

CPA: A web-based platform for Consensus Pathway Analysis and interactive visualization

Hung Nguyen
Computer Science and Engineering (CSE)
University of Nevada, Reno

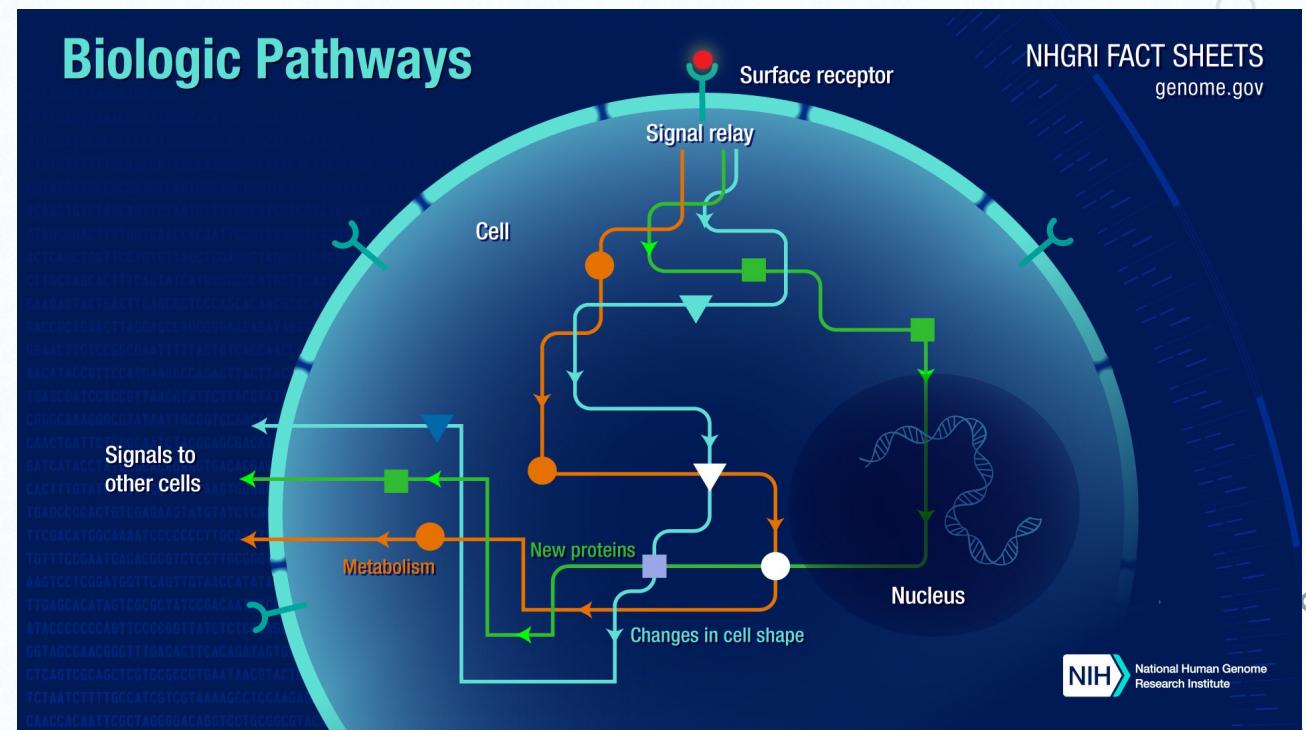
Hung Nguyen, Duc Tran, Jonathan M. Galazka, Sylvain V. Costes, Afshin Beheshti, Juli Petereit, Sorin Draghici, and Tin Nguyen. "CPA: a web-based platform for consensus pathway analysis and interactive visualization." *Nucleic Acids Research* (2021).

Background

What is a biological pathway?

A biological pathway is a series of interactions among molecules in a cell that leads to a certain product or a change in a cell. Such a pathway can trigger the assembly of new molecules, such as a fat or protein. Pathways can also turn genes on and off, or spur a cell to move.

