

RediscoveryDB :

Connecting high-throughput data
with Testable hypotheses

Dawei Lin, Ph.D. et al.

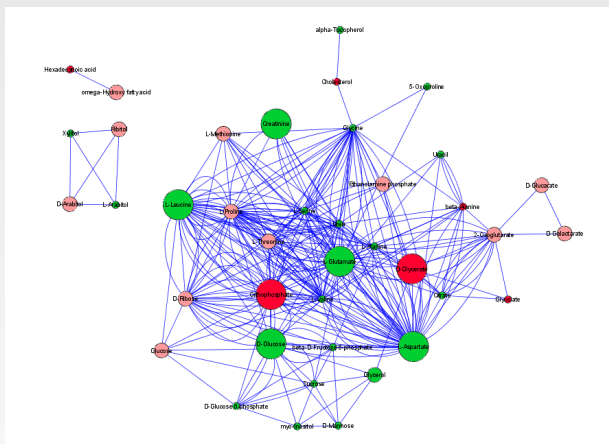
Director of Bioinformatics Core
Genome Center, UC Davis

Jan 15, 2010, GMOD 2010

The Goal of Bioinformatics is to Generate Testable Hypotheses

	retention_l	quantmass	PubChem	WT	redAA	WT	redAA	WT	redAA	WT	redAA	WT	redAA
1													
2													
3													
4													
5	threonic acid	497167	292	199262	5262933	3271	3402	4084	6936	4561	5003		
6	isoleucic acid	787359	117	199195	5261	25443	21146	25480	22393	28060	19396		
7	phosphoric acid	344674	299	199165	1004	238181	275149	299409	231676	295226	230659		
8	pentadecanoic acid	674704	132	203296	13849	144	189	285	113	401	103		
9	pelargonic acid	399163	117	201810	9159	1000	1213	2295	1072	1903	1139		
10	palmitoleic acid	790299	129	203265	5202745	2476	434	1737	695	1417	1451		
11	palmitic acid	713542	313	199173	985	7498	4327	6992	4498	6812	5530		
12	oxoprolin	486399	156	199593	7405	66652	84204	67103	93562	66705	118010		
13	methanophosphate	290941	241	199609	13130	1026	895	1169	723	979	770		
14	malate	462908	233	199166	525	2334	2225	1711	2056	2081	2081		
15	inositol myo-	723867	305	199164	892	31231	33572	47161	31684	39549	38010		
16	heptadecanoic acid	759645	117	200399	10465	537	527	690	394	843	377		
17	glycolic acid	221407	177	202167	357	178	231	408	272	477	243		
18	glycine	369041	174	199588	750	51285	66936	48601	48406	41763	34812		
19	glycerolphosphate alpha	591357	259	199419	754	11086	11122	10200	9695	9864	9185		
20	glucose 1	649920	160	199413	5793	12796	12378	13846	15614	9703	11391		
21	erythronic acid lactone	407545	147	200445	6427169	2121	2454	2918	1918	2626	2048		
22	cholesterol	1077630	129	199589	5957	72830	59988	63078	59639	77430	69956		
23	inoleic acid	777102	337	199240	5280450	805	184	449	288	577	466		
24	citric acid	617289	273	199168	311	6026	5140	5168	5629	5643	5133		
25	lumarate	399999	245	199199	444972	1633	1054	1974	2041	2018	1854		
26	fructose 2	642325	307	199181	5984	71	91	74	66	33	62		
27	2-hydroxybutanoic acid	278929	191	199774	11266	8673	6765	12145	7316	7316	2174		
28	threonine 2TMS	350345	117	199780	6288	5308	8163	5935	6011	4614	5571		
29	threonine	499403	117	199626	6288	722	940	698	699	493	760		
30	fructose	915457	271	202121	5968	45	22	22	38	46	18		
31	stigmastanol	1105016	129	199250	5280794	610	580	503	575	567	582		

Metabolomics Data



Pathway Mapping and Statistical Analysis

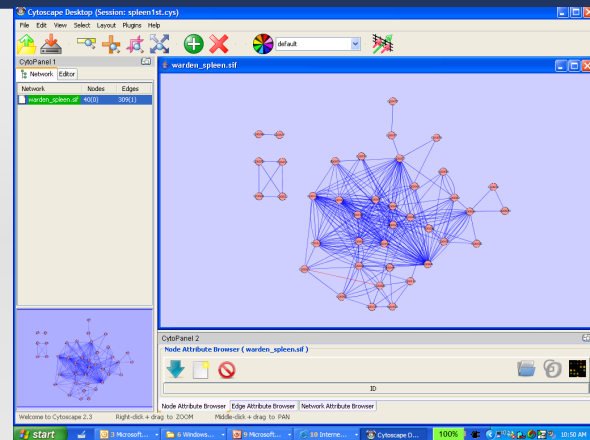


Color: Red: up-regulated, Blue: down-regulated
 Node Size: Big: high confidence, Small: Low confidence
 Edges: two metabolites share same pathway
 Edge color: same color same pathway

