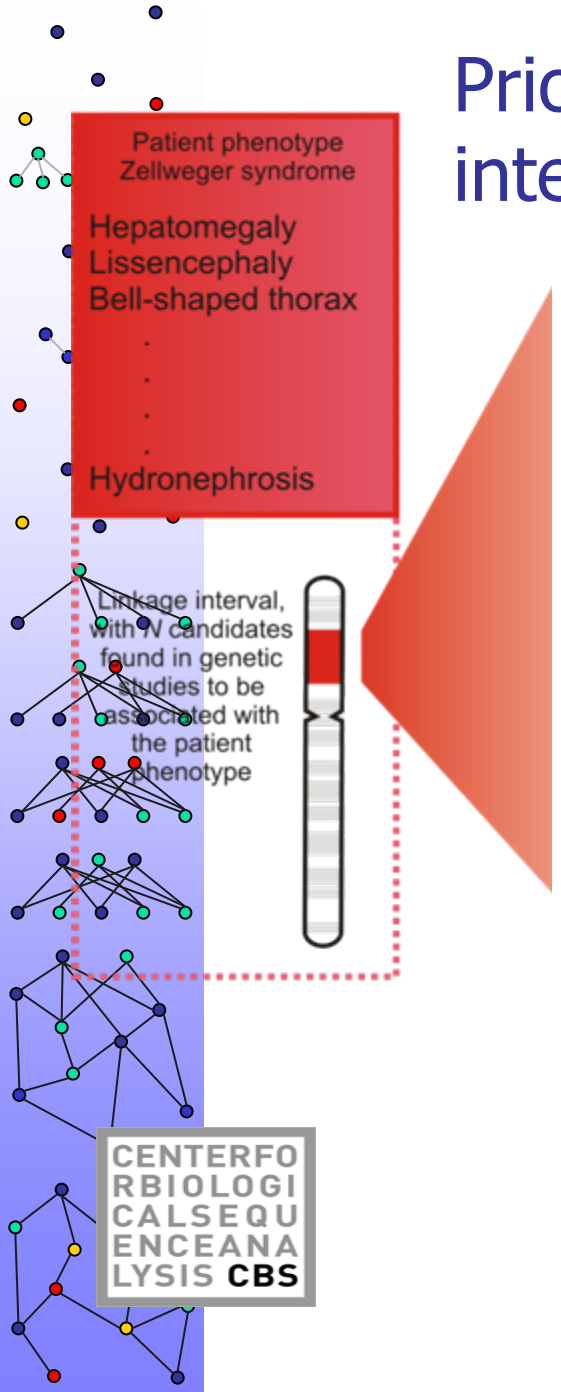
Area Under Curve (AUC)