Image source: <https://www.genome.gov>



Background

Common types of biological pathways

Metabolic pathways: a step-by-step series of interconnected biochemical reactions that convert a substrate molecule or molecules through a series of metabolic intermediates, eventually yielding a final product or products

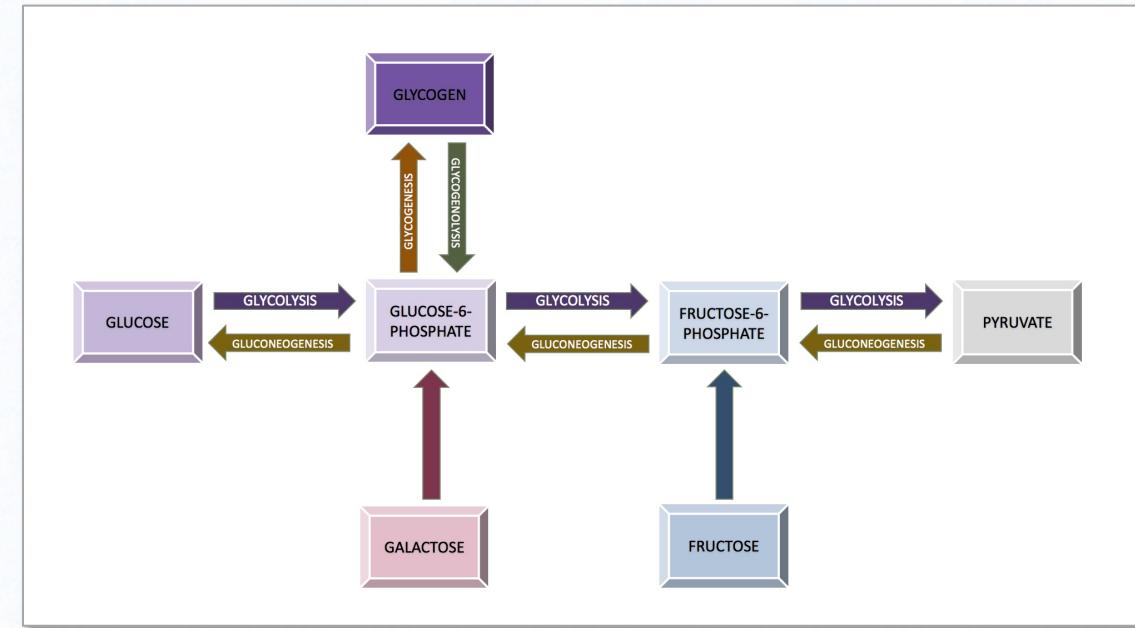


Image source: Wikipedia

Background

Common types of biological pathways

Gene-regulation pathways: a collection of molecular regulators that interact with each other and with other substances in the cell to govern the gene expression levels of mRNA and proteins which, in turn, determine the function of the cell.