1

2

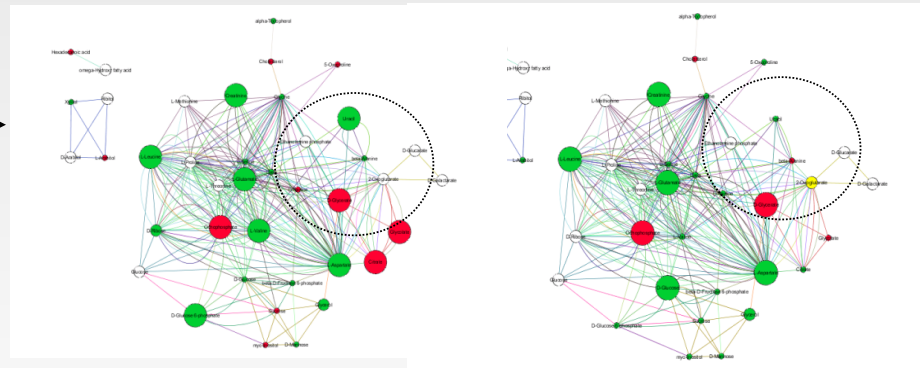
3



Network Construction

Knockout gene 1

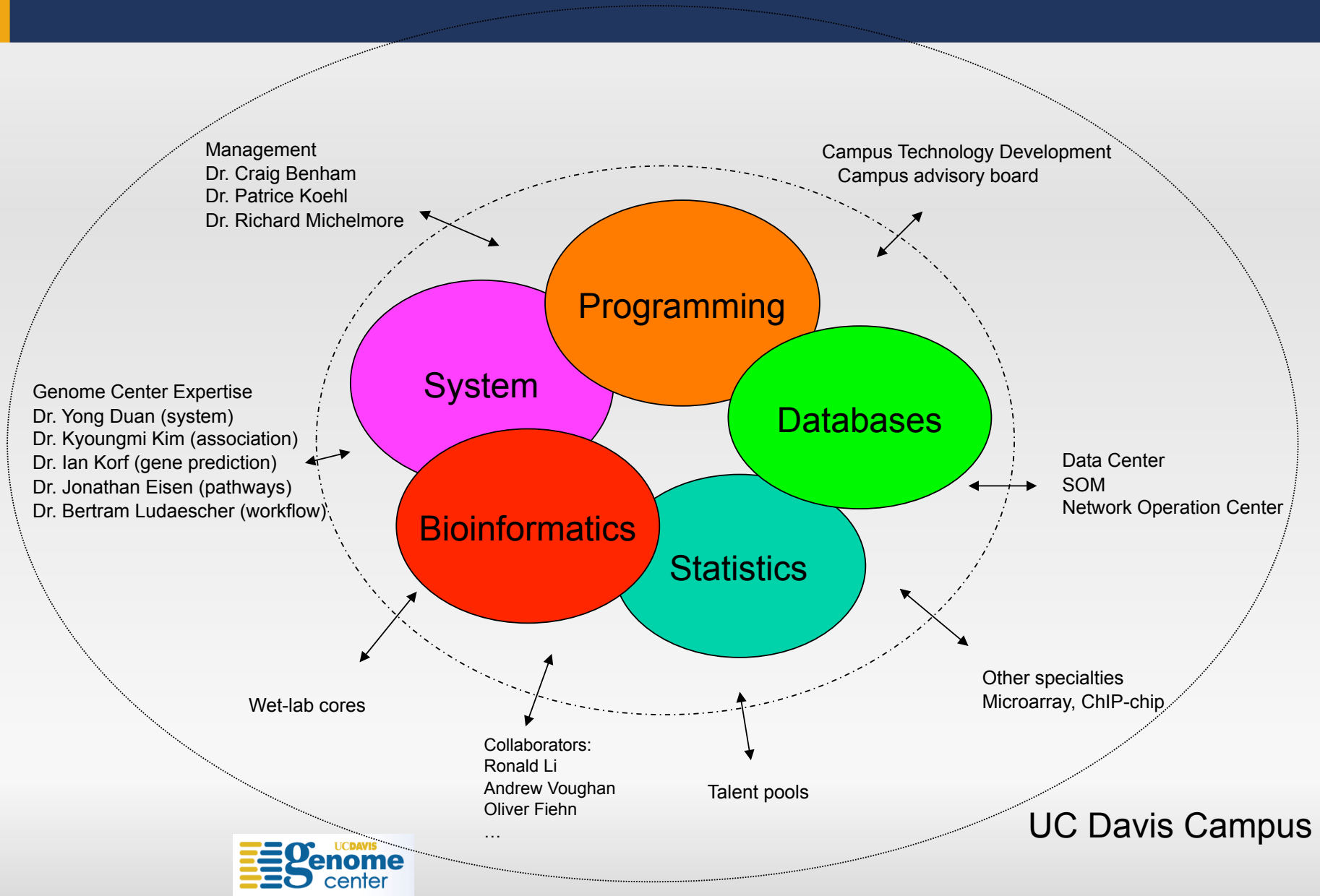
Knockout gene 2



Pathway Comparison

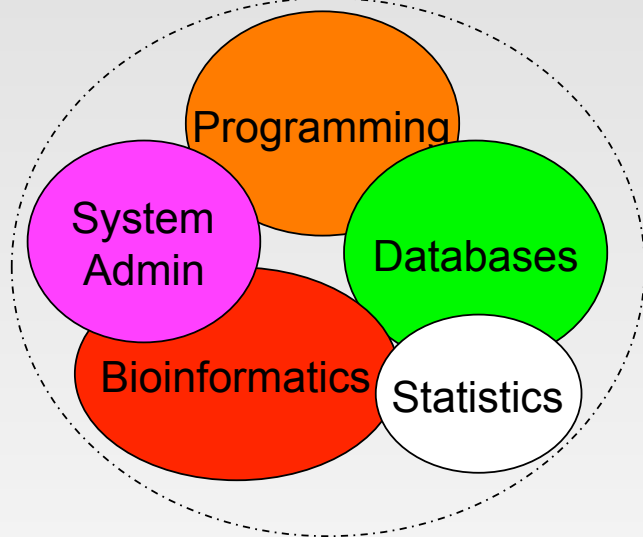
Craig Warden
 Oliver Fiehn

Bioinformatis 2.0



Example: Center for Molecular and Genomic Imaging (CMGI)