or 1-AUC

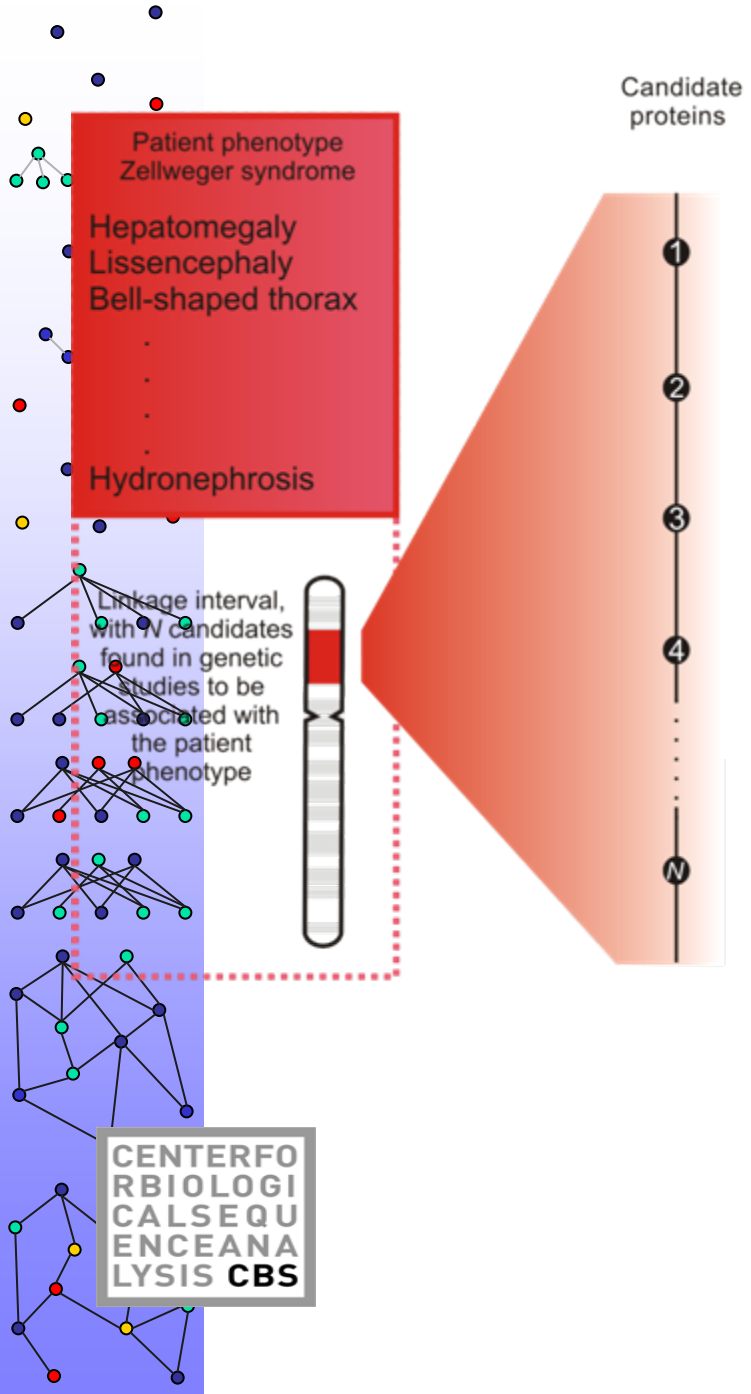


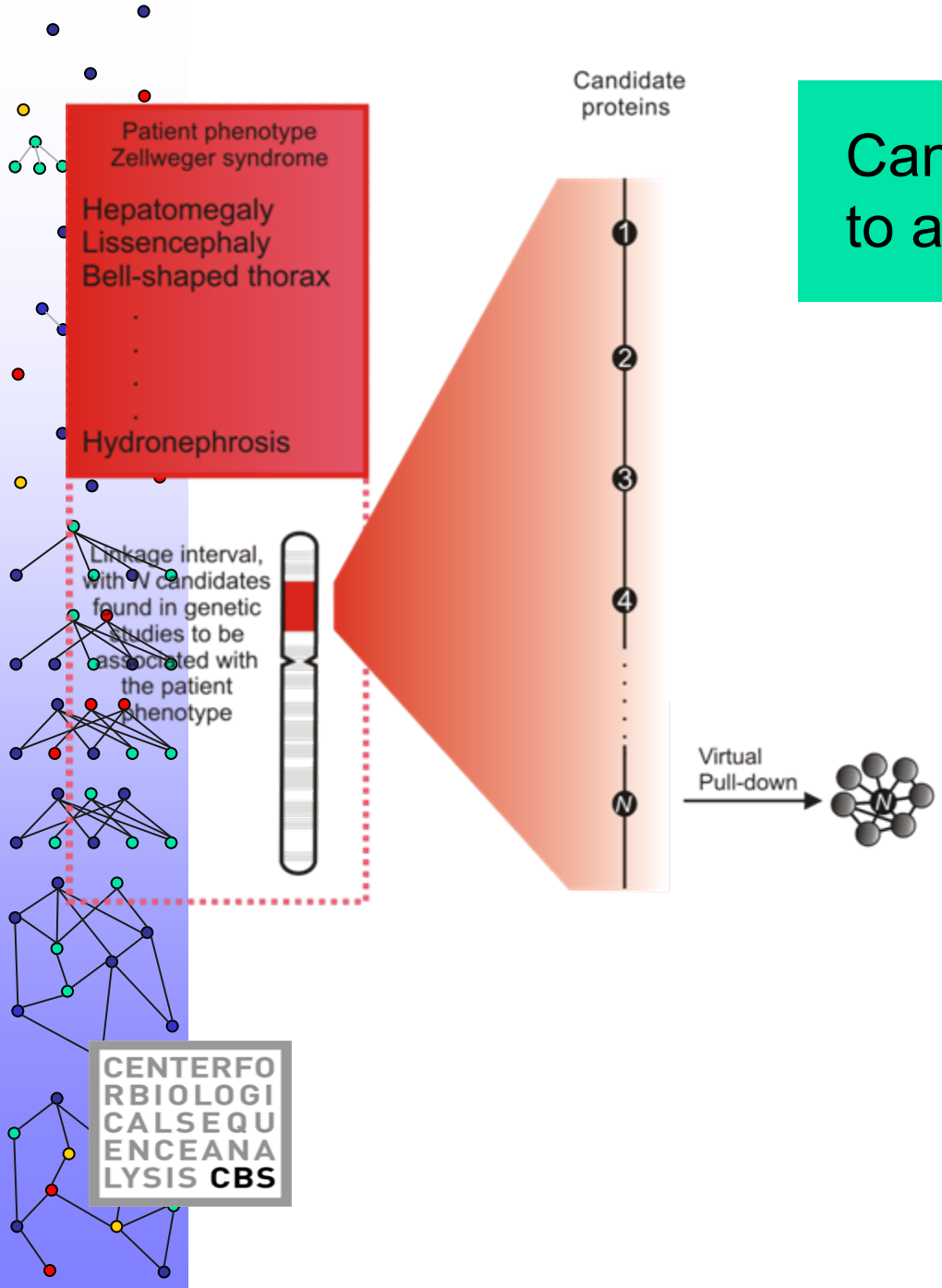
Prioritization by virtual pulldown

Prioritization by virtual protein-protein interaction pulldown and text mining



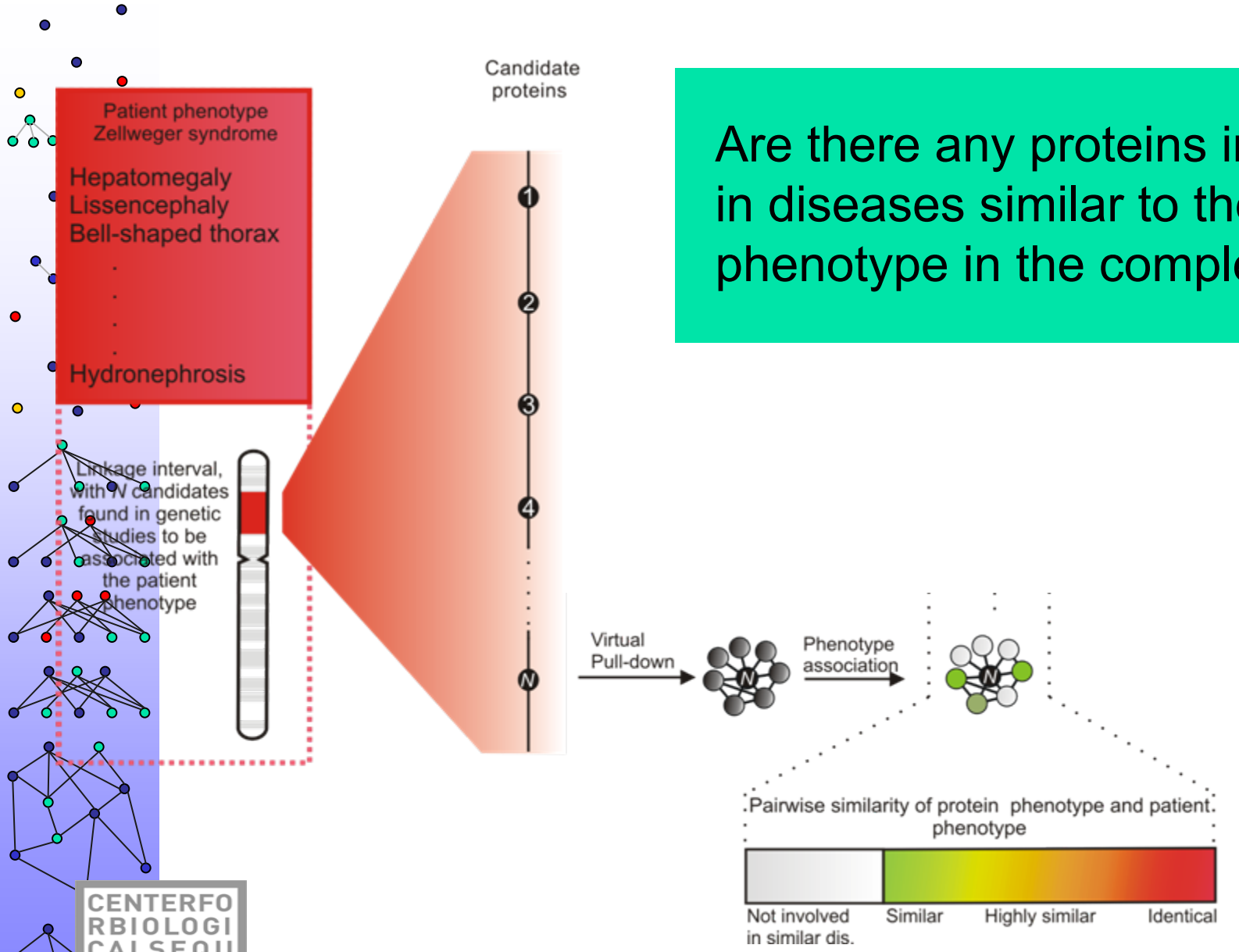
- *Lage et al. Nature Biotech. March 2007*


Candidate proteins
Patient phenotype
Zellweger syndrome
Hepatomegaly
Lissencephaly
Bell-shaped thorax
Hydronephrosis
Linkage interval, with N candidates found in genetic studies to be associated with the patient phenotype
1
2
3
4
N
CENTERFOR BIOLOGICAL SEQUENCE ANALYSIS CBS



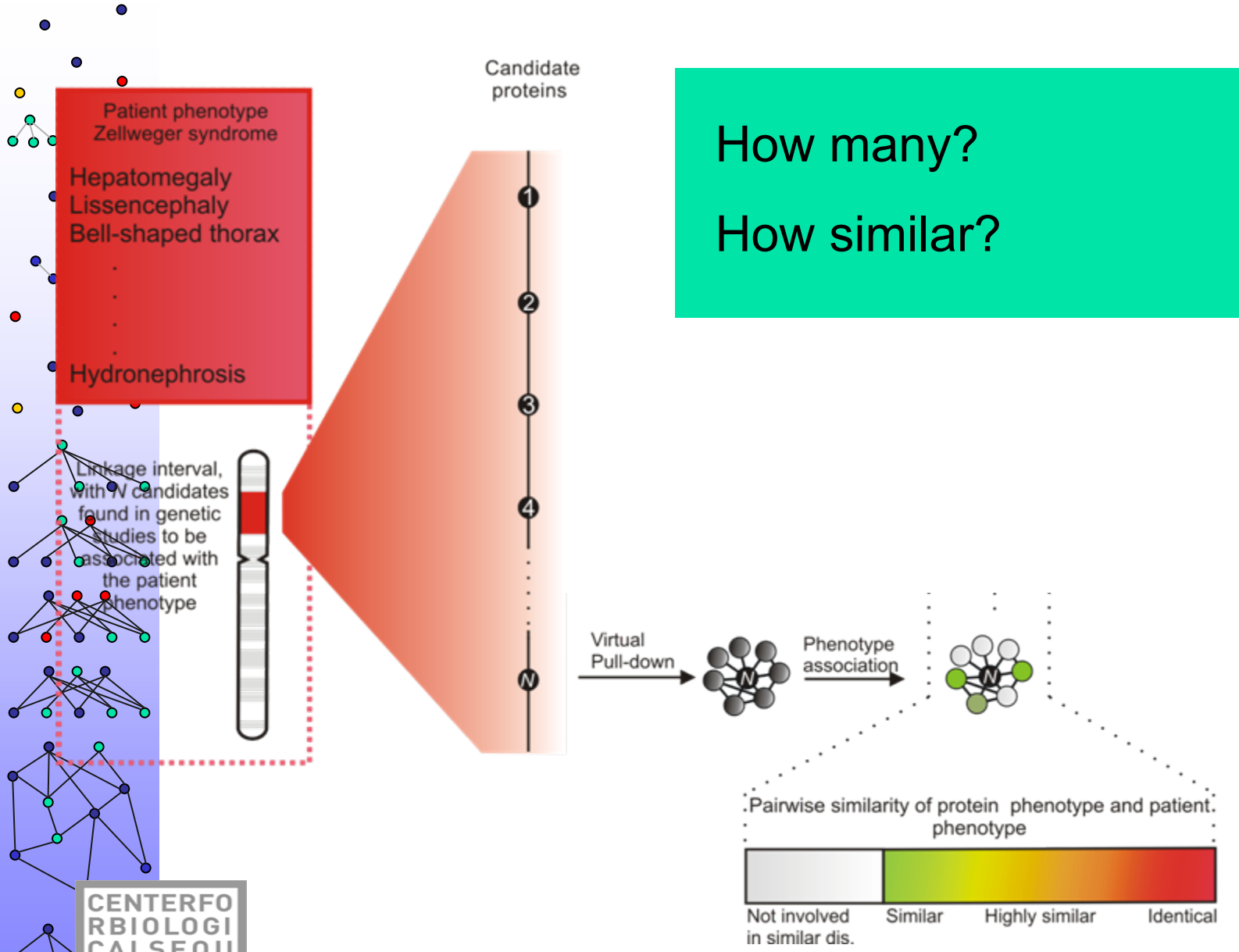
Can the candidate be assigned to a protein complex?