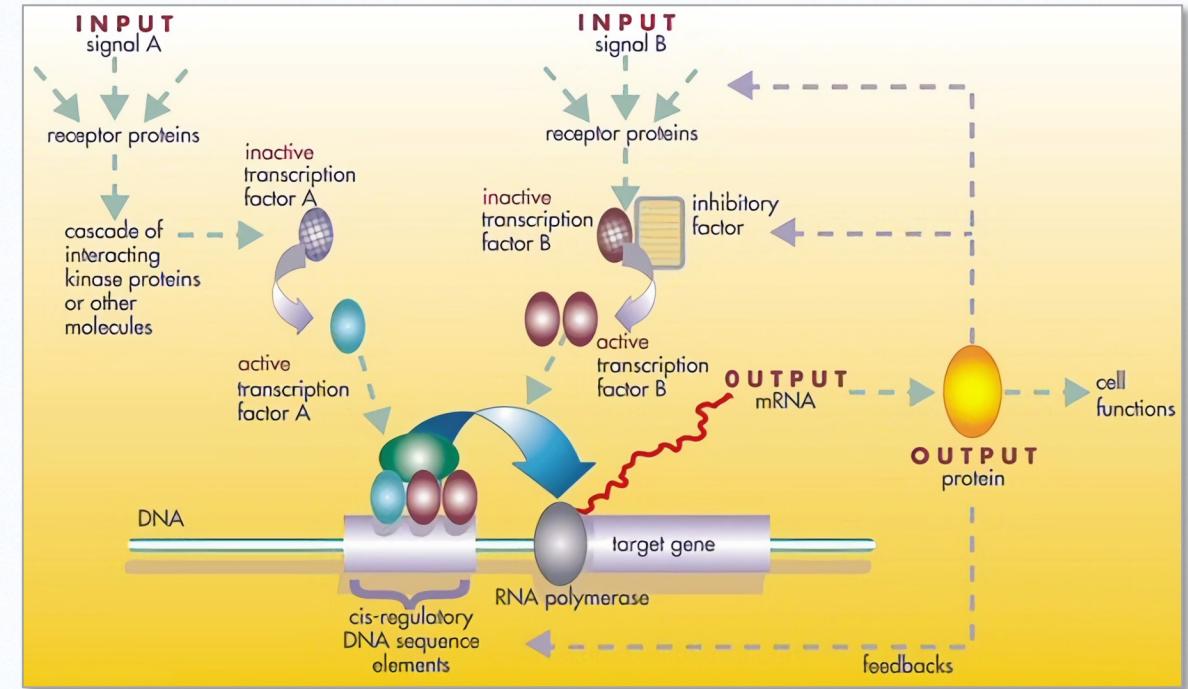


Image source: Wikipedia

Background

What is pathway analysis?

Example: One buys two of the same plant, keeps them for a year and uses different watering regimens. Plant A is watered once a week and plant B is watered twice a week. After a year, they stop watering the two plants all together and observe that plant A survive much longer than plant B.

→ What is the biological mechanism that causes plant A to have better water stress tolerance than plant B?