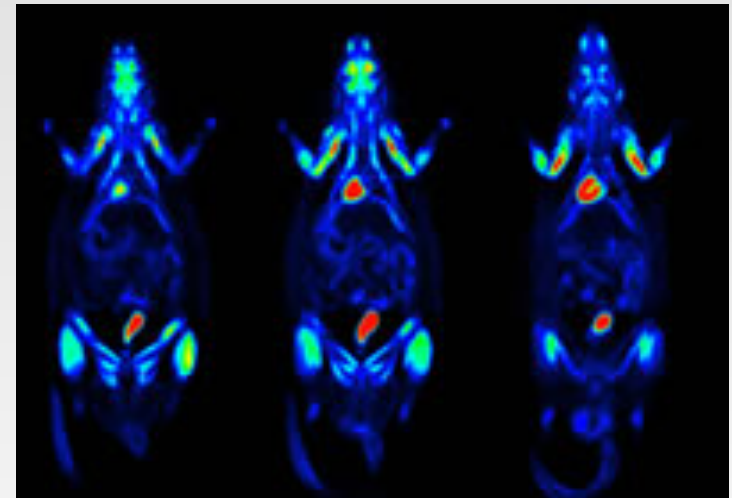
Bioinformatics Core



Predictable revenue
Critical mass of staff
Skill redundancy

Virtual member
←
Recharge

Imaging Center



No hiring or training
Access to diverse skills
No management
No interruption

Bioinformatics Core Members

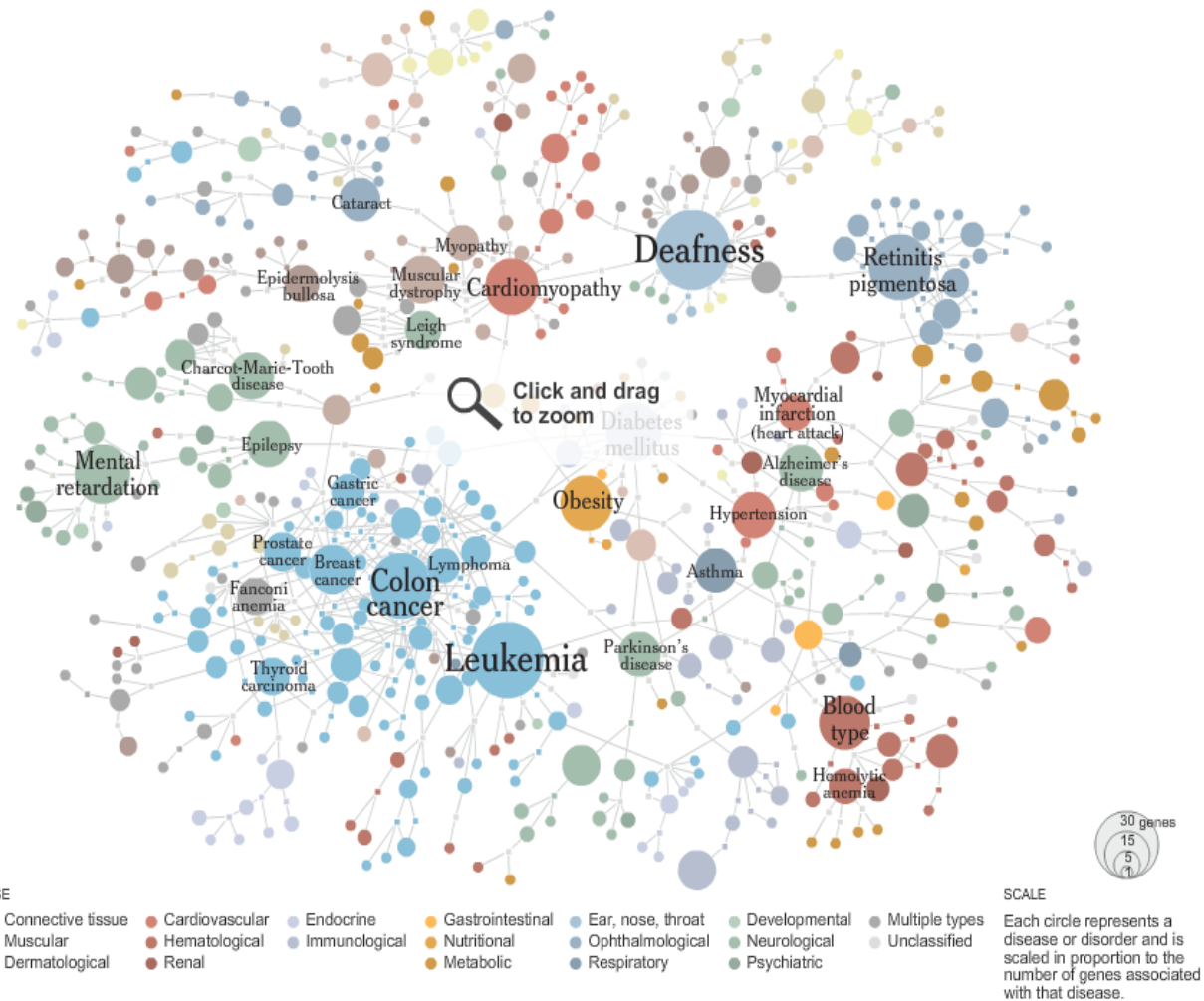


Dawei Lin, Vince Buffalo, Joe Fass, Monica Britton, Jose Boveda, Nikhil Joshi, Zhiwei Lu



