Gene Function Prediction

Exploiting Ontology Graph for Predicting Sparsely Annotated Gene Function

Sheng Wang[1,†], Hyunghoon Cho[2,†], ChengXiang Zhai[1], Bonnie Berger[2,3]and Jian Peng[1,*]

1 Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

2 Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

3 Department of Mathematics, MIT, Cambridge, MA, USA

SUPPLEMENTARY INFORMATION

1 Two cluster structures discussed in Section 3.3

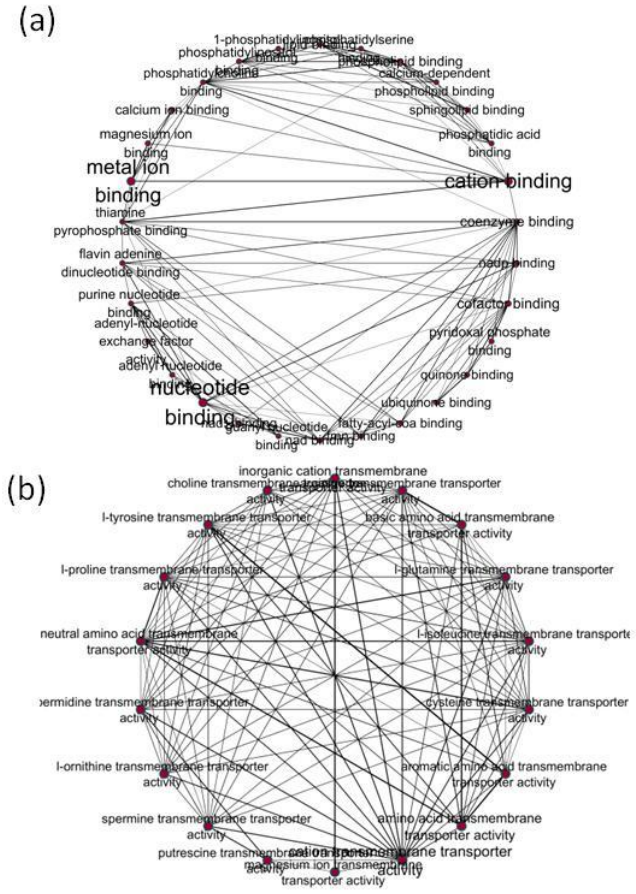
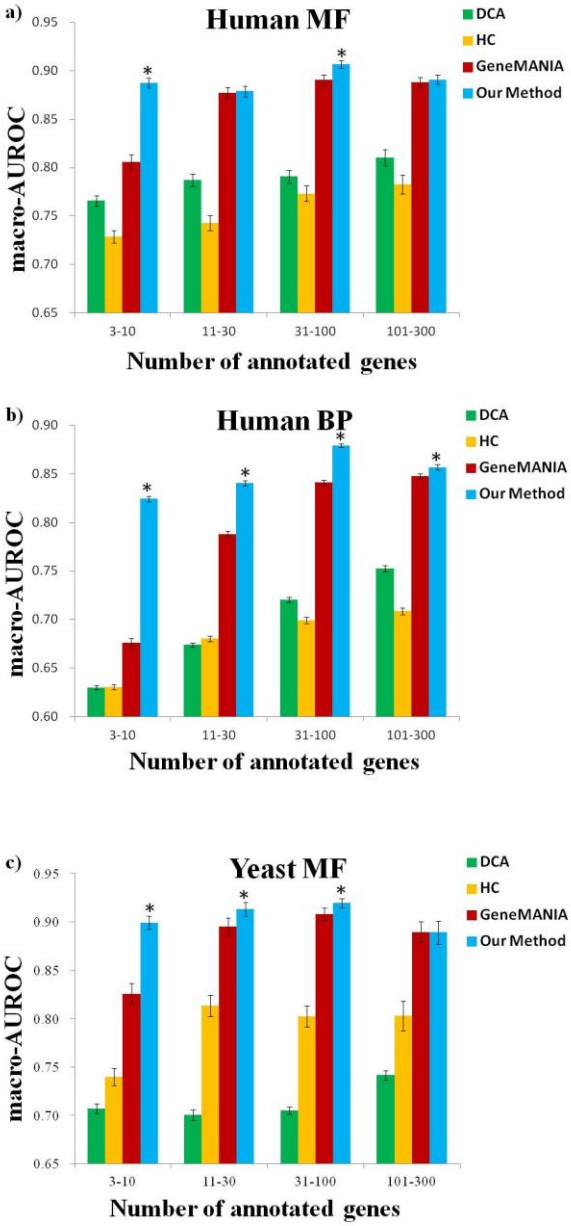


Figure S1: (a) The cluster structure of several binding functions. (b) The cluster structure of GO labels related to transmembrane transporting.

2 Comparison of performance in terms of macro-AUROC



[†]These authors equally contribute to this work.

[*]To whom correspondence should be addressed. *jianpeng@illinois.edu*

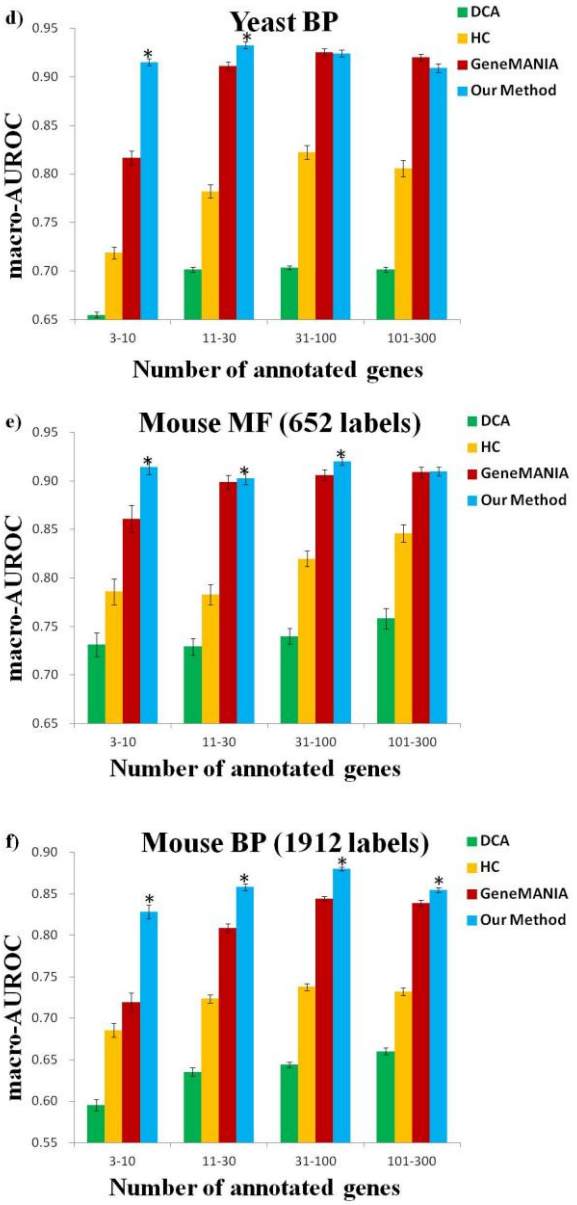


Figure S2: Comparison of our approach with other methods in terms of micro-AUROC. * indicates that our approach is statistically significant in comparison with GeneMANIA. Performance is evaluated for different subsets of GO labels with varying sparsity levels as shown on the x-axis.

3 Comparison of number of dimensions discussed in Section 3.7

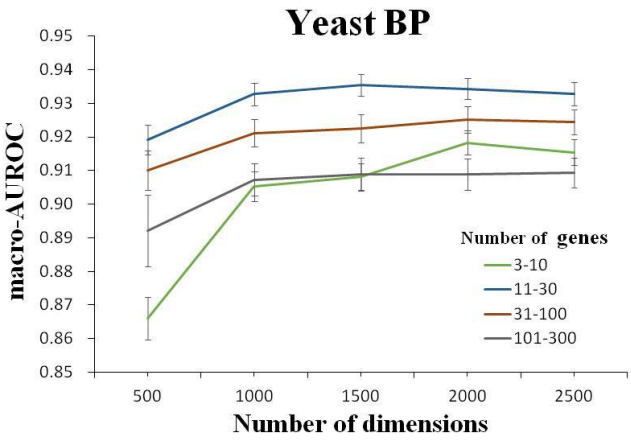
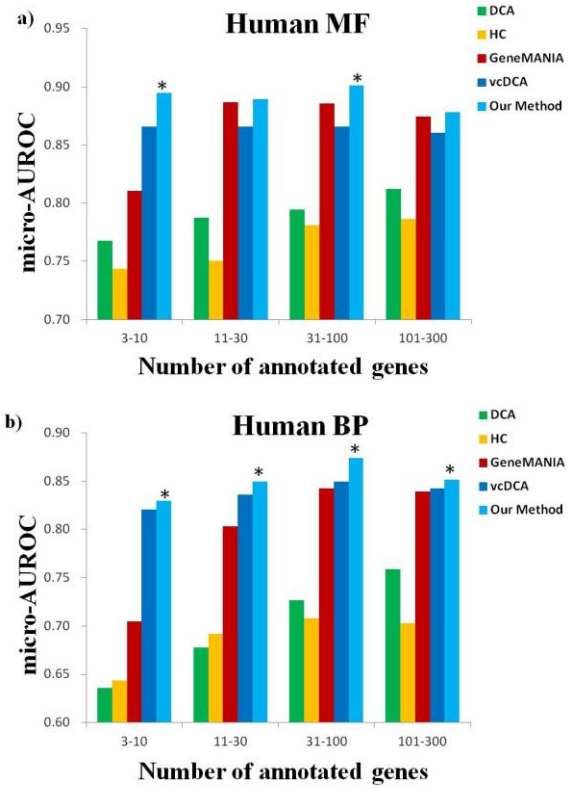


Figure S3: Comparison of number of dimensions in terms of macro-AUROC of biological process in yeast.