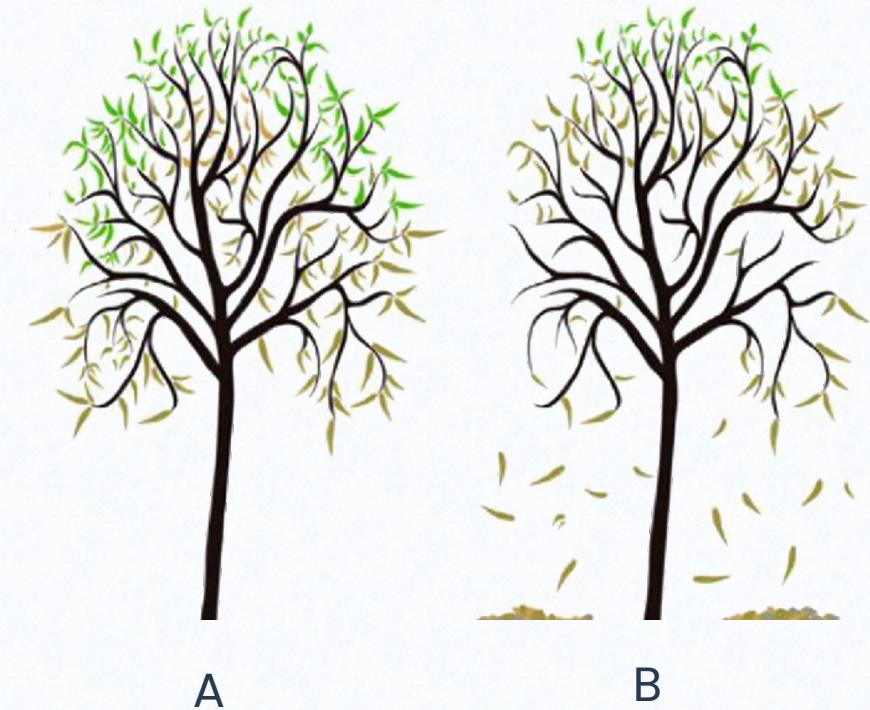


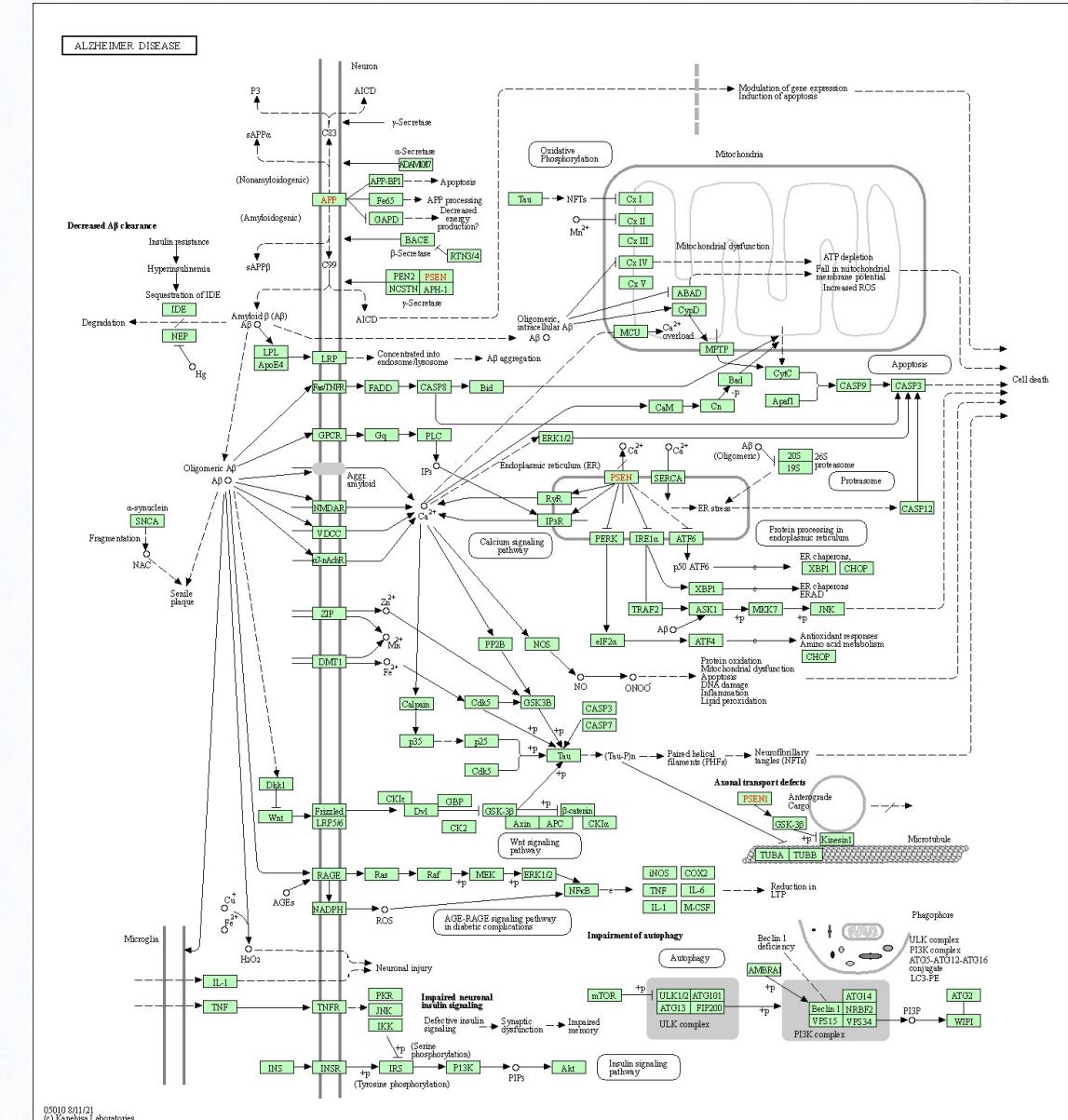
Image credit: Gely et al., Biological Reviews

Background

What is pathway analysis?

Pathway analysis is a comparative analysis that aims at identifying biological pathways that are significantly impacted in a given phenotype.

Image source: <https://www.genome.jp/pathway/hsa05010>



Background

Gene expression measurement

Ideally, measurement of expression is done by detecting the final gene product (for many genes, this is the protein); however, it is often easier to detect one of the precursors, typically mRNA and to infer gene-expression levels from these measurements.