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>

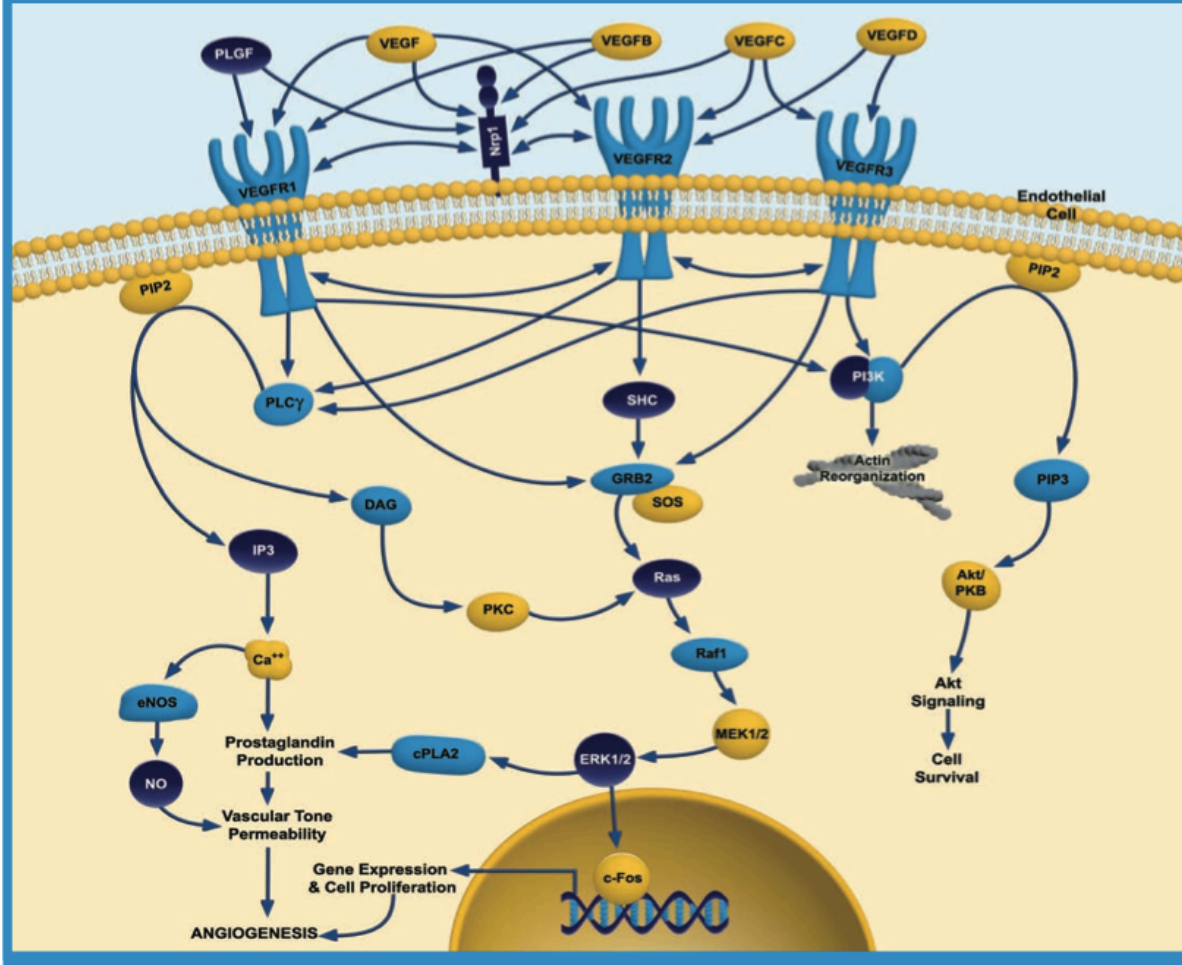
Disease Network



Mapping Human Diseases, New York Times, May 6, 2008

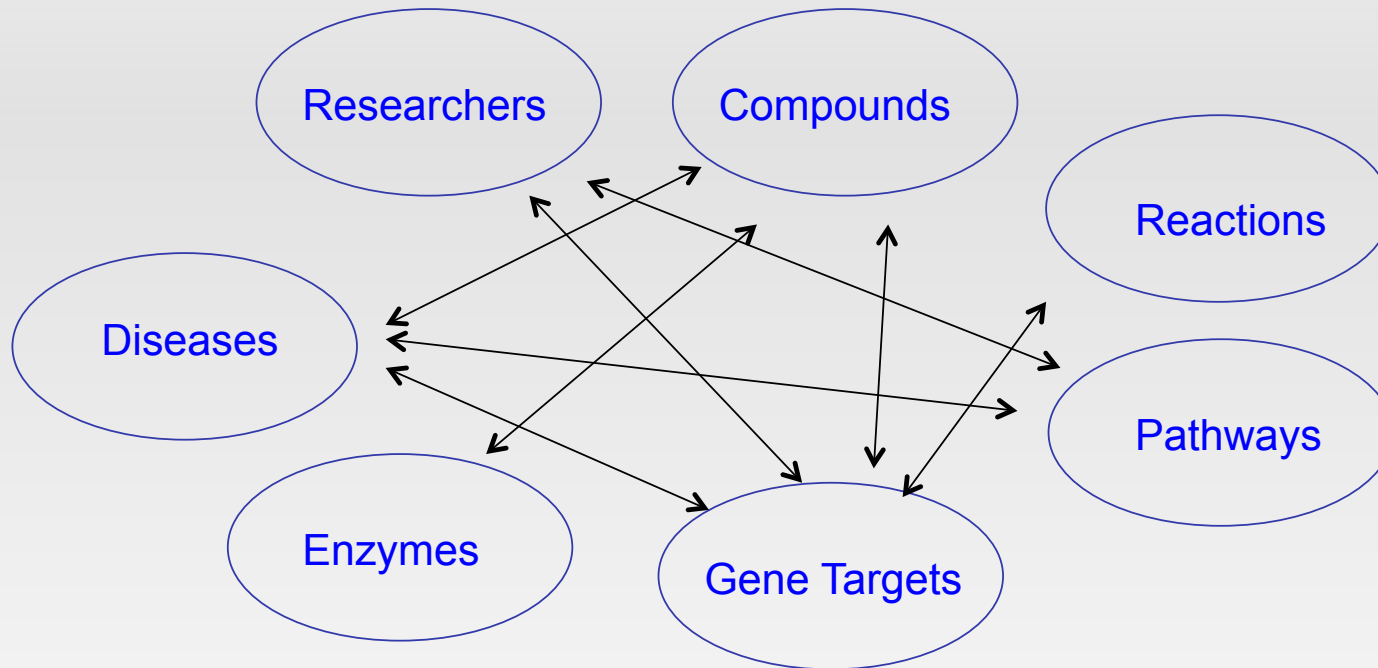
Vascular Endothelial Growth Factor (VEGF) Pathway

VEGF Family ligands and Receptor interactions



<http://www.abcam.com/>

Data Integration and Complex Queries



Demo

Live Demo – Beta version

RediscoveryDB Home Page



COMPOUNDS DISEASES ENZYMES PATHWAYS REACTIONS RESEARCHERS TARGETS **HELP**

Welcome to RedB

Rediscovery Database (RedB or 'red bee') is a project aimed at allowing researchers to explore the relationships between genes, diseases, pathways, compounds, and the researchers who are interested in them.

Many plant diseases and health traits are connected through shared molecular functions in which multiple genes participate. Hence, a global consideration of cell biology is needed to understand the mechanism of diseases and to conduct an effective drug development. How to integrate system level data from various large biological databases and be able to ask biological meaningful questions are both a challenge and an opportunity.

RediscoveryDB is addressing the issue by compiling and interrelating a large array of information including genes and their annotations, pathways, compounds, reactions, protein-protein interactions, disease/gene associations, and publications into a coherent form that greatly facilitate cross queries among the aforementioned resources. The commonly shared pathways among large number of genes or metabolites are presented in a table that can be sorted and filtered through a clean and informative interface using a Web Browser. It can present complex relationships into a network that can be visualized via Cytoscape and queried across molecular resources through Gaggle (<http://gaggle.systemsbiology.org/docs/>).

RediscoveryDB uses a unique layered approach to separate the original data source from data tables used for queries. This makes it easy to import and standardize data, as well as updating information. RediscoveryDB's schema and database backend is optimized for responsiveness necessary for Web-based interactions.