4 Performance of clustering based on learned label vectors

Besides clustering GO labels based on the sparseness, we explored to cluster the GO labels based on learned label vectors. We denoted this method as vcDCA. We show the performance of this method in **Figure S4** and **Figure S5.**



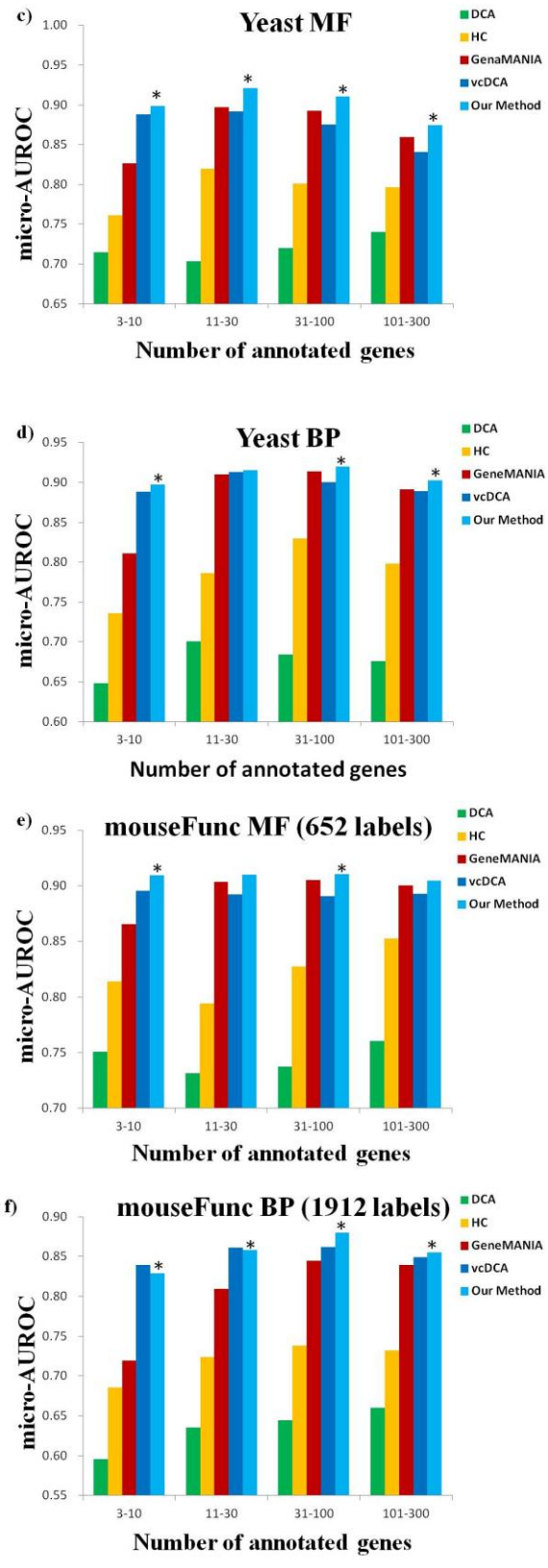
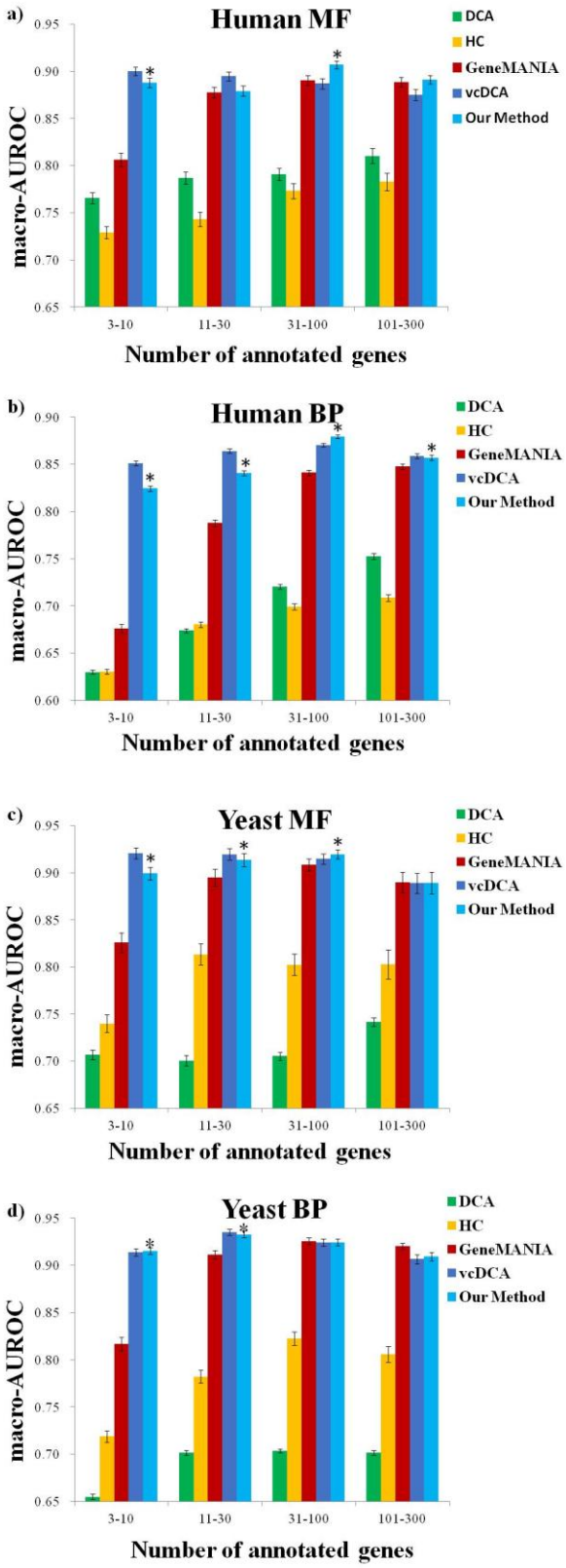


Figure S4: Comparison of our approach with other methods in terms of micro-AUROC. * indicates that our approach is statistically significant in comparison with GeneMANIA. Performance is evaluated for different subsets of GO labels with varying sparsity levels as shown on the x-axis.



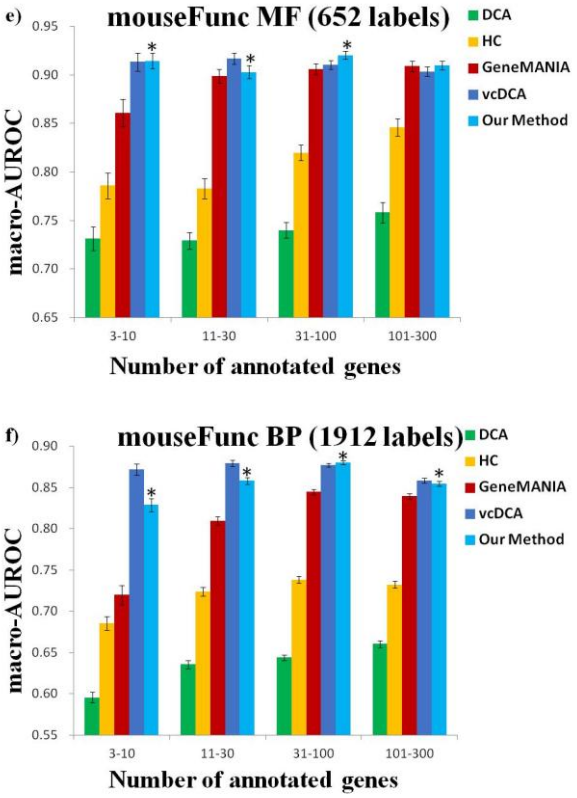


Figure S5: Comparison of our approach with other methods in terms of macro-AUROC. * indicates that our approach is statistically significant in comparison with GeneMANIA. Performance is evaluated for different subsets of GO labels with varying sparsity levels as shown on the x-axis.

5 Comparison of performance in terms of AP@10 and macro-APRUC

We show the comparison of performance in terms of AP@10 and macro-APRUC in **Table S1** and **Table S2.** In human, our method achieved 0.1317 AUPRC on BP labels with 31-100 annotations, which is slightly higher than 0.1172 AUPRC for GeneMANIA. In yeast, our method achieved 0.3353 AUPRC on MF labels with 101-300 annotations, which is higher than 0.3198 AUPRC for GeneMANIA.

Table S1: Comparison of our approach with other methods in terms of AP@10.