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS



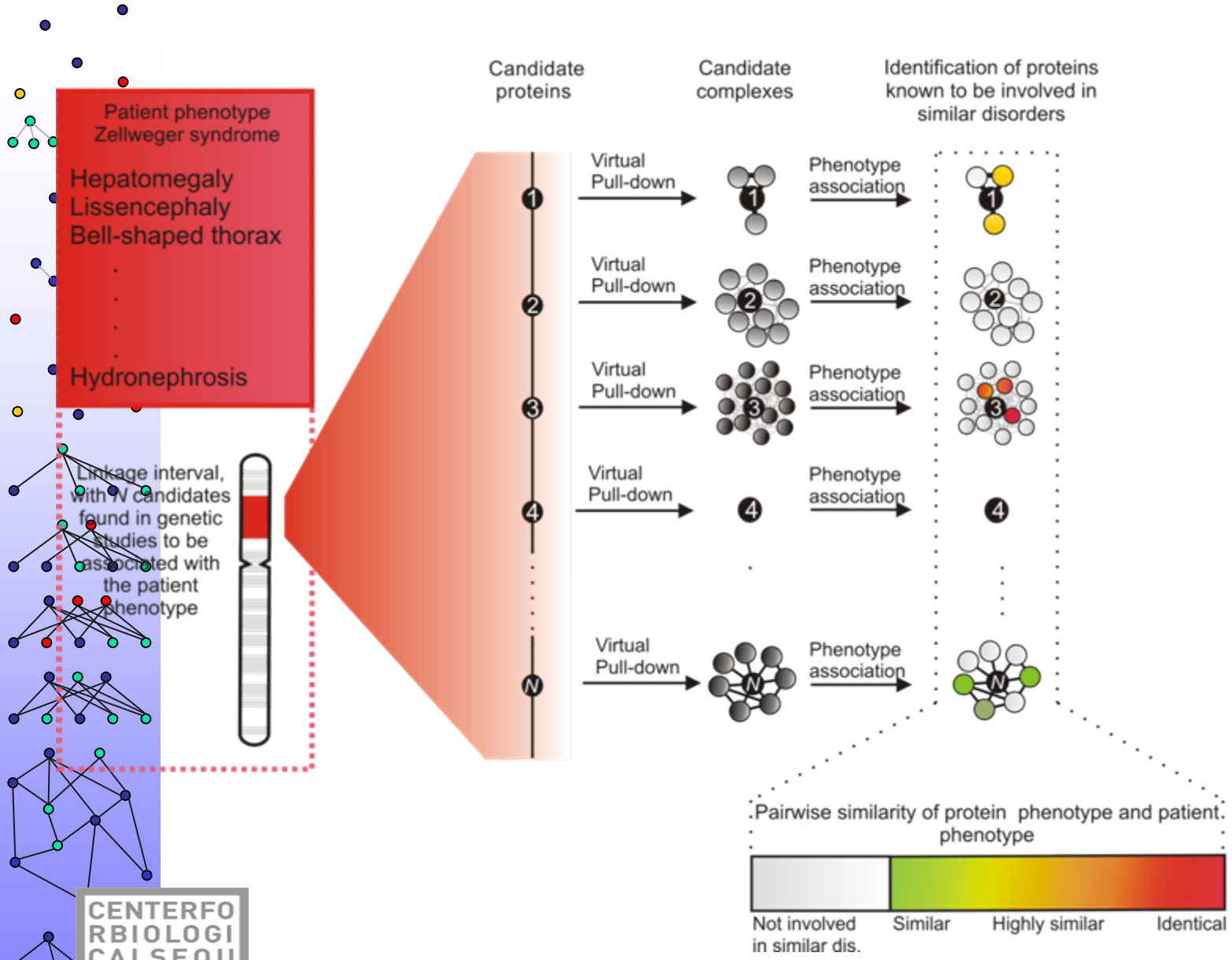
Are there any proteins involved in diseases similar to the patient phenotype in the complex?

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

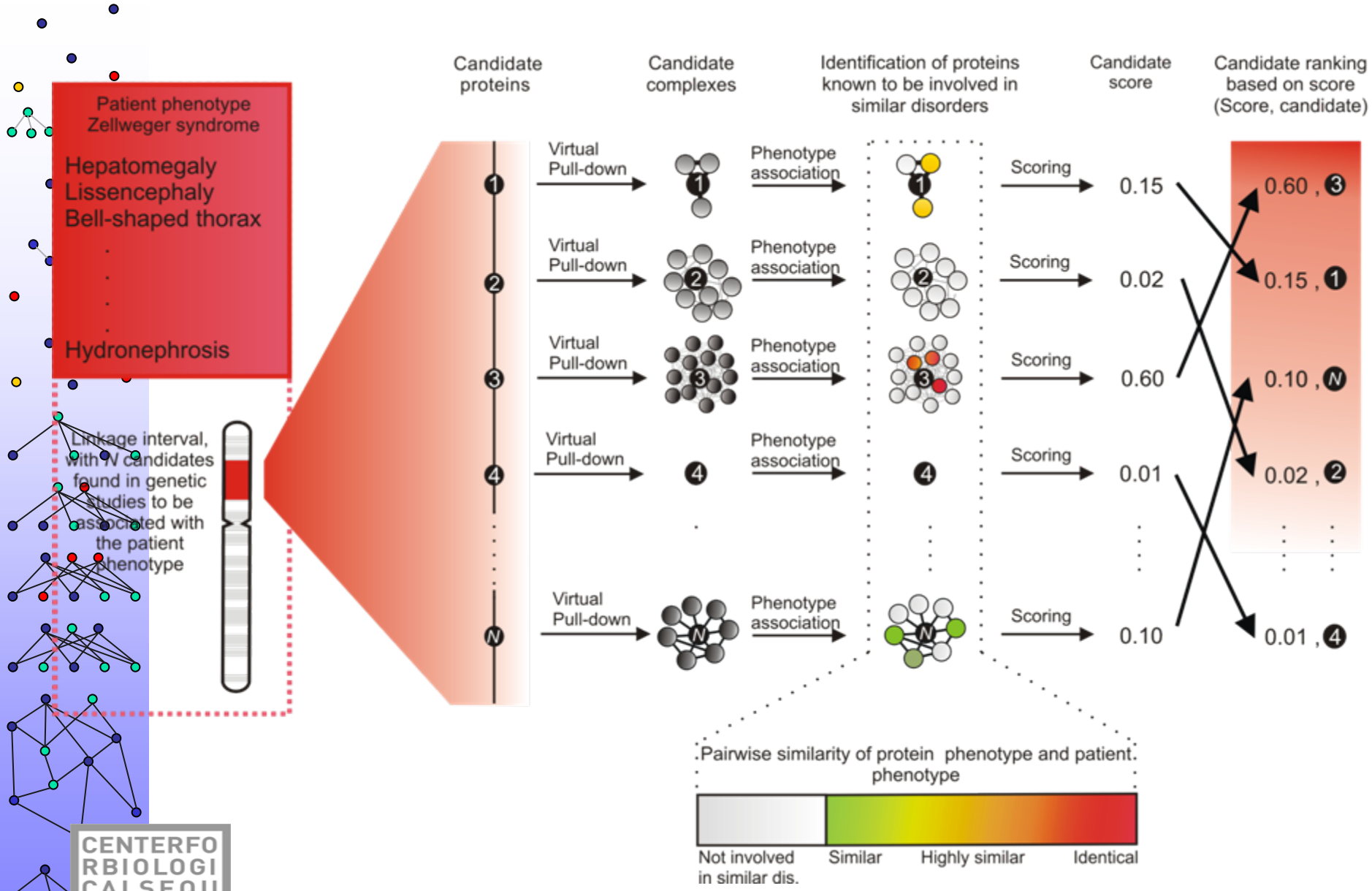


How many?