RediscoveryDB is the perfect tool for automatically annotating large data set from proteomics and metabolomics experiments. The RediscoveryDB is currently available at <http://rediscovery.bioinformatics.ucdavis.edu>

© 2010 UC Davis Bioinformatics Core
part of the University of California, Davis Genome Center



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>

Query with disease names

ReDiscovery_{database} COMPOUNDS **DISEASES** ENZYMES PATHWAYS REACTIONS RESEARCHERS TARGETS HELP

LIST MANAGEMENT

HISTORY
History Empty

DISEASES ... click to add terms → Targets [v] →

[Advanced Options](#)

<input checked="" type="checkbox"/>	Name ↕	MeSH ID ↕
<input checked="" type="checkbox"/>	PARKINSON_DISEASE	D010300
<input checked="" type="checkbox"/>	ATAXIA_TELANGIECTASIA	D001260
<input checked="" type="checkbox"/>	ALZHEIMER_DISEASE	D000544

Genes associated with the diseases based on OMIM

LIST MANAGEMENT

Export to: [Cytoscape SIF](#) [Gaggle Network](#)
Export Spreadsheet: [CSV](#), [HTML](#)

HISTORY

[diseases\(3\)](#) -> [target\(37\)](#)
at: 18:24:59

DISEASES » related TARGETS

... click to add terms Diseases

Link to this result set: http://rediscovery_dev.bioinformatics.ucdavis.edu/target/diseases?loadReq=kjrlp1263349499&history=1
[Advanced Options](#)

<input checked="" type="checkbox"/>	HGNC ID	Symbol	Name	Diseases
<input checked="" type="checkbox"/>	HGNC:7	A2M	alpha-2-macroglobulin	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:185	AD5	Alzheimer disease 5	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:188	ADAM10	ADAM metallopeptidase domain 10	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:582	APBB2	amyloid beta (A4) precursor protein-binding, family B, member 2	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:613	APOE	apolipoprotein E	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:620	APP	amyloid beta (A4) precursor protein	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:1059	BLMH	bleomycin hydrolase	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:2707	ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:7218	MPO	myeloperoxidase	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:7876	NOS3	nitric oxide synthase 3 (endothelial cell)	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:8624	PAXIP1	PAX interacting (with transcription-activation domain) protein 1	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:9052	PLAU	plasminogen activator, urokinase	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:9508	PSEN1	presenilin 1	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:9509	PSEN2	presenilin 2 (Alzheimer disease 4)	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:13666	AAAS	achalasia, adrenocortical insufficiency, alacrimia (Allgrove, triple-A)	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:26088	NBPF1	neuroblastoma breakpoint family, member 1	ALZHEIMER_DISEASE
<input checked="" type="checkbox"/>	HGNC:336	AGTR1	angiotensin II receptor, type 1	ATAXIA_TELANGIECTASIA
<input checked="" type="checkbox"/>	HGNC:795	ATM	ataxia telangiectasia mutated	ATAXIA_TELANGIECTASIA
<input checked="" type="checkbox"/>	HGNC:7230	MRE11A	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)	ATAXIA_TELANGIECTASIA
<input checked="" type="checkbox"/>	HGNC:15979	TP63	tumor protein p63	ATAXIA_TELANGIECTASIA
<input checked="" type="checkbox"/>	HGNC:251	ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide	PARKINSON_DISEASE
<input checked="" type="checkbox"/>	HGNC:3677	FGF20	fibroblast growth factor 20	PARKINSON_DISEASE

Shared Pathways Among Diseases based on KEGG



COMPOUNDS DISEASES ENZYMES **PATHWAYS** REACTIONS RESEARCHERS TARGETS HELP

LIST MANAGEMENT

Export to: [Cytoscape SIF](#) [Gaggle Network](#)

Export Spreadsheet: [CSV](#), [HTML](#)

HISTORY

[targets\(37\)](#)-> [pathway\(36\)](#)
at: 18:25:33

TARGETS » related PATHWAYS

... click to add terms

Targets

Link to this result set: http://rediscovery_dev.bioinformatics.ucdavis.edu/pathway/targets?loadReq=oexbc1263349532&history=1
[Advanced Options](#)

<input checked="" type="checkbox"/>	Pathway ID ↕	Description ↕	KEGG ↕	Targets ↕
<input checked="" type="checkbox"/>	path:hsa00010	Glycolysis / Gluconeogenesis - Homo sapiens (human)	hsa00010	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00071	Fatty acid metabolism - Homo sapiens (human)	hsa00071	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00120	Bile acid biosynthesis - Homo sapiens (human)	hsa00120	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00190	Oxidative phosphorylation - Homo sapiens (human)	hsa00190	HGNC:7717
<input checked="" type="checkbox"/>	path:hsa00330	Arginine and proline metabolism - Homo sapiens (human)	hsa00330	HGNC:7876
<input checked="" type="checkbox"/>	path:hsa00350	Tyrosine metabolism - Homo sapiens (human)	hsa00350	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00624	1- and 2-Methylnaphthalene degradation - Homo sapiens (human)	hsa00624	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00641	3-Chloroacrylic acid degradation - Homo sapiens (human)	hsa00641	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00790	Folate biosynthesis - Homo sapiens (human)	hsa00790	HGNC:30213
<input checked="" type="checkbox"/>	path:hsa00830	Retinol metabolism - Homo sapiens (human)	hsa00830	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00980	Metabolism of xenobiotics by cytochrome P450 - Homo sapiens (human)	hsa00980	HGNC:251
<input checked="" type="checkbox"/>	path:hsa00982	Drug metabolism - cytochrome P450 - Homo sapiens (human)	hsa00982	HGNC:251
<input checked="" type="checkbox"/>	path:hsa03022	Basal transcription factors - Homo sapiens (human)	hsa03022	HGNC:11588



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>

Shared Reactions Among Diseases Based on KEGG

ReDiscovery database

COMPOUNDS DISEASES ENZYMES PATHWAYS **REACTIONS** RESEARCHERS TARGETS HELP

PATHWAYS » related REACTIONS ... click to add terms Enzymes

Link to this result set: http://rediscovery_dev.bioinformatics.ucdavis.edu/reaction/pathways?loadReq=df1bc1263349570&history=1
[Advanced Options](#)