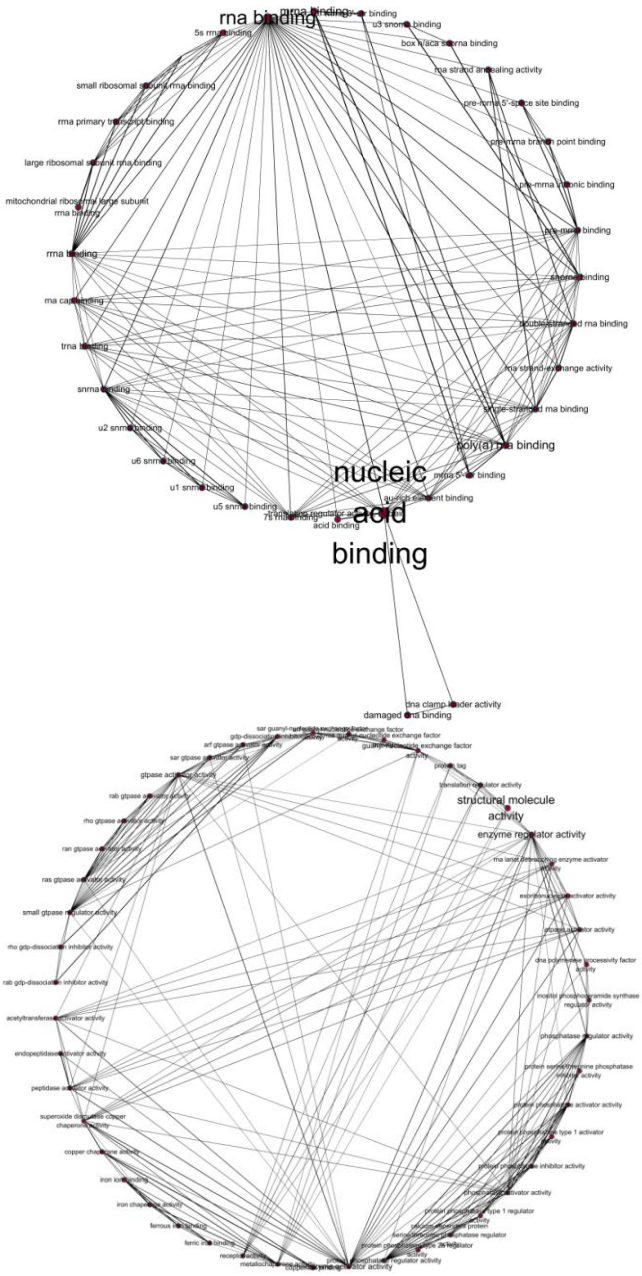
	#annotated genes	HC	GeneMANIA	clusDCA
Human MF	3-10	0.0665	0.1265	**0.1520**
	11-30	0.1517	0.2227	**0.2778**
	31-100	0.2802	0.3307	**0.3944**
	101-300	0.2710	0.4642	**0.5357**
Human BP	3-10	0.0253	0.0541	**0.0674**
	11-30	0.0658	0.1169	**0.1338**
	31-100	0.1229	0.2457	**0.3034**
	101-300	0.1829	0.3860	**0.4617**
Yeast MF	3-10	0.0860	0.1187	**0.2748**
	11-30	0.2232	0.2521	**0.3832**
	31-100	0.3285	0.3911	**0.5830**
	101-300	0.3177	0.4451	**0.6985**
Yeast BP	3-10	0.0861	0.0973	**0.2253**
	11-30	0.2003	0.2312	**0.3666**
	31-100	0.3594	0.3837	**0.5476**
	101-300	0.4184	0.4953	**0.6960**
mouseFunc MF	3-10	0.0409	0.0993	**0.2331**
	11-30	0.1782	0.2130	**0.3283**
	31-100	0.3061	0.2637	**0.4727**
	101-300	0.3242	0.3629	**0.6012**
mouseFunc BP	3-10	0.0203	0.0406	**0.0984**
	11-30	0.0927	0.0958	**0.1802**
	31-100	0.1272	0.1678	**0.3105**
	101-300	0.1648	0.2504	**0.3794**

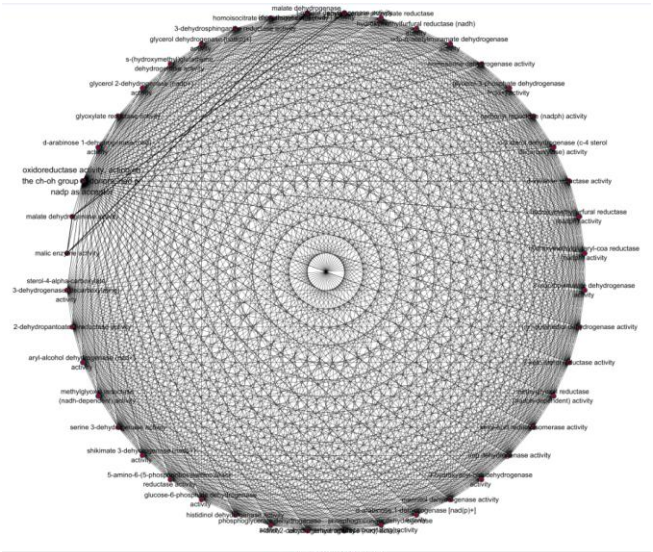
Table S2: Comparison of our approach with other methods in terms of macro-AUPRC.

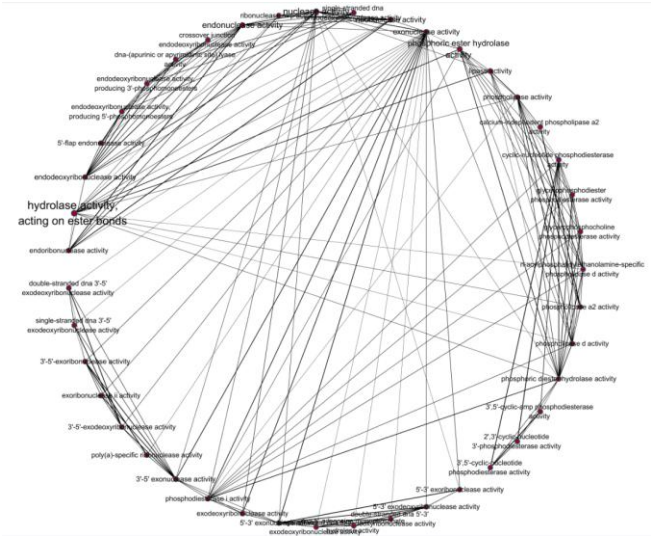
	#annotated genes	HC	GeneMANIA	clusDCA
Human MF	3-10	0.0286	0.0882	**0.0988**
	11-30	0.0527	0.1430	**0.1671**
	31-100	0.0602	0.1678	**0.1690**
	101-300	0.0603	**0.2367**	0.2133
Human BP	3-10	0.0118	0.0368	**0.0412**
	11-30	0.0285	0.0662	**0.0698**
	31-100	0.0366	0.1172	**0.1317**
	101-300	0.0436	**0.1764**	0.1721
Yeast MF	3-10	0.0403	0.0946	**0.1850**
	11-30	0.0740	0.1768	**0.2366**
	31-100	0.0730	0.2466	**0.2721**
	101-300	0.0778	0.3198	**0.3354**
Yeast BP	3-10	0.0412	0.0760	**0.1405**
	11-30	0.0867	0.1634	**0.2208**
	31-100	0.1075	0.2610	**0.2791**
	101-300	0.1158	**0.4010**	0.3805
mouseFunc MF	3-10	0.0150	0.0804	**0.1535**
	11-30	0.0534	0.1519	**0.1950**
	31-100	0.0567	0.1664	**0.2129**
	101-300	0.0596	**0.2738**	0.2663
mouseFunc BP	3-10	0.0085	0.0313	**0.0615**
	11-30	0.0323	0.0564	**0.0865**
	31-100	0.0345	0.0860	**0.1120**
	101-300	0.0437	0.1270	**0.1328**