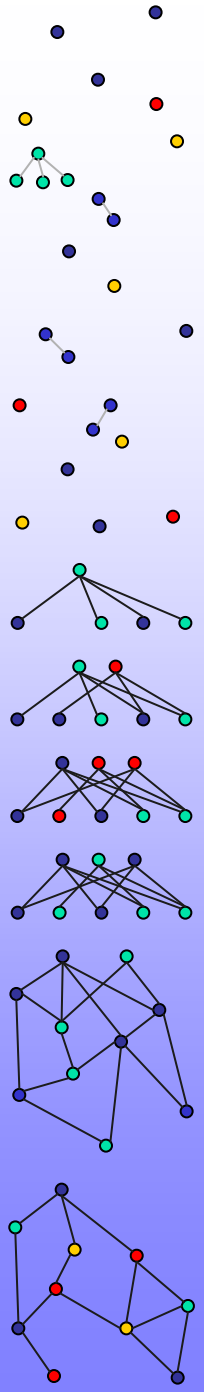
How similar?



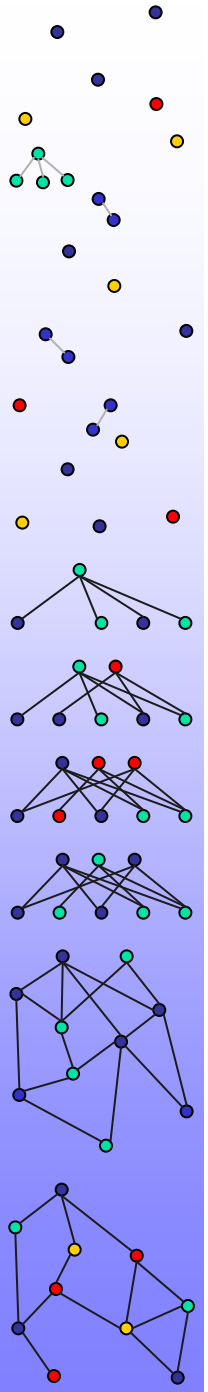
CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS



CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS



Prioritization by example



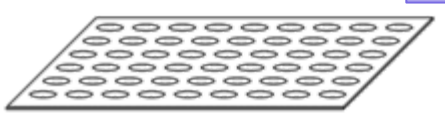
Prioritization by novelty detection

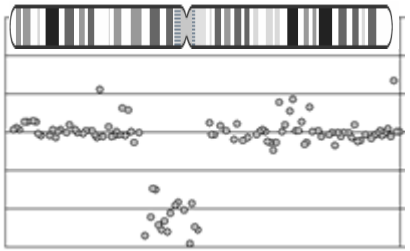
- Terminology:
  - Training set = disease-related genes
  - Test set = candidate genes
- Algorithm learns what makes a 'gene' a 'disease gene' based on the training set
- Test the learning algorithm on the test set, prioritize
- Rely on a vector representation of the genes

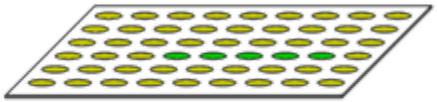
Array CGH

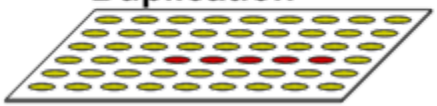
Child with e.g. heart defect and learning disabilities

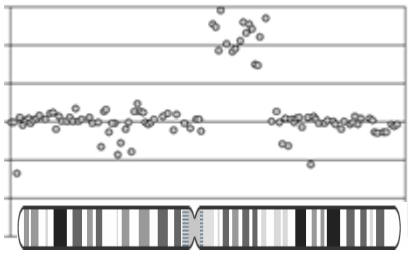
Sample is collected and sent to genetic center











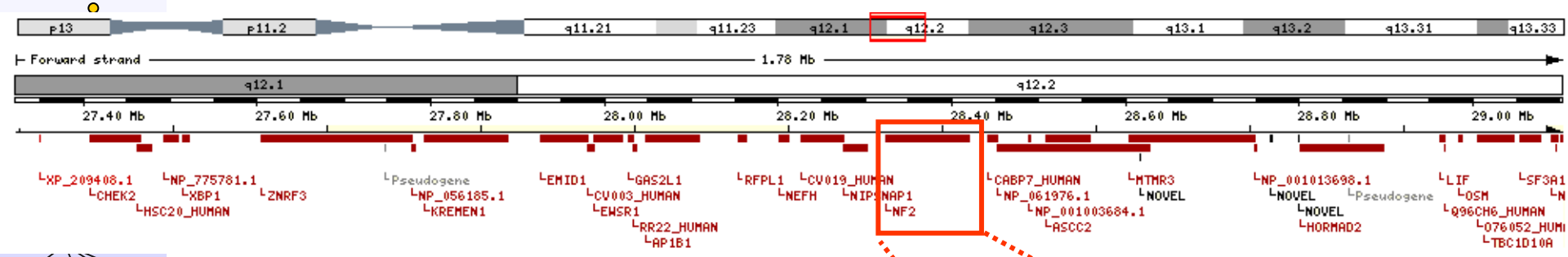
Cytogenetic diagnostic

- 2-3% of live birth with major congenital anomaly
  - 15-25% recognized genetic causes
  - 8-12% environmental factors
  - 20-25% multifactorial
  - 40-60% unknown
    - *15-20% of those resolved by array CGH*
- Importance of diagnosis
  - Usually limited therapeutic impact BUT
  - Reduce family distress
    - End of "diagnostic odyssey"
  - Estimate risk of recurrence
    - De novo aberration vs. familial mutation
  - Knowledge of disorder evolution (life planning)
  - Prevent complications
  - Future therapies (e.g., fragile-X, Rett + gene therapy)

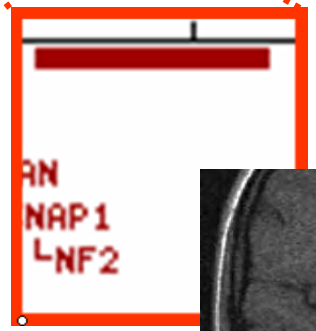
Deletion del(22)(q12.2)

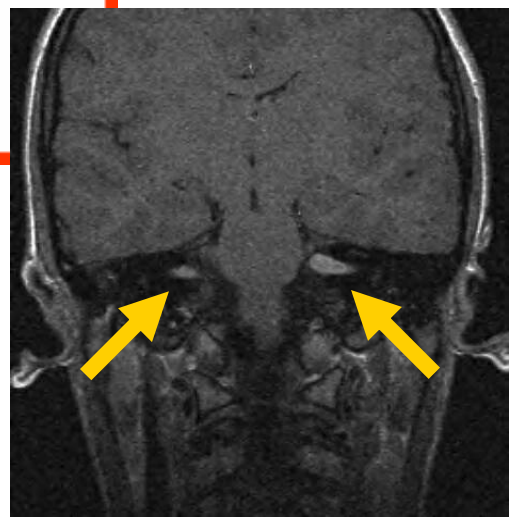
- Patient
  - Pulmonary valve stenosis
  - Cleft uvula
  - Mild dysmorphism
  - Mild learning difficulties
  - High myopia

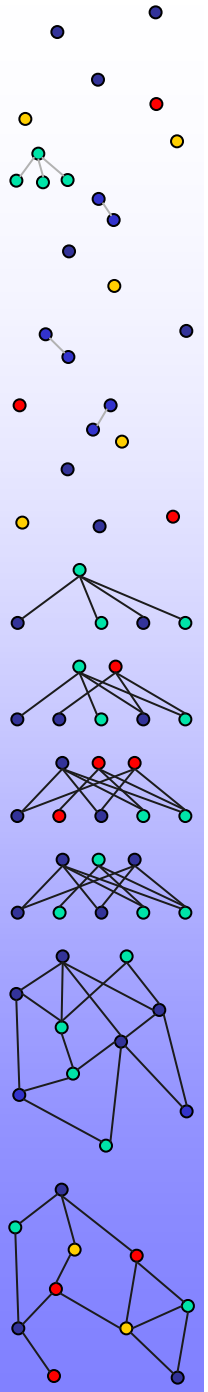
Deletion del(22)(q12.2)



- Deletion on Chromosome 22
  - ~0.8Mb
- Deletion contains NF2
  - NF2 ↔ acoustic neurinomas
  - Benign tumor, BUT
    - Hard to diagnose
    - Severe complications







Array CGH: from diagnosis to gene discovery

1. Processing of array CGH data
2. Databasing and mining of patient descriptions
3. Genotype-phenotype correlation
4. Candidate gene prioritization
5. Experimental validation of candidate genes