<input checked="" type="checkbox"/>	Reaction ID	Name	Definition	Equation	Pathways
<input checked="" type="checkbox"/>	R00014	pyruvate:thiamin diphosphate acetaldehydetransferase (decarboxylating)	Pyruvate + Thiamin diphosphate \rightleftharpoons 2-(alpha-Hydroxyethyl)thiamine diphosphate + CO ₂	C00022 + C00068 \rightleftharpoons C05125 + C00011	path:hsa00010
<input checked="" type="checkbox"/>	R00200	ATP:pyruvate 2-O-phosphotransferase	ATP + Pyruvate \rightleftharpoons ADP + Phosphoenolpyruvate	C00002 + C00022 \rightleftharpoons C00008 + C00074	path:hsa00010
<input checked="" type="checkbox"/>	R00235	Acetate:CoA ligase (AMP-forming)	ATP + Acetate + CoA \rightleftharpoons AMP + Diphosphate + Acetyl-CoA	C00002 + C00033 + C00010 \rightleftharpoons C00020 + C00013 + C00024	path:hsa00010
<input checked="" type="checkbox"/>	R00431	GTP:oxaloacetate carboxy-lyase (adding GTP;phosphoenolpyruvate-forming); GTP:oxaloacetate carboxy-lyase (transphosphorylating)	GTP + Oxaloacetate \rightleftharpoons GDP + Phosphoenolpyruvate + CO ₂	C00044 + C00036 \rightleftharpoons C00035 + C00074 + C00011	path:hsa00010
<input checked="" type="checkbox"/>	R00658	2-phospho-D-glycerate hydro-lyase (phosphoenolpyruvate-forming)	2-Phospho-D-glycerate \rightleftharpoons Phosphoenolpyruvate	C000631 \rightleftharpoons C00074 + C00001	path:hsa00010
<input checked="" type="checkbox"/>	R00703	(S)-Lactate:NAD+ oxidoreductase	(S)-Lactate + NAD+ \rightleftharpoons Pyruvate + NADH + H+	C00186 + C00003 \rightleftharpoons C00022 + C00004 + C00080	path:hsa00010
<input checked="" type="checkbox"/>	R00710	Acetaldehyde:NAD+ oxidoreductase	Acetaldehyde + NAD+ + H ₂ O \rightleftharpoons Acetate + NADH + H+	C00084 + C00003 + C00001 \rightleftharpoons C00033 + C00004 + C00080	path:hsa00010
<input checked="" type="checkbox"/>	R00711	Acetaldehyde:NADP+ oxidoreductase	Acetaldehyde + NADP+ + H ₂ O \rightleftharpoons Acetate + NADPH + H+	C00084 + C00006 + C00001 \rightleftharpoons C00033 + C00005 + C00080	path:hsa00010
<input checked="" type="checkbox"/>	R00726	ITP:oxaloacetate carboxy-lyase (adding ITP; phosphoenolpyruvate-forming); ITP:oxaloacetate carboxy-lyase (transphosphorylating)	ITP + Oxaloacetate \rightleftharpoons IDP + Phosphoenolpyruvate + CO ₂	C00081 + C00036 \rightleftharpoons C00104 + C00074 + C00011	path:hsa00010
<input checked="" type="checkbox"/>	R00746	Ethanol:NADP+ oxidoreductase	Ethanol + NADP+ \rightleftharpoons Acetaldehyde + NADPH + H+	C00469 + C00006 \rightleftharpoons C00084 + C00005 + C00080	path:hsa00010
<input checked="" type="checkbox"/>	R00754	Ethanol:NAD+ oxidoreductase	Ethanol + NAD+ \rightleftharpoons Acetaldehyde + NADH + H+	C00469 + C00003 \rightleftharpoons C00084 + C00004 + C00080	path:hsa00010

Shared Metabolites Among Diseases Based on KEGG

ReDiscoverydatabase

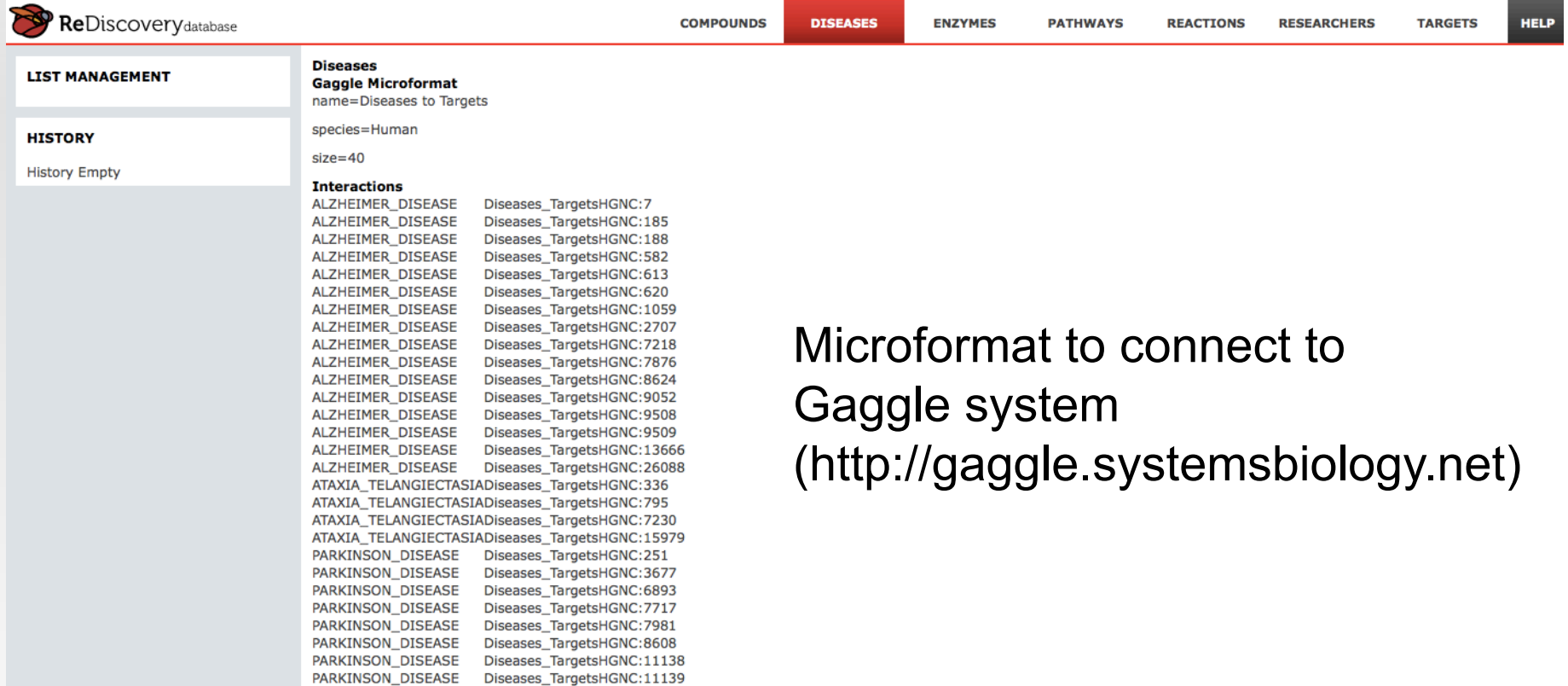
COMPOUNDS DISEASES ENZYMES PATHWAYS REACTIONS RESEARCHERS TARGETS HELP

ENZYMES » related COMPOUNDS ... click to add terms Reactions

Link to this result set: http://rediscovery_dev.bioinformatics.ucdavis.edu/compound/enzymes?loadReq=qbeoc1263398205&history=1
[Advanced Options](#)