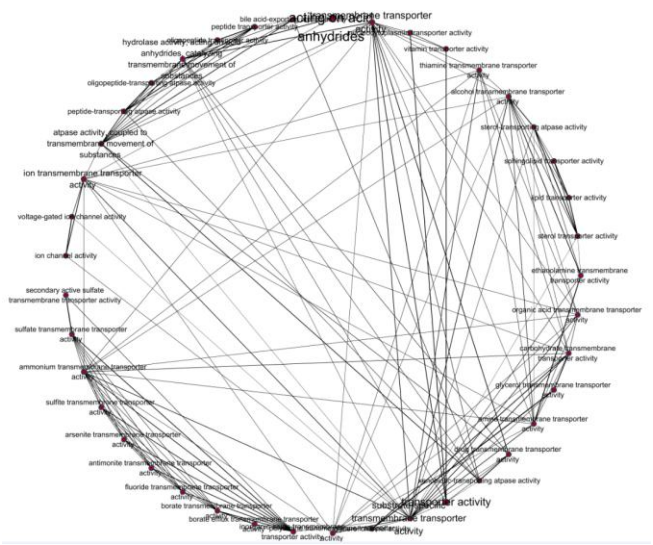
6 The complete list of clusters

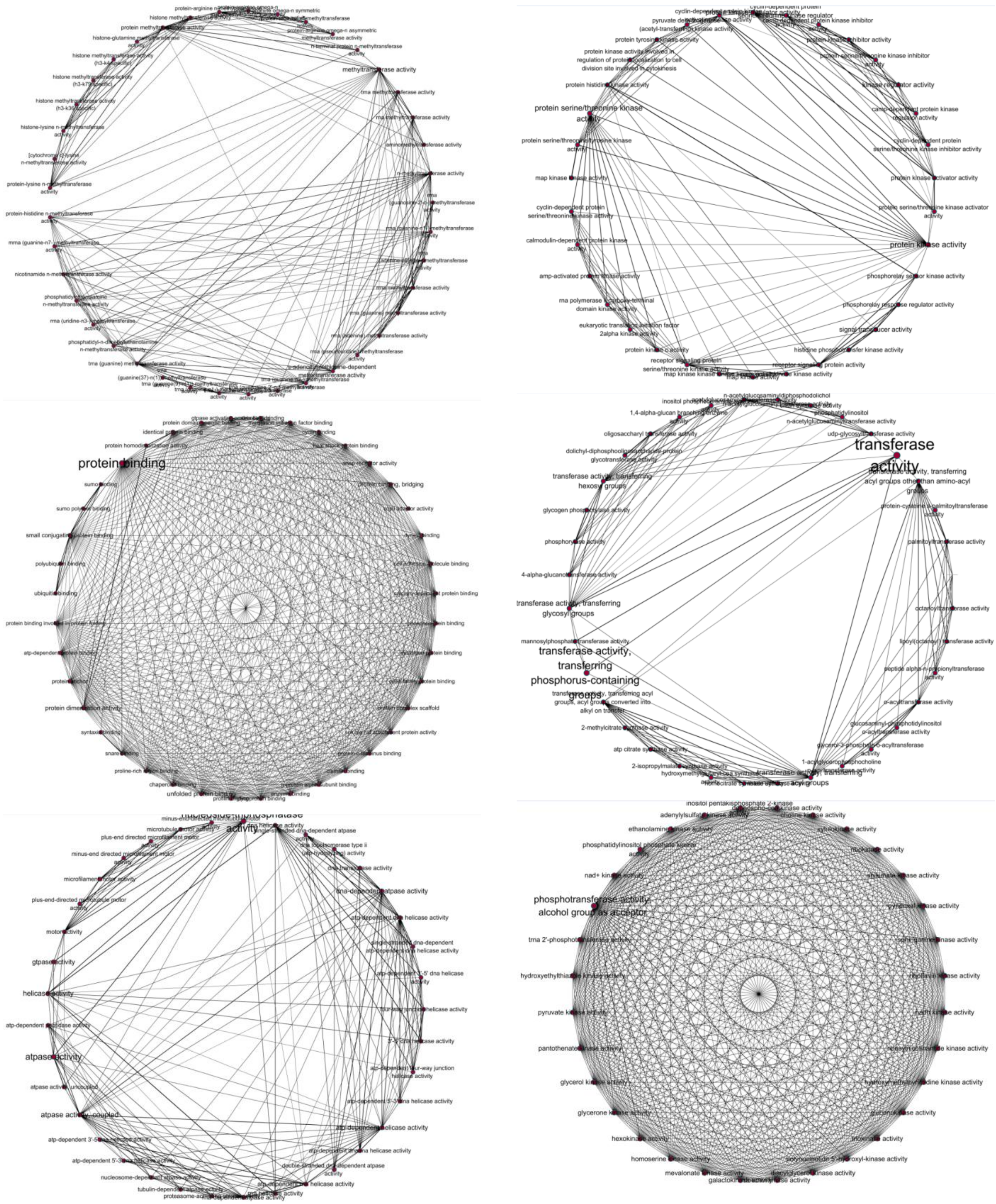
We show the complete list of clusters here.

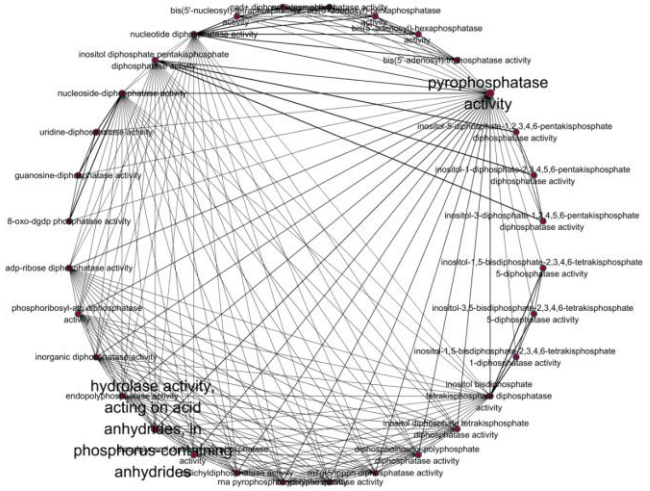


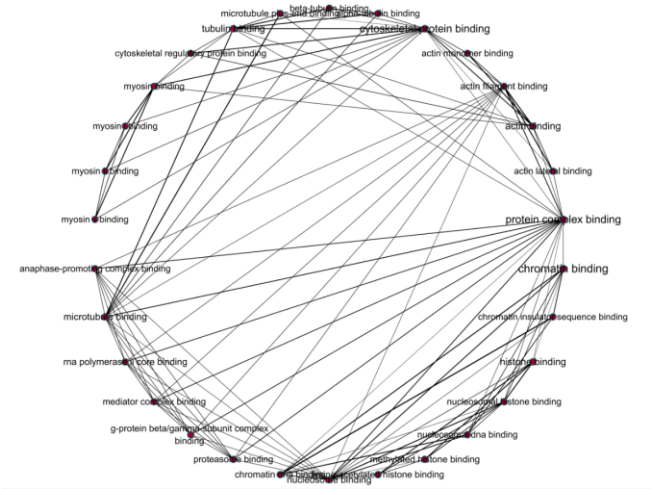


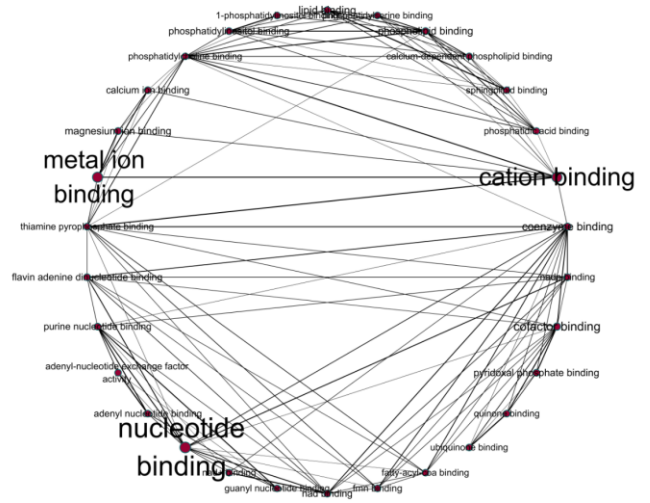


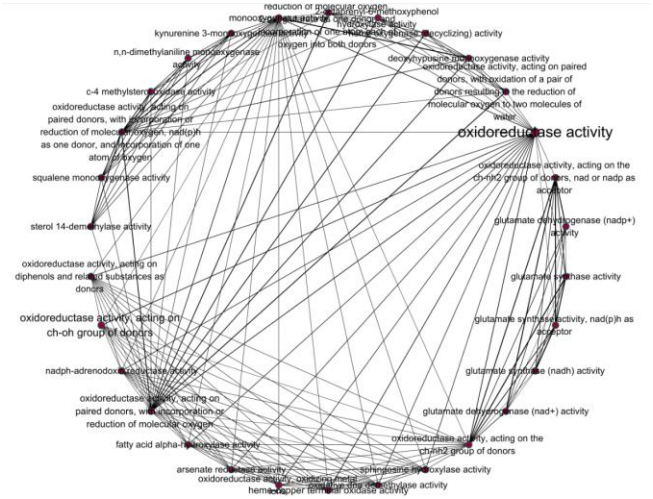


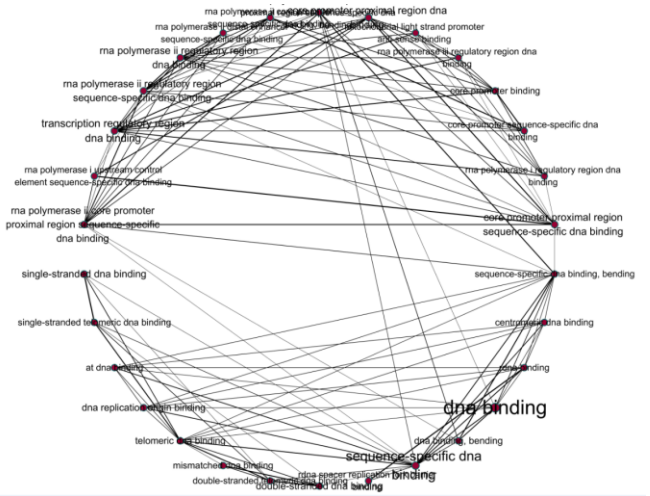

methyltransferase activity
protein serine/threonine kinase activity
protein kinase activity
protein binding
transferase activity
transferase activity, transferring glycosyl groups
transferase activity, transferring phosphorus-containing groups
transferase activity, transferring acyl groups
helicase activity
atpase activity
atpase activity, coupled
phosphotransferase activity, alcohol group as acceptor

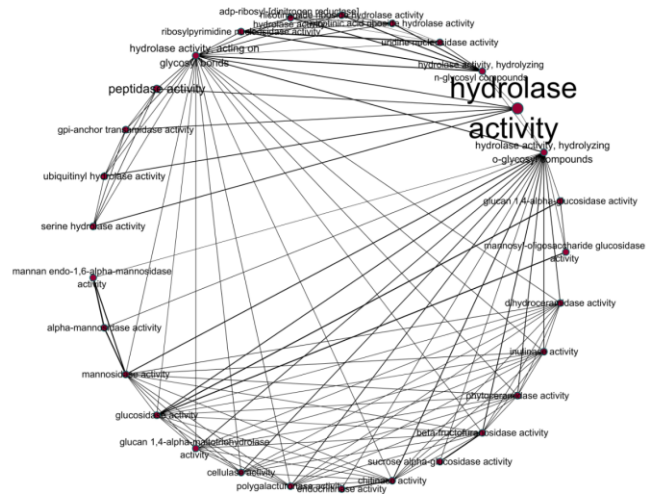

pyrophosphatase activity
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides

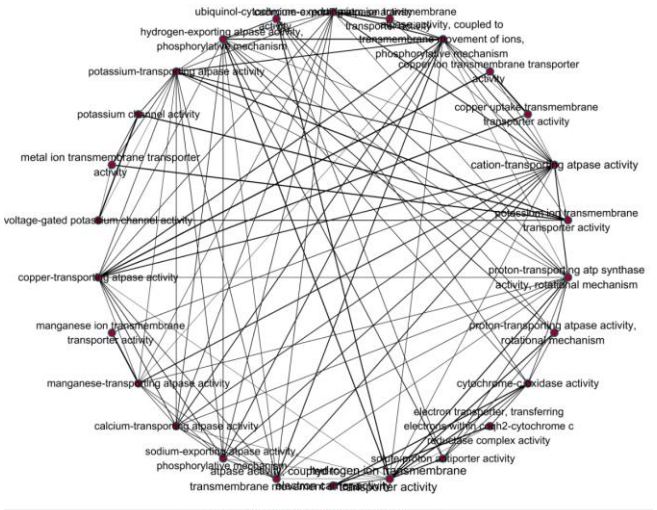

cytoskeletal protein binding
protein complex binding
chromatin binding
tubulin binding
actin filament binding
myosin binding
microtubule binding
rna polymerase ii core binding
histone binding

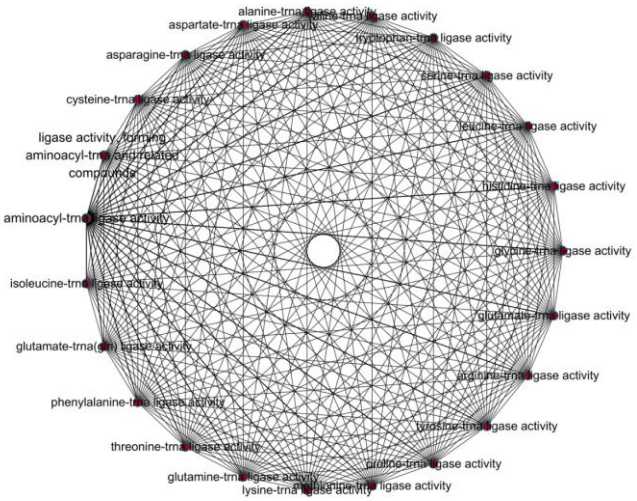

metal ion binding
cation binding
nucleotide binding
coenzyme binding
cofactor binding
lipid binding

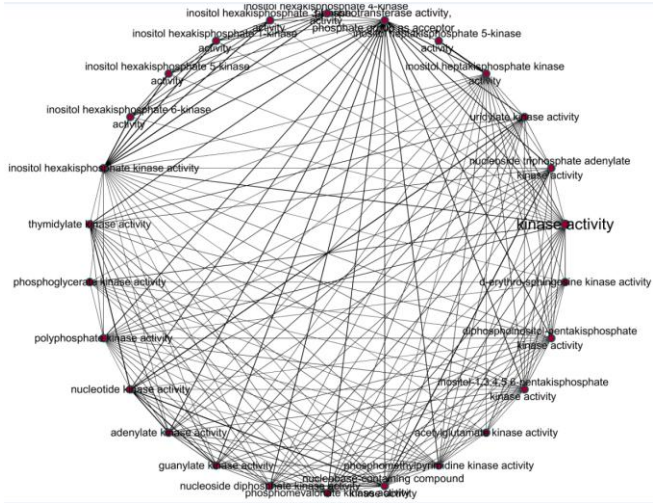

oxidoreductase activity
oxidoreductase activity, acting on ch-oh group of donors
oxidoreductase activity, acting on the ch-nh2 group of donors

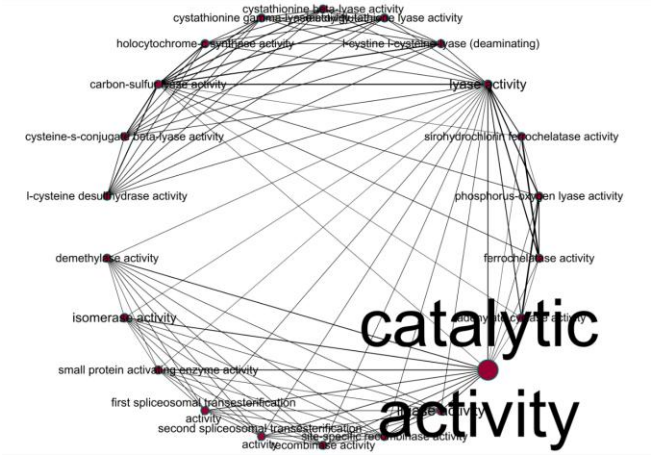

dna binding
sequence-specific dna binding
core promoter proximal region sequence-specific dna binding
rna polymerase ii core promoter proximal region dna binding

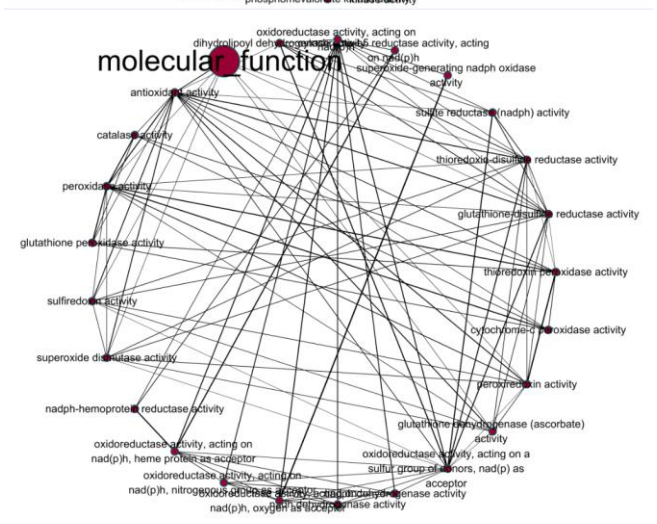

hydrolase activity
peptidase activity
hydrolase activity, acting on glycosyl bonds
hydrolase activity, hydrolyzing o-glycosyl compounds

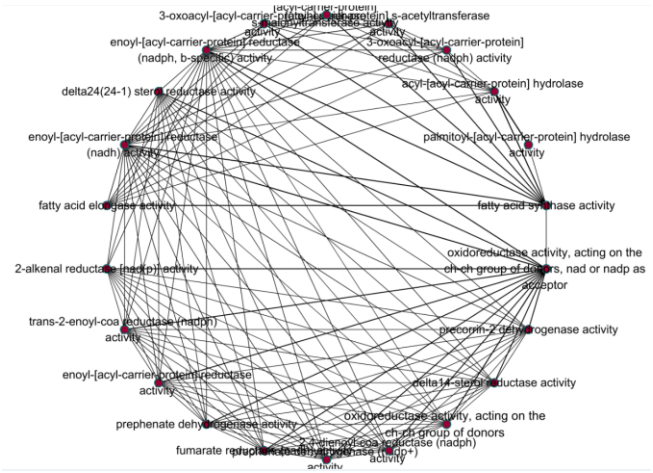

ubiquinol-cytochrome-c reductase activity
hydrogen-exporting atpase activity, phosphorylative mechanism
transporter activity, coupled to transmembrane movement of ions, phosphorylative mechanism
potassium-transporting atpase activity
copper ion transmembrane transporter activity
potassium channel activity
copper uptake transmembrane transporter activity
metal ion transmembrane transporter activity
cation-transporting atpase activity
voltage-gated potassium channel activity
potassium ion transmembrane transporter activity
copper-transporting atpase activity
proton-transporting atp synthase activity, rotational mechanism
manganese ion transmembrane transporter activity
proton-transporting atpase activity, rotational mechanism
manganese-transporting atpase activity
cytochrome-c oxidase activity
electron transporter, transferring electrons within cd1 heme protein complex activity
calcium-transporting atpase activity
sodium-exporting atpase activity
phosphorylative mechanism
sodium:proton antiporter activity
atpase activity, coupled to transmembrane movement of ions
hydrogen ion transmembrane transporter activity

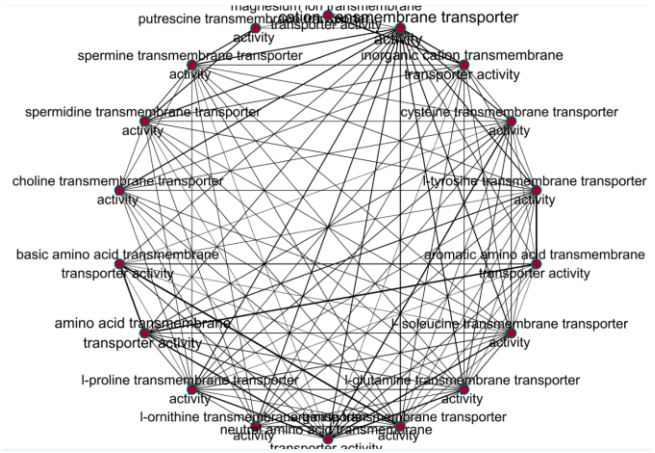

alanine-trna ligase activity
aspartate-trna ligase activity
tryptophan-trna ligase activity
asparagine-trna ligase activity
serine-trna ligase activity
cysteine-trna ligase activity
leucine-trna ligase activity
ligase activity, forming aminoacyl-trna and related compounds
histidine-trna ligase activity
aminoacyl-trna ligase activity
glycine-trna ligase activity
isoleucine-trna ligase activity
glutamate-trna ligase activity
glutamate-trna(gln) ligase activity
arginine-trna ligase activity
phenylalanine-trna ligase activity
tyrosine-trna ligase activity
threonine-trna ligase activity
proline-trna ligase activity
glutamine-trna ligase activity
lysine-trna ligase activity
alanine-trna ligase activity