Genotype-phenotype correlation

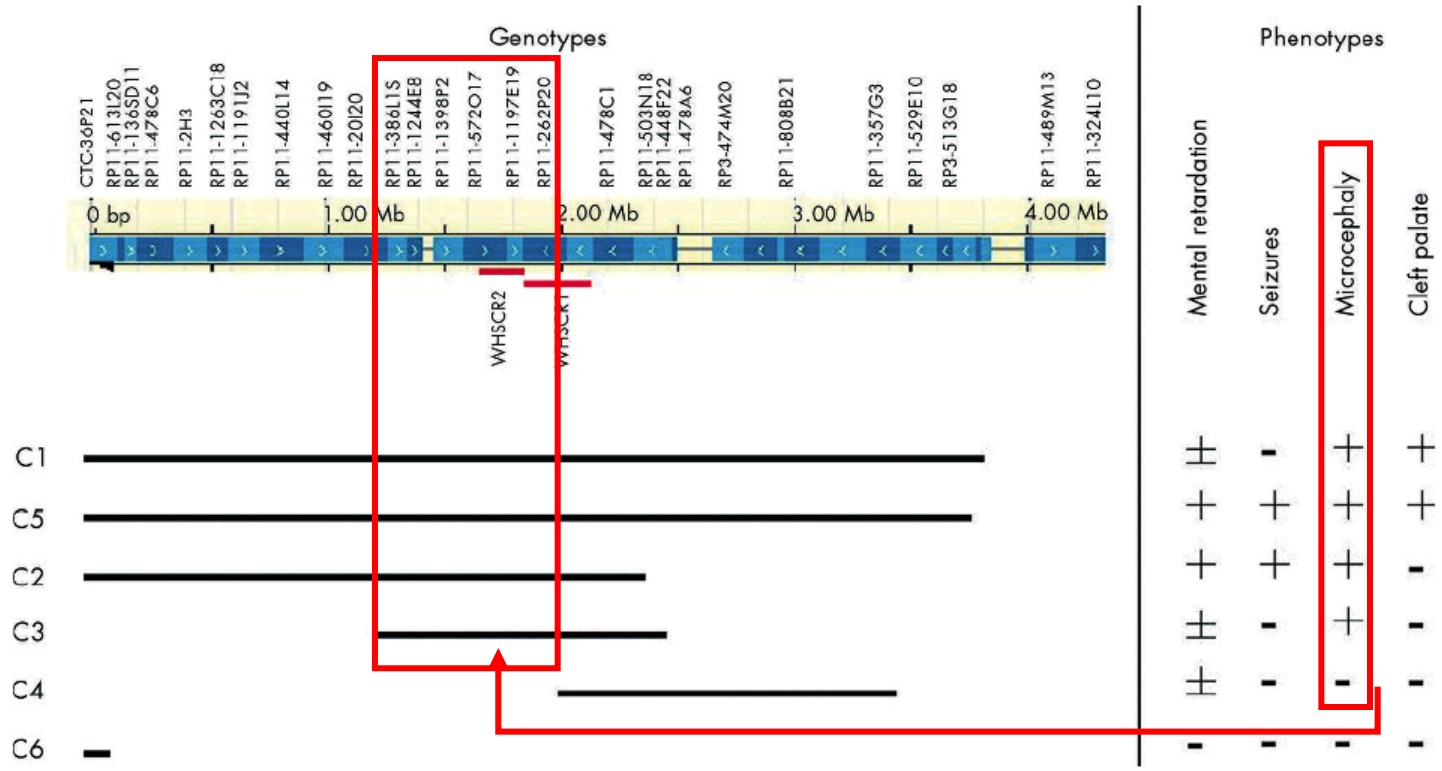
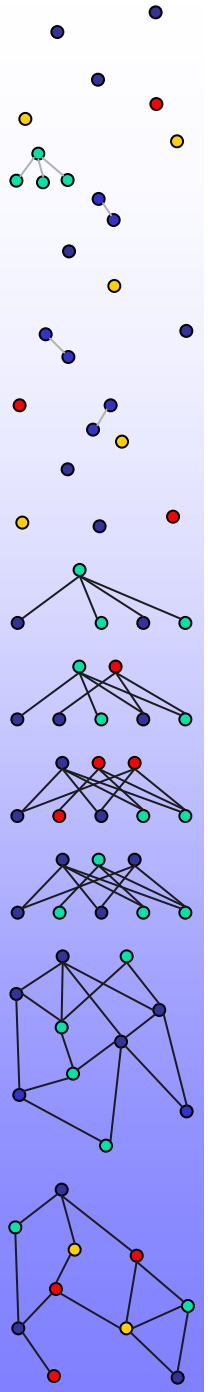
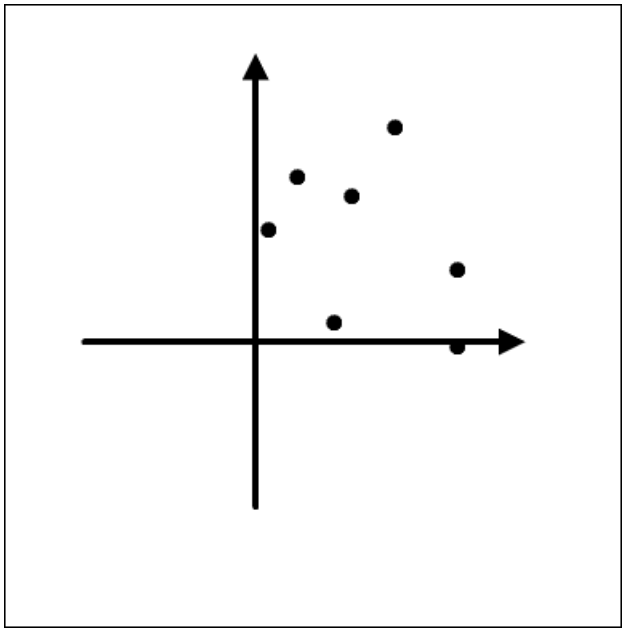


Figure 1 Genotypes and phenotypes of the patients analysed in this study. The top part shows the clones represented on the array from the telomeric 4 Mb together with the DNA contig representation of Ensembl (01/2004). Clones in italics are not represented in the Golden Path sequence. The Wolf-Hirschhorn critical regions WHSCR1 and WHSCR2 are indicated with the lines under the Ensembl contig representation. The bottom shows a summary of the genotypes of all the patients analysed in this study. The lines indicate the sizes of the 4p deletions. On the right, the main phenotypic features discussed in the text are presented.



Kernel-based novelty detection

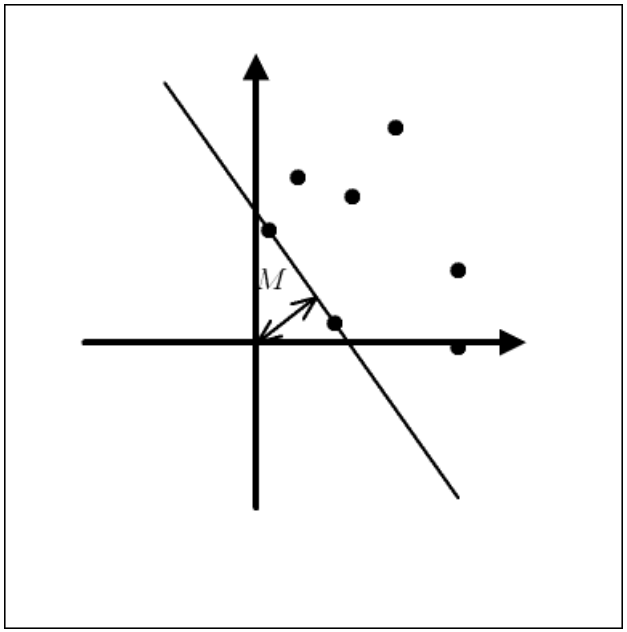
Prioritization as machine learning



- Training set = disease-related genes
- Test set = candidate genes

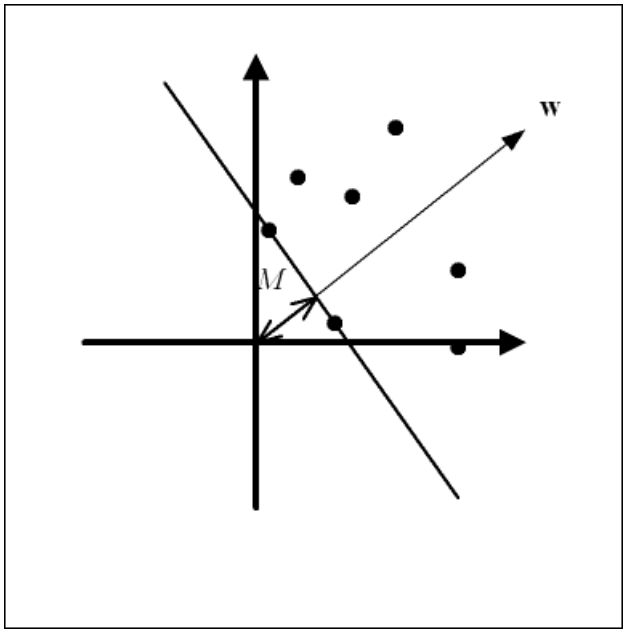
- Represent all training genes in a vector space
  - Expression data, vector space model for text, sequence, etc.
  - Potentially very high-dimensional
- Identification of *negative* examples not straightforward