<input checked="" type="checkbox"/>	KEGG ID	Name	Formula	INCHI
<input checked="" type="checkbox"/>	C00001	H2O; Water	H2O	1/H2O/h1H2
<input checked="" type="checkbox"/>	C00003	NAD ⁺ ; NAD; Nicotinamide adenine dinucleotide; DPN; Diphosphopyridine nucleotide; Nadide	C21H28N7O14P2	1/C21H27N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37/h1-4,7-8,10-11,13-16,20-21,29-32H,5-6H2,(H5-,22,23,24,25,33,34,35,36,37)/p+1/t10-,11-,13-,14
<input checked="" type="checkbox"/>	C00004	NADH; DPNH	C21H29N7O14P2	1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-
<input checked="" type="checkbox"/>	C00005	NADPH; TPNH	C21H30N7O17P3	1/C21H30N7O17P3/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(44-46(33,34)35)14(30)11(43-21)6/h1,3-4,7-8,10-11,13-16,20-21,29-31H,2,5-6H2,(H2,23,32)(H,36,37)(H,38,39)(H2,22,24,25)(H2,33,
<input checked="" type="checkbox"/>	C00006	NADP ⁺ ; NADP; Nicotinamide adenine dinucleotide phosphate; beta-Nicotinamide adenine dinucleotide phosphate; TPN; Triphosphopyridine nucleotide	C21H29N7O17P3	1/C21H28N7O17P3/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(44-46(33,34)35)14(30)11(43-21)6/h1-4,7-8,10-11,13-16,20-21,29-31H,5-6H2,(H7-,22,23,24,25,32,33,34,35,36,37,38,39)/p+1/t10-,11-
<input checked="" type="checkbox"/>	C00038	Zinc; Zn2 ⁺ ; Zn(II)	Zn	1/Zn
<input checked="" type="checkbox"/>	C00069	Alcohol	HOR	
<input checked="" type="checkbox"/>	C00071	Aldehyde; RCHO	CHOR	
<input checked="" type="checkbox"/>	C00080	H ⁺	H	1/p+1
<input checked="" type="checkbox"/>	C00084	Acetaldehyde; Ethanal	C2H4O	1/C2H4O/c1-2-3/h2H,1H3
<input checked="" type="checkbox"/>	C00114	Choline; Bilineurine	C5H14NO	1/C5H14NO/c1-6(2,3)4-5-7/h7H,4-5H2,1-3H3/q+1
<input checked="" type="checkbox"/>	C00116	Glycerol; Glycerin; 1,2,3-Trihydroxypropane; 1,2,3-Propanetriol	C3H8O3	1/C3H8O3/c4-1-3(6)2-5/h3-6H,1-2H2
<input checked="" type="checkbox"/>	C00226	Primary alcohol; 1-Alcohol	CH3OR	

Text results to be broadcast to Cytoscape via a simple click

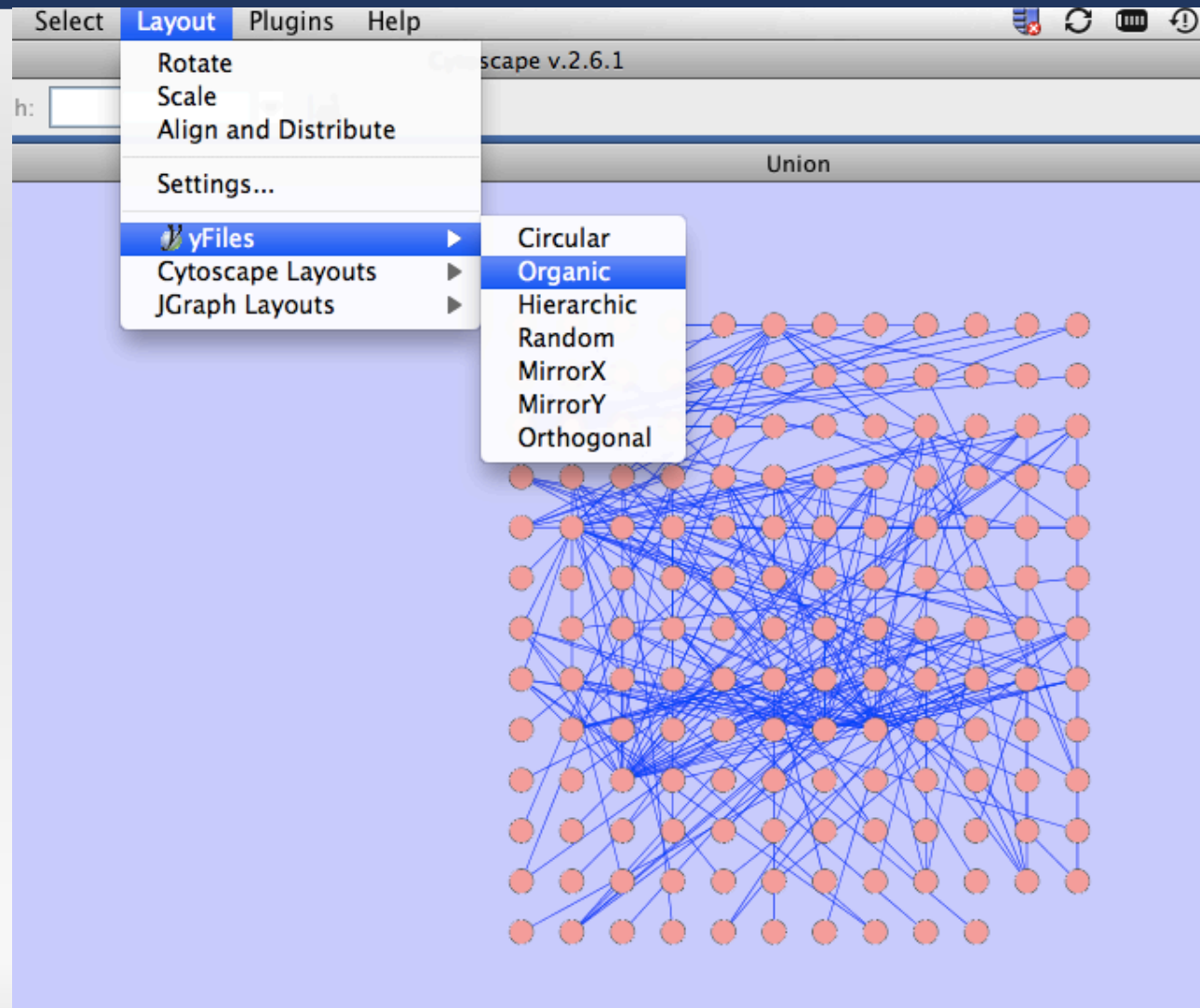


The screenshot shows the ReDiscovery database interface. The top navigation bar includes 'COMPOUNDS', 'DISEASES' (highlighted), 'ENZYMES', 'PATHWAYS', 'REACTIONS', 'RESEARCHERS', 'TARGETS', and 'HELP'. On the left, there are sections for 'LIST MANAGEMENT' and 'HISTORY' (History Empty). The main content area displays 'Diseases Gaggle Microformat' with details: 'name=Diseases to Targets', 'species=Human', and 'size=40'. Below this is a table of interactions:

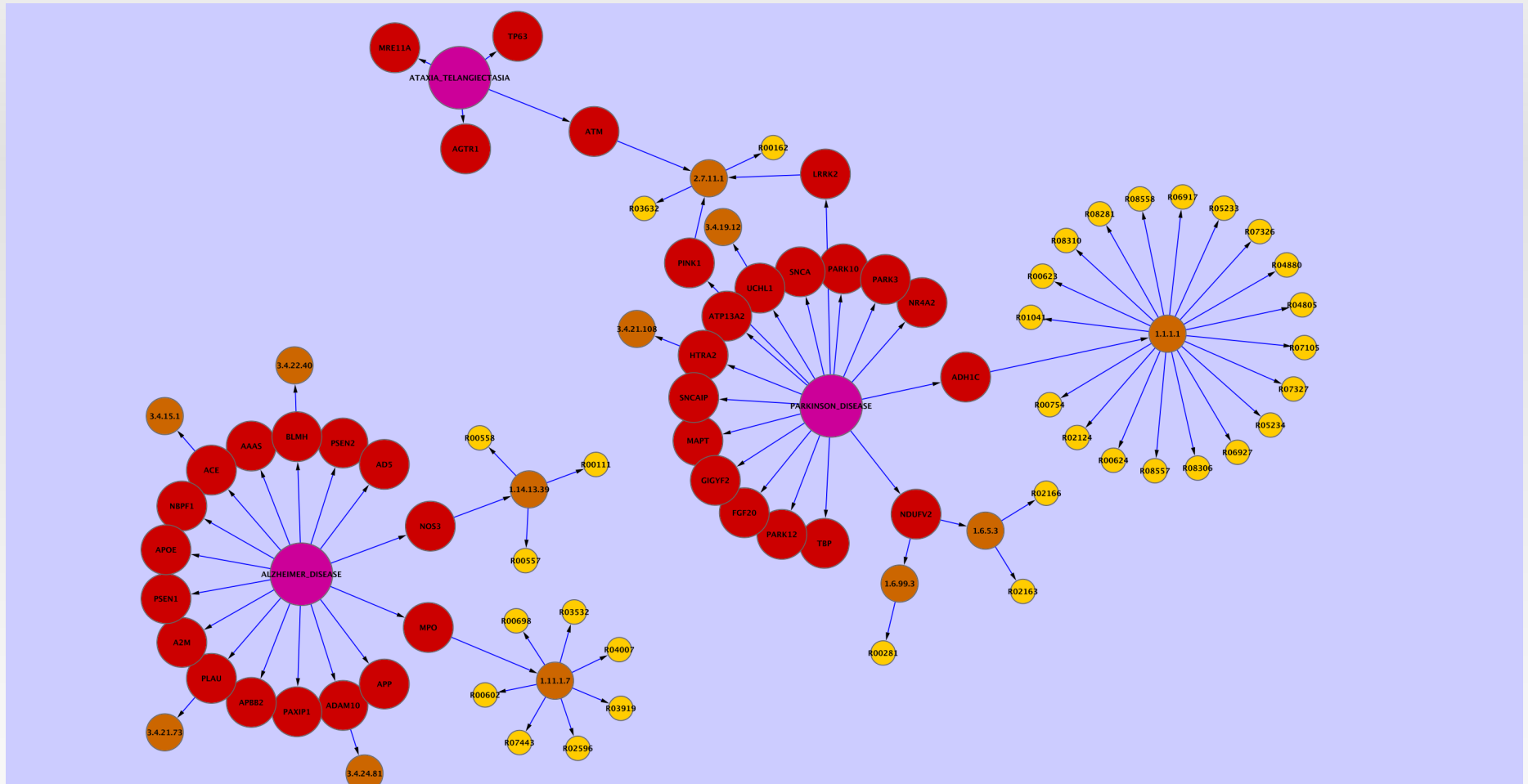
Disease	Target
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:7
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:185
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:188
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:582
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:613
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:620
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:1059
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:2707
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:7218
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:7876
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:8624
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:9052
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:9508
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:9509
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:13666
ALZHEIMER_DISEASE	Diseases_TargetsHGNC:26088
ATAXIA_TELANGIECTASIA	Diseases_TargetsHGNC:336
ATAXIA_TELANGIECTASIA	Diseases_TargetsHGNC:795
ATAXIA_TELANGIECTASIA	Diseases_TargetsHGNC:7230
ATAXIA_TELANGIECTASIA	Diseases_TargetsHGNC:15979
PARKINSON_DISEASE	Diseases_TargetsHGNC:251
PARKINSON_DISEASE	Diseases_TargetsHGNC:3677
PARKINSON_DISEASE	Diseases_TargetsHGNC:6893
PARKINSON_DISEASE	Diseases_TargetsHGNC:7717
PARKINSON_DISEASE	Diseases_TargetsHGNC:7981
PARKINSON_DISEASE	Diseases_TargetsHGNC:8608
PARKINSON_DISEASE	Diseases_TargetsHGNC:11138
PARKINSON_DISEASE	Diseases_TargetsHGNC:11139

Microformat to connect to
Gaggle system
(<http://gaggle.systemsbiology.net>)

Broadcast data directly to Cytoscape from Web queries



Network visualization reveals non-obvious connections among diseases



<http://rediscovery.bioinformatics.ucdavis.edu/>



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>

Acknowledgement

- Prof. Oliver Fiehn, The Genome Center, UC Davis
- Prof. Lars Berglund, Dr. Kate Marusina, The CTSC, UC Davis Medical School
- Prof. Aaron Cohen, DMICE, Oregon Health & Science University

- NIH, UCD
- <http://rediscovery.bioinformatics.ucdavis.edu/>