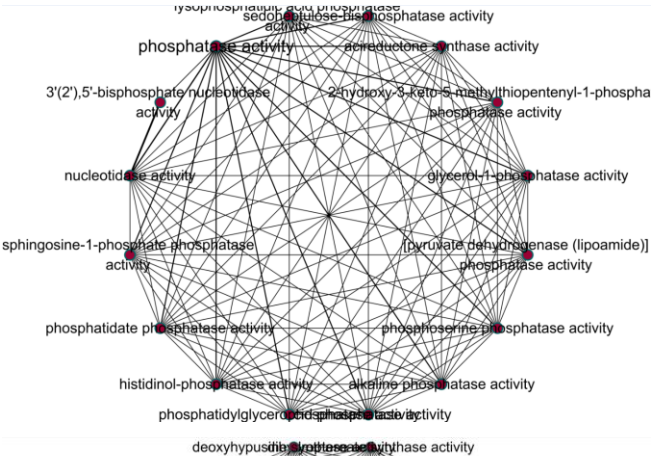

inositol hexakisphosphate 4-kinase activity
inositol hexakisphosphate 5-phosphotransferase activity
inositol hexakisphosphate 5-kinase activity
inositol hexakisphosphate 5-kinase activity
inositol heptakisphosphate kinase activity
inositol hexakisphosphate 6-kinase activity
uridylate kinase activity
nucleoside triphosphate adenylate kinase activity
inositol hexakisphosphate kinase activity
kinase activity
thymidylate kinase activity
phosphoglycerate kinase activity
d-erythro-sphinganine kinase activity
diphosphoinositol-pentakisphosphate kinase activity
polyphosphate kinase activity
inositol-1,3,4,5,6-pentakisphosphate 2-kinase activity
nucleotide kinase activity
adenylate kinase activity
acetylglutamate kinase activity
guanylate kinase activity
phosphomethylpyrimidine kinase activity
nucleoside diphosphate kinase activity
nucleobase-containing compound phosphonate kinase activity

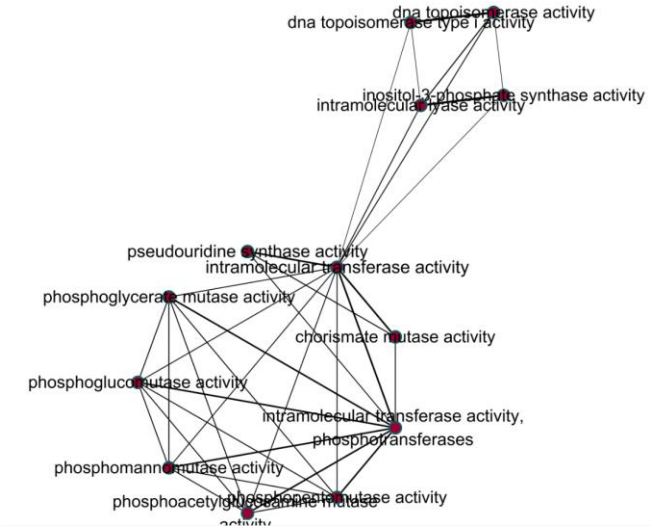

cystathionine beta-lyase activity
cystathionine gamma-lyase activity
holocytochrome-c synthase activity
cysteine lyase activity
carbon-sulfur lyase activity
lyase activity
cysteine-s-conjugate beta-lyase activity
sirohydrochlorin ferrochelatase activity
l-cysteine desulfhydrase activity
phosphorus-oxygen lyase activity
demethylase activity
ferrochelatase activity
isomerase activity
catalytic activity
small protein activating enzyme activity
first spliceosomal transesterification activity
second spliceosomal transesterification activity
recombinase activity

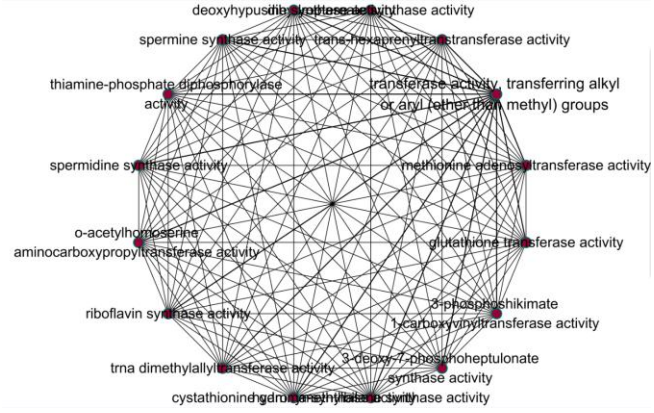

molecular_function
oxidoreductase activity, acting on dihydrolipoyl dehydrogenase (nadh) activity, acting on nad(p)h
superoxide-generating nadph oxidase activity
antioxidant activity
sulfite reductase (nadph) activity
catalase activity
thioredoxin-disulfide reductase activity
peroxidase activity
glutathione-disulfide reductase activity
glutathione peroxidase activity
thiol peroxidase activity
sulfiredoxin activity
cytochrome-c peroxidase activity
superoxide dismutase activity
peroxiredoxin activity
nadph-hemoprotein reductase activity
glutathione dehydrogenase (ascorbate) activity
oxidoreductase activity, acting on nad(p)h, heme protein as acceptor
oxidoreductase activity, acting on a sulfur group of donors, nad(p) as acceptor
oxidoreductase activity, acting on nad(p)h, nitrogenous group as acceptor
nad(p)h dehydrogenase activity
nad(p)h, oxygen as acceptor

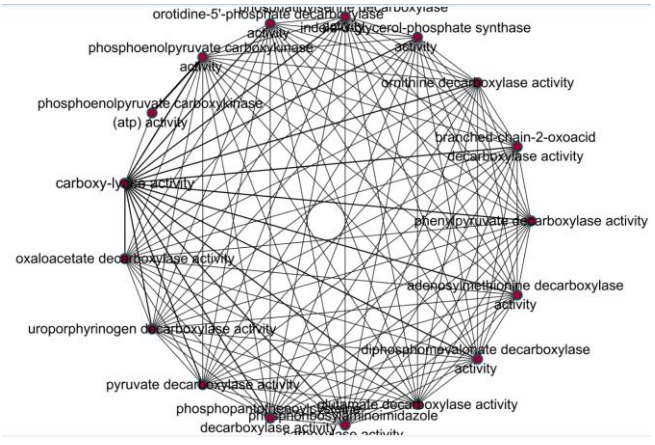

[acyl-carrier-protein] s-acetyltransferase activity
3-oxoacyl-[acyl-carrier-protein] reductase (nadph) activity
enoyl-[acyl-carrier-protein] reductase (nadph, b-specific) activity
acyl-[acyl-carrier-protein] hydrolase activity
delta24(24-1) sterol reductase activity
palmitoyl-[acyl-carrier-protein] hydrolase activity
enoyl-[acyl-carrier-protein] reductase (nadh) activity
fatty acid elongase activity
fatty acid synthase activity
oxidoreductase activity, acting on the ch-ch group of donors, nad or nadp as acceptor
2-alkenal reductase [nad(p)+] activity
trans-2-enoyl-coa reductase (nadph) activity
precorrin-2 dehydrogenase activity
enoyl-[acyl-carrier-protein] reductase activity
delta14-sterol reductase activity
prephenate dehydrogenase activity
oxidoreductase activity, acting on the ch-ch group of donors
2,4-dienoyl-coa reductase (nadph) activity

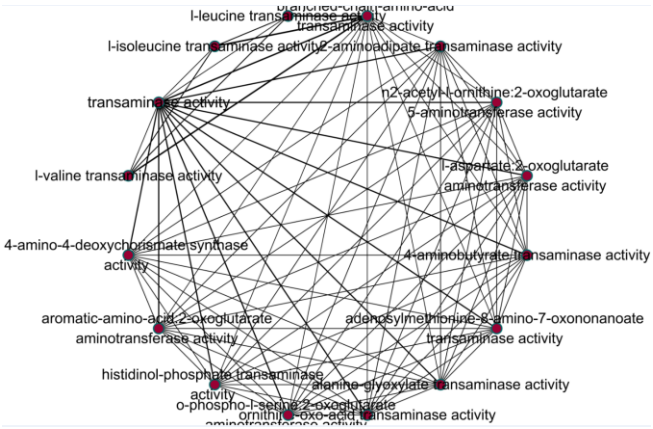

putrescine transmembrane transporter activity
spermine transmembrane transporter activity
inorganic cation transmembrane transporter activity
spermidine transmembrane transporter activity
cysteine transmembrane transporter activity
choline transmembrane transporter activity
l-tyrosine transmembrane transporter activity
basic amino acid transmembrane transporter activity
aromatic amino acid transmembrane transporter activity
amino acid transmembrane transporter activity
l-isoleucine transmembrane transporter activity
l-proline transmembrane transporter activity
l-glutamine transmembrane transporter activity