Kernel-based novelty detection



- Formulate problem as novelty detection
  - Does not use *negative* examples
- Find a *hyperplane* separating these from origin
- The further (the larger M), the more *homogeneous* the training set

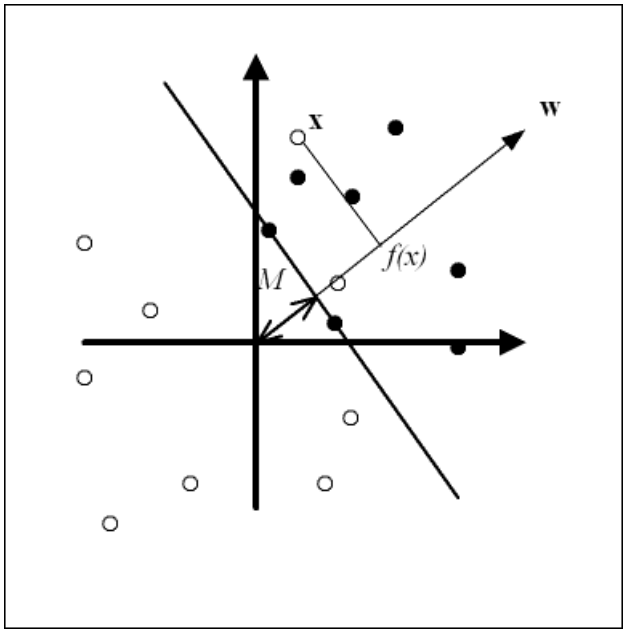
Kernel-based novelty detection



- Hyperplane is parameterized by a (unit norm) *weight vector w*
- Optimization problem

$$\max_w M$$
$$\Leftrightarrow \max_w (\min_i w'x_i)$$
$$\Leftrightarrow \max_{w,M} M \quad \text{s.t.} \quad M \leq w'x_i$$

Kernel-based novelty detection



- Further from origin along w → more 'like a disease gene'
- Scoring function:

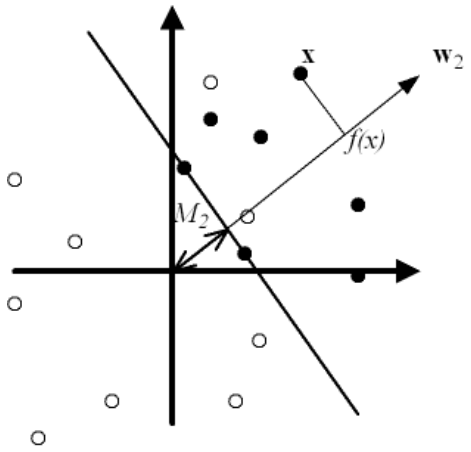
$$f(x) = w'x$$

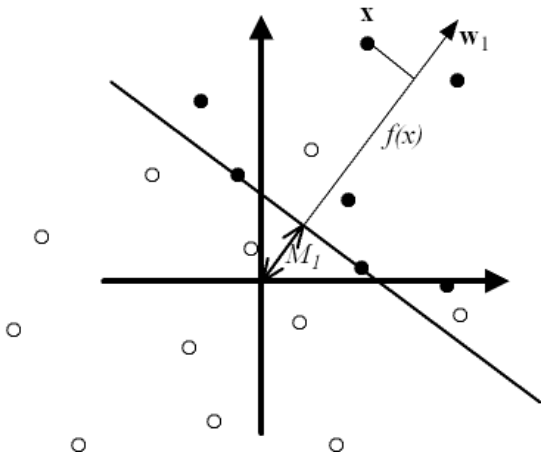
= distance from origin along w

- Sort in decreasing value of f
- Genes "similar" to training genes will rank highly

Which representation, which similarity?

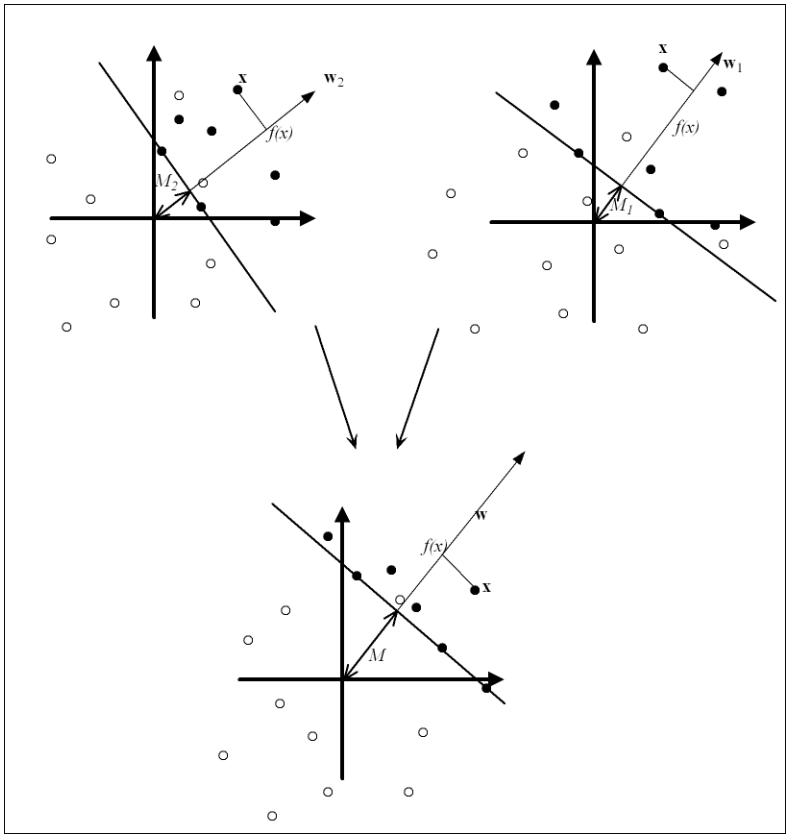
- *Representation is arbitrary*
  - Sequence, expression, interaction, annotation…
  - Which one to use? Select the one with largest M?
- Perhaps we can integrate!





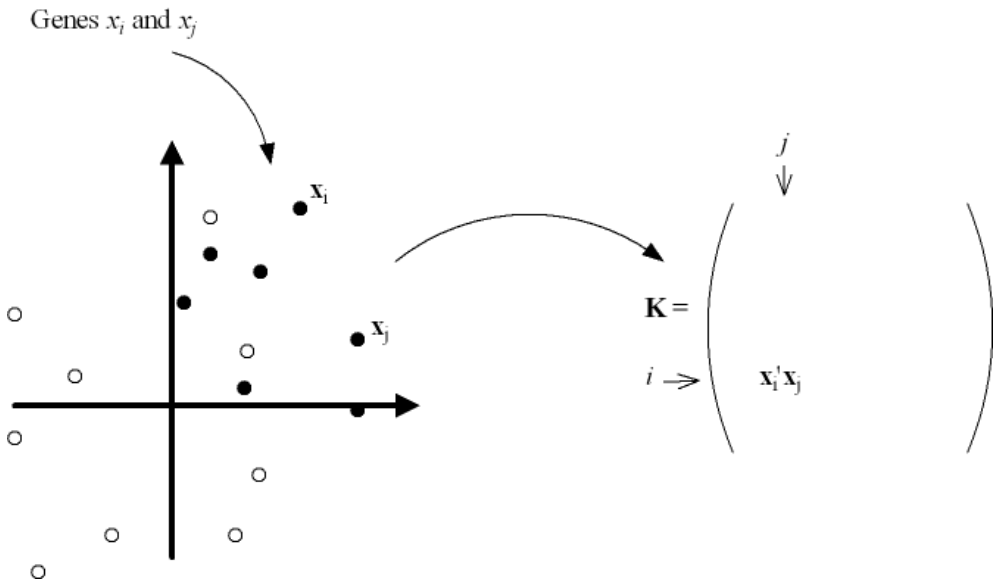
Kernel-based data fusion

- Given two or more vector representations
- Integrate into one vector representation…
  … such that *training set is maximally coherent* (i.e., M as large as possible)



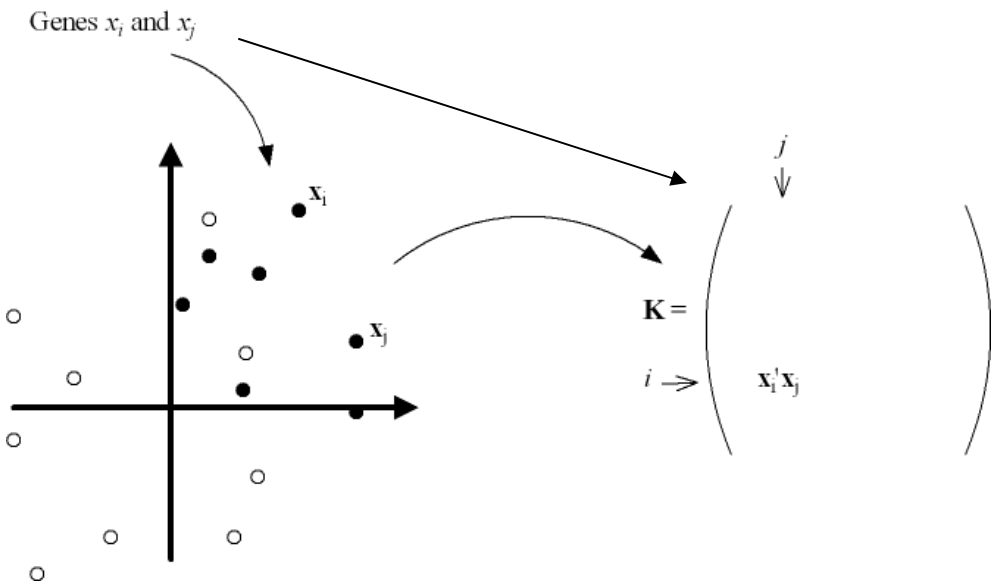
The kernel trick

- Kernel methods ideally suited for this…
- Represent vectors indirectly, by means of *all pairwise inner products*
- Inner product matrix = *kernel matrix* K
- Contains inner product $K_{i,j}=x_i'x_j$ at position (i,j)



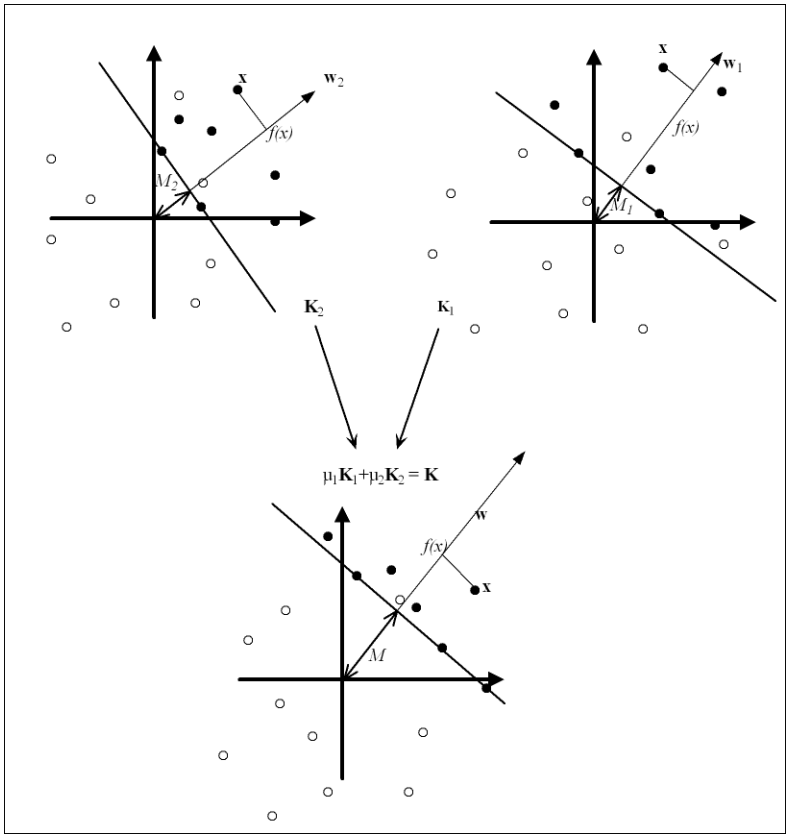
The kernel trick

- Inner product (kernel) = *measure of similarity*
- Often easier to specify than the vector representation
- Vector representation is implicit, no need to make explicit, since …
- … kernel is *sufficient to compute* w *and* $f(x)$



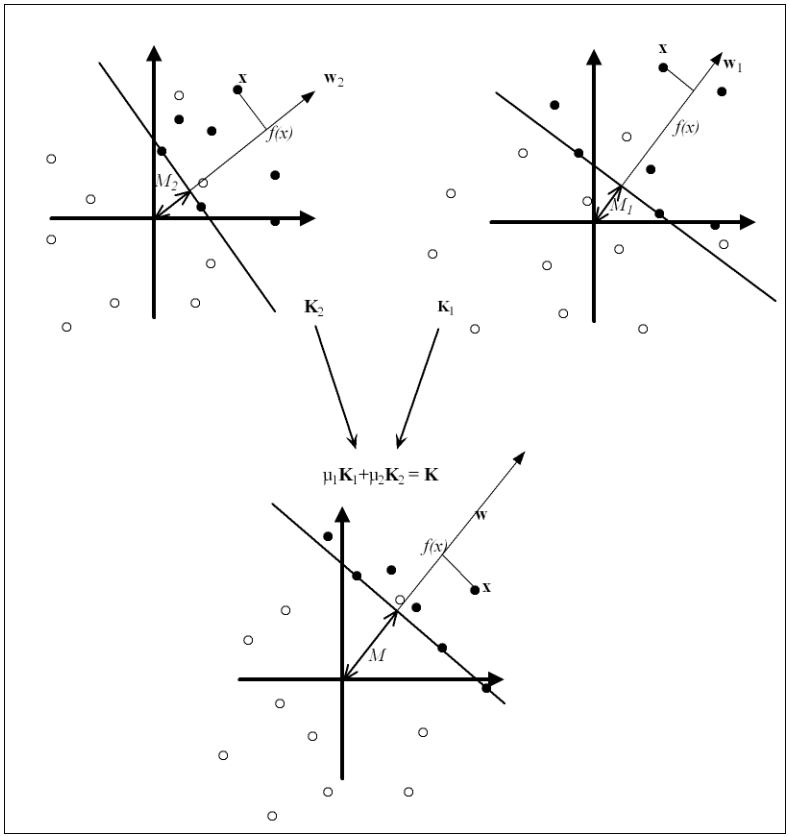
Kernel-based data fusion

- For each gene representation j, a *kernel matrix* K_j
- Given m kernels K_j
- Compute one integrating kernel as
  $K = \mu_1 K_1 + \ldots + \mu_m K_m$
  (e.g., Lanckriet et al., *Bioinformatics* 2004)
- μ_j?



Kernel-based data fusion

- How to choose μ_j?
- Such that M is maximal: $\max_{\mu j,w} \min_i w'x_i$
- μ_j *guided by the data!*
- Efficient *convex optimization* problem (~seconds)
- Efficient $f(x)$ evaluation

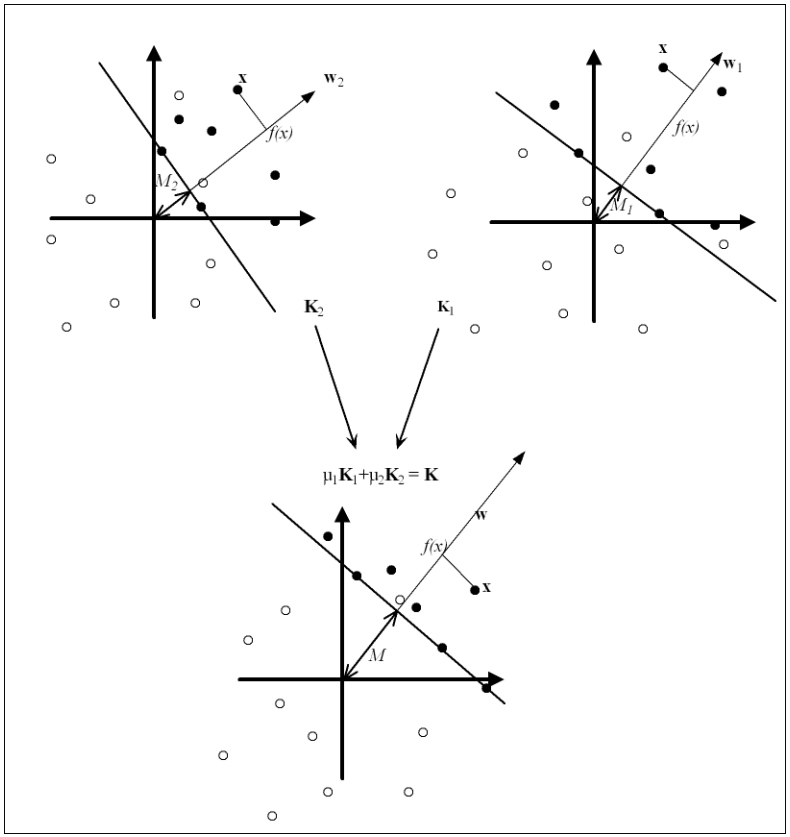


Kernel-based data fusion

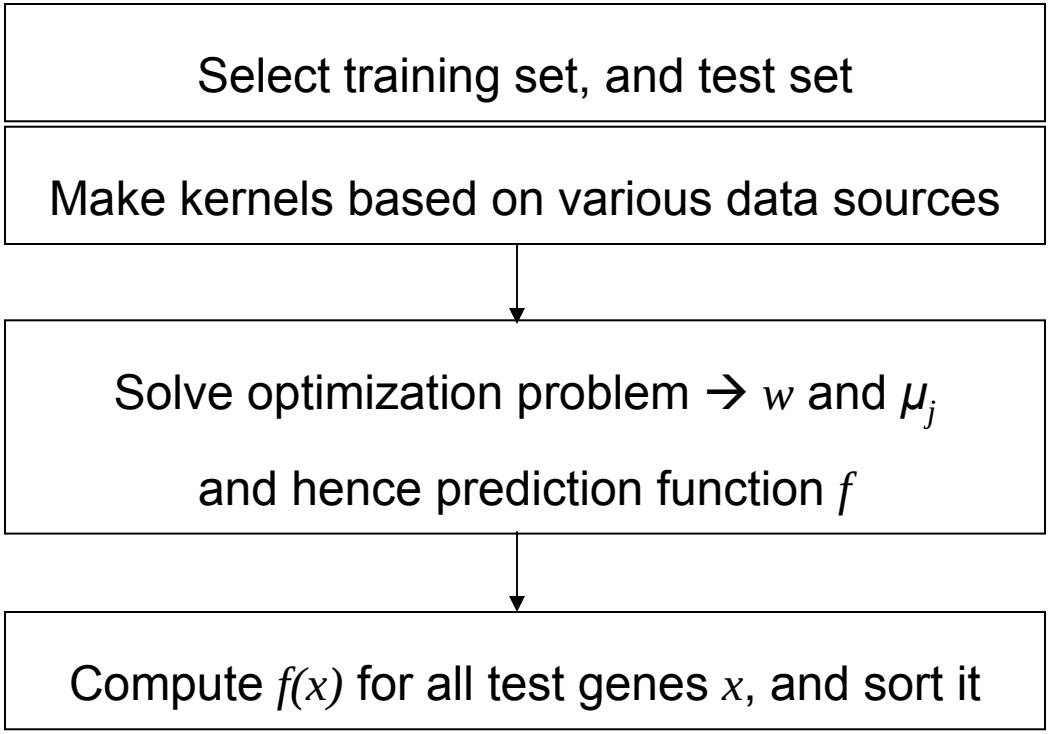
- Optimization problem

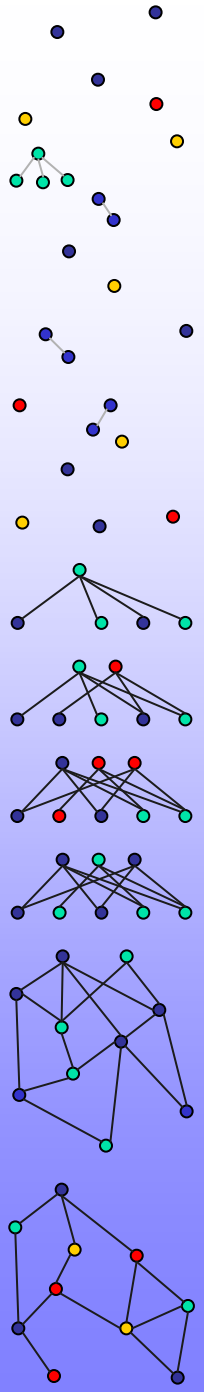
$$\max_{\mu j,w} \min_i w'x_i$$

- Risk of overfitting with large number of kernels
- *Regularization:* impose lower bound on the μ_j
  - All kernels contribute at least a bit



Global strategy



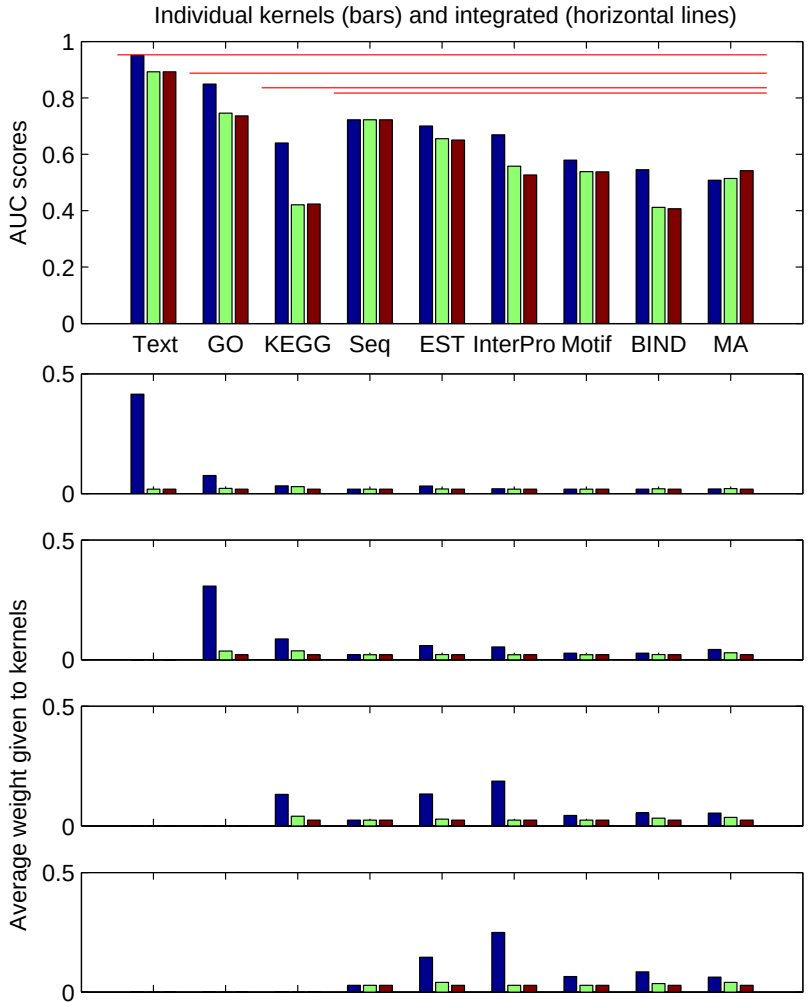


Experimental results

- 29 diseases (same as in ENDEAVOUR paper)
  - Between 4 and 113 genes associated to each
- 9 data sources used
  - Text, GO, KEGG, Seq, EST, InterPro, Motif, BIND, MA
- 3 kernels per source (corresponding to different vector representations)
- Sources evaluated separately, after fusion, and in presence of noise

Experimental results

- Performs well for data sources separately
- Integration performs better than individual data sources



Experimental results

1-AUC	All	No Text	No Text, GO	No Text, GO, KEGG
ENDEAVOUR	0.0833	0.1290	0.1698	0.1698
Kernel method	**0.0686**	**0.1043**	**0.1491**	**0.1675**
p-value	7.4e-10	7.5e-11	3.3e-7	2.4e-1

- Performs better than ENDEAVOUR
  - Significantly so
  - Also faster (at run-time)

Experimental results

1-AUC		$\mu_{\min} = 0$	$\mu_{\min} = 0.5$	$\mu_{\min} = 1$
All data sources	No noise	0.0505	**0.0477**	0.0686
	4× noise	0.0596	**0.0579**	0.0950
	8× noise	0.0656	**0.0644**	0.1144
	16× noise	0.0702	**0.0694**	0.1420
No Text	No noise	0.1241	0.1121	**0.1043**
	4× noise	0.1411	**0.1330**	0.1395
	8× noise	0.1520	**0.1444**	0.1629
	16× noise	0.1624	**0.1566**	0.1943
No Text, no GO	No noise	0.1902	0.1644	**0.1491**
	4× noise	0.2186	0.2034	**0.2005**
	8× noise	0.2375	**0.2257**	0.2275
	16× noise	0.2554	**0.2496**	0.2599
No Text, no GO, no KEGG	No noise	0.2121	0.1828	**0.1675**
	4× noise	0.2410	**0.2245**	0.2296
	8× noise	0.2626	**0.2500**	0.2612
	16× noise	0.2825	**0.2770**	0.2963

- For different levels of regularization
- Different features used
- Different amounts of noise

Endeavour architecture