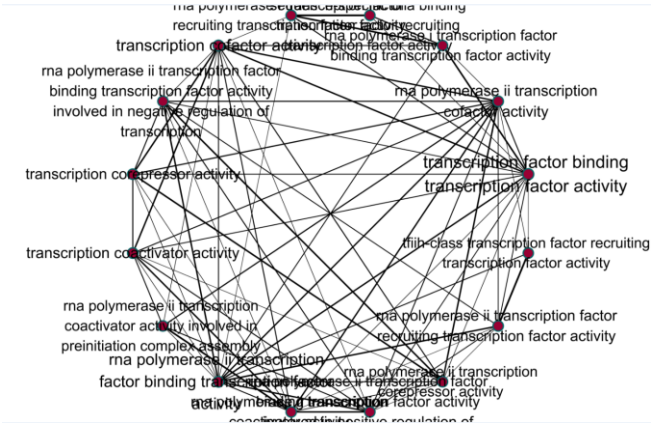

phosphatase activity
3'(2'),5'-bisphosphate nucleotidase activity
2-hydroxy-3-keto-5-methylthiopentenyl-1-phospha phosphatase activity
nucleotidase activity
glycerol-1-phosphatase activity
sphingosine-1-phosphate phosphatase activity
[pyruvate dehydrogenase (lipoamide)] phosphatase activity
phosphatidate phosphatase activity
phosphoserine phosphatase activity
histidinol-phosphatase activity
alkaline phosphatase activity

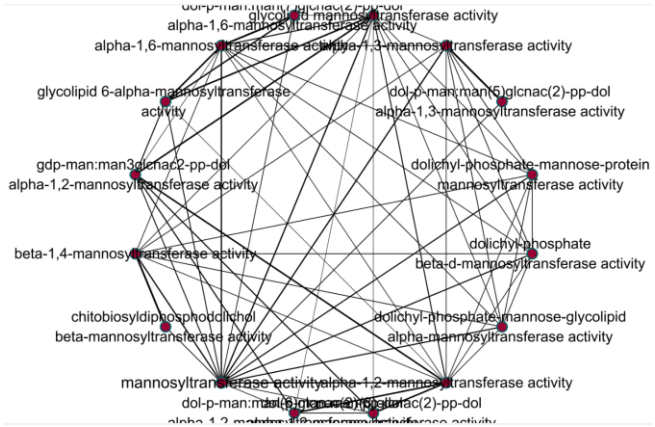

dna topoisomerase activity
dna topoisomerase type i activity
inositol-3-phosphate synthase activity
intramolecular lyase activity
pseudouridine synthase activity
intramolecular transferase activity
phosphoglycerate mutase activity
chorismate mutase activity
phosphoglucomutase activity
intramolecular transferase activity, phosphotransferases
phosphomannomutase activity
phosphoacetylglucosamine mutase activity

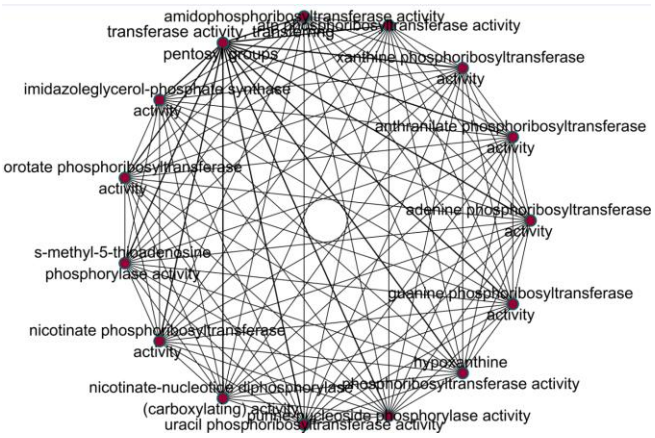

spermine synthase activity
thiamine-phosphate diphosphorylase activity
transferase activity, transferring alkyl or aryl (other than methyl) groups
spermidine synthase activity
methionine adenosyltransferase activity
o-acetylhomoserine aminocarboxypropyltransferase activity
glutathione transferase activity
riboflavin synthase activity
3-phosphoshikimate 1-carboxyvinyltransferase activity
trna dimethylallyltransferase activity
3-deoxy-7-phosphoheptulonate synthase activity

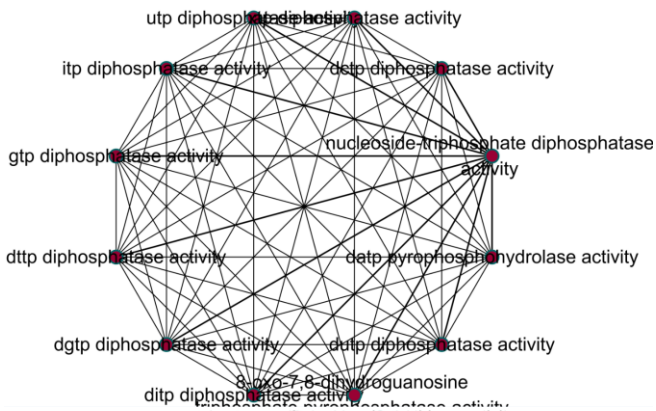

orotidine-5'-phosphate decarboxylase activity
phosphoenolpyruvate carboxykinase activity
ornithine decarboxylase activity
phosphoenolpyruvate carboxykinase (atp) activity
branched-chain-2-oxoacid decarboxylase activity
carboxy-lyase activity
phenylpyruvate decarboxylase activity
oxaloacetate decarboxylase activity
adenosylmethionine decarboxylase activity
uroporphyrinogen decarboxylase activity
diphosphomevalonate decarboxylase activity
pyruvate decarboxylase activity
glutamate decarboxylase activity

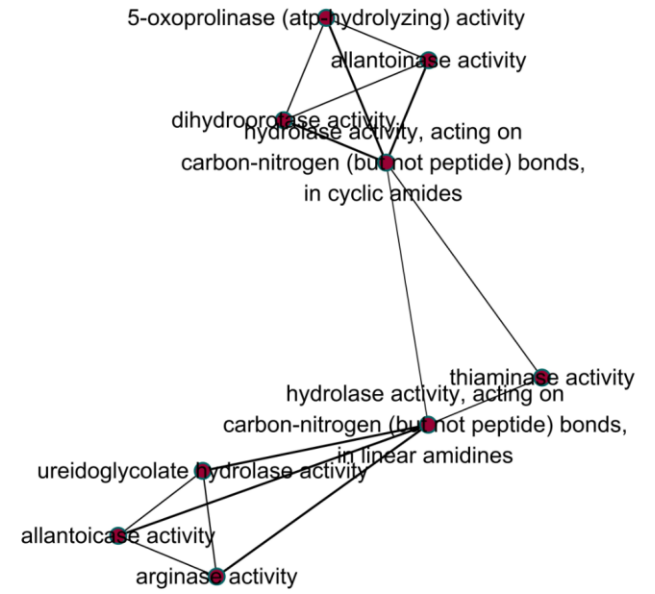

l-leucine transaminase activity
l-isoleucine transaminase activity
2-aminoadipate transaminase activity
transaminase activity
n2-acetyl-l-ornithine:2-oxoglutarate 5-aminotransferase activity
l-valine transaminase activity
l-asparate 2-oxoglutarate aminotransferase activity
4-amino-4-deoxychorismate synthase activity
4-aminobutyrate transaminase activity
aromatic-amino-acid 2-oxoglutarate aminotransferase activity
adenosylmethionine-8-amino-7-oxononanoate transaminase activity
histidinol-phosphate transaminase activity
alanine glyoxylate transaminase activity
o-phospho-l-serine 2-oxoglutarate

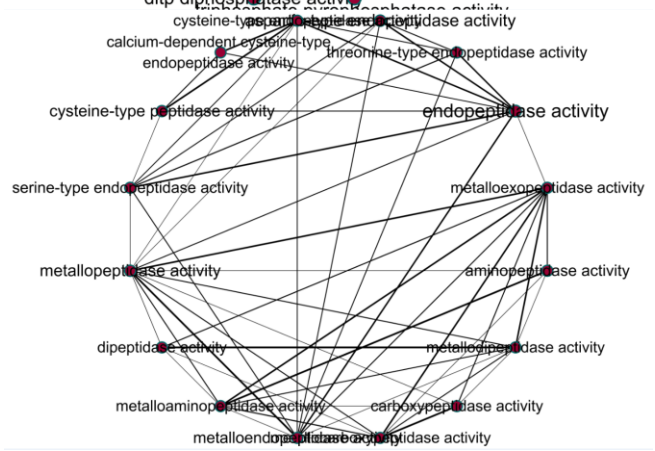

transcription cofactor activity
rna polymerase ii transcription factor binding transcription factor activity
rna polymerase ii transcription cofactor activity
transcription corepressor activity
transcription factor binding transcription factor activity
transcription coactivator activity
rna polymerase ii transcription coactivator activity involved in preinitiation complex assembly
rna polymerase ii transcription factor recruiting transcription factor activity

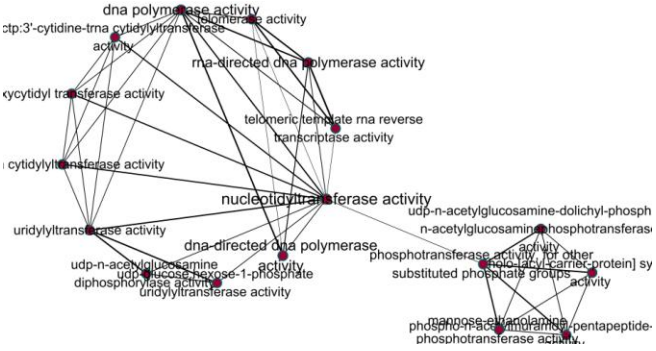

alpha-1,6-mannosyltransferase activity
alpha-1,6-mannosyltransferase activity
alpha-1,3-mannosyltransferase activity
glycolipid 6-alpha-mannosyltransferase activity
dol-p-man:man(5)glcnac(2)-pp-dol alpha-1,3-mannosyltransferase activity
gdp-man:man3glcnac2-pp-dol alpha-1,2-mannosyltransferase activity
dolichyl-phosphate-mannose-protein mannosyltransferase activity
beta-1,4-mannosyltransferase activity
dolichyl-phosphate beta-d-mannosyltransferase activity
chitobiosyldiphosphodolichol beta-mannosyltransferase activity
dolichyl-phosphate-mannose-glycolipid alpha-mannosyltransferase activity
mannosyltransferase activity
alpha-1,2-mannosyltransferase activity

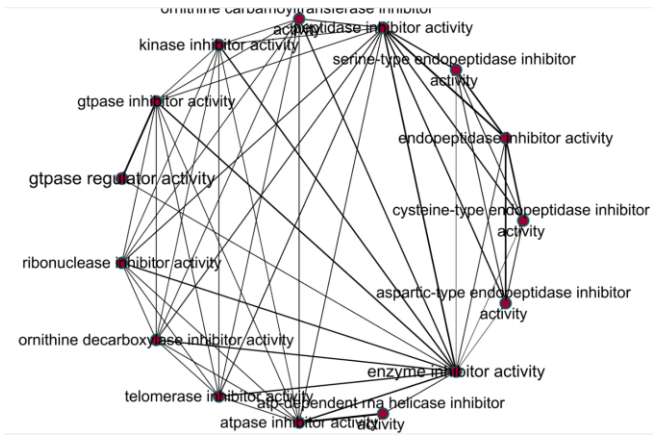

amidophosphoribosyltransferase activity
transferase activity, transferring
pentosyl groups
xanthine phosphoribosyltransferase activity
imidazoleglycerol-phosphate synthase activity
anthranilate phosphoribosyltransferase activity
orotate phosphoribosyltransferase activity
adenine phosphoribosyltransferase activity
s-methyl-5-thioadenosine phosphorylase activity
guanine phosphoribosyltransferase activity
nicotinate phosphoribosyltransferase activity
hypoxanthine phosphoribosyltransferase activity
nicotinate-nucleotide diphosphorylase (carboxylating) activity
uracil phosphoribosyltransferase activity

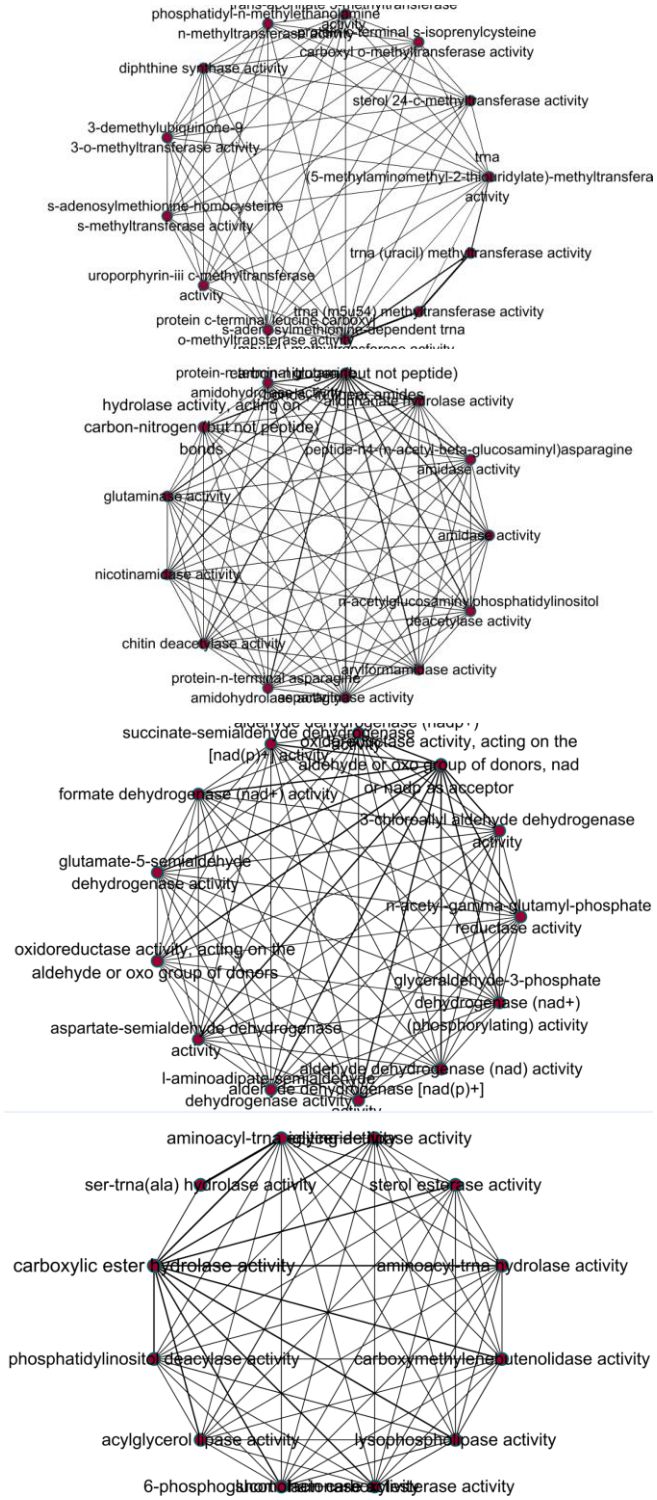

itp diphosphatase activity
dctp diphosphatase activity
gtp diphosphatase activity
nucleoside-triphosphate diphosphatase activity
dttp diphosphatase activity
datp pyrophosphohydrolase activity
dgtp diphosphatase activity
dutp diphosphatase activity
8-oxo-7,8-dihydroguanosine


5-oxoprolinase (atp-hydrolyzing) activity
allantoinase activity
dihydroorotase activity
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amides
thiaminase activity
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amidines
ureidoglycolate hydrolase activity
allantoicase activity
arginase activity


calcium-dependent cysteine-type endopeptidase activity
threonine-type endopeptidase activity
cysteine-type peptidase activity
endopeptidase activity
serine-type endopeptidase activity
metalloexopeptidase activity
metallopeptidase activity
aminopeptidase activity
dipeptidase activity
metalloendopeptidase activity
metalloaminopeptidase activity
carboxypeptidase activity


dna polymerase activity
telomerase activity
rna-directed dna polymerase activity
telomeric template rna reverse transcriptase activity
nucleotidyltransferase activity
uridylyltransferase activity
dna-directed dna polymerase activity
udp-n-acetylglucosamine diphosphorylase activity
udp-glucose hexose-1-phosphate uridylyltransferase activity
udp-n-acetylglucosamine-dolichyl-phosph
n-acetylglucosamine phosphotransferase activity
phosphotransferase activity, for other substituted phosphate groups


kinase inhibitor activity
serine-type endopeptidase inhibitor activity
gtpase inhibitor activity
endopeptidase inhibitor activity
gtpase regulator activity
cysteine-type endopeptidase inhibitor activity
ribonuclease inhibitor activity
aspartic-type endopeptidase inhibitor activity
ornithine decarboxylase inhibitor activity
enzyme inhibitor activity
telomerase inhibitor activity
atp-dependent rna helicase inhibitor activity
atpase inhibitor activity


phosphatidyl-n-methylethanolamine
n-methyltransferase activity
carboxyl o-methyltransferase activity
diphthine synthase activity
sterol 24-c-methyltransferase activity
3-demethylubiquinone-9
3-o-methyltransferase activity
trna
(5-methylaminomethyl-2-thiouridylate)-methyltransfera
activity
s-adenosylmethionine-homocysteine
s-methyltransferase activity
trna (uracil) methyltransferase activity
uroporphyrin-iii c-methyltransferase
activity
trna (m5u54) methyltransferase activity
protein c-terminal leucine carboxyl
o-methyltransferase activity
hydrolase activity, acting on
carbon-nitrogen
bonds
peptide-n4-(n-acetyl-beta-glucosaminyl)asparagine
amidase activity
glutaminase activity
amidase activity
nicotinamidase activity
n-acetylglucosaminylphosphatidylinositol
deacetylase activity
chitin deacetylase activity
protein-n-terminal asparagine
succinate-semialdehyde dehydrogenase
[nad(p)+] activity
aldehyde or oxo group of donors, nad
or nadp as acceptor
formate dehydrogenase (nad+) activity
3-chloroallyl aldehyde dehydrogenase
activity
glutamate-5-semialdehyde
dehydrogenase activity
n-acetyl-gamma-glutamyl-phosphate
reductase activity
oxidoreductase activity, acting on the
aldehyde or oxo group of donors
glyceraldehyde-3-phosphate
dehydrogenase (nad+)
(phosphorylating) activity
aspartate-semialdehyde dehydrogenase
activity
l-aminoadipate-semialdehyde
dehydrogenase activity
aldehyde dehydrogenase (nad) activity
aldehyde dehydrogenase [nad(p)+]
aminoacyl-trna
ser-trna(ala) hydrolase activity
sterol esterase activity
carboxylic ester hydrolase activity
aminoacyl-trna hydrolase activity
phosphatidylinositol deacylase activity
carboxymethylenebutenolidase activity
acylglycerol lipase activity
lysophospholipase activity
6-phosphogluconolactonase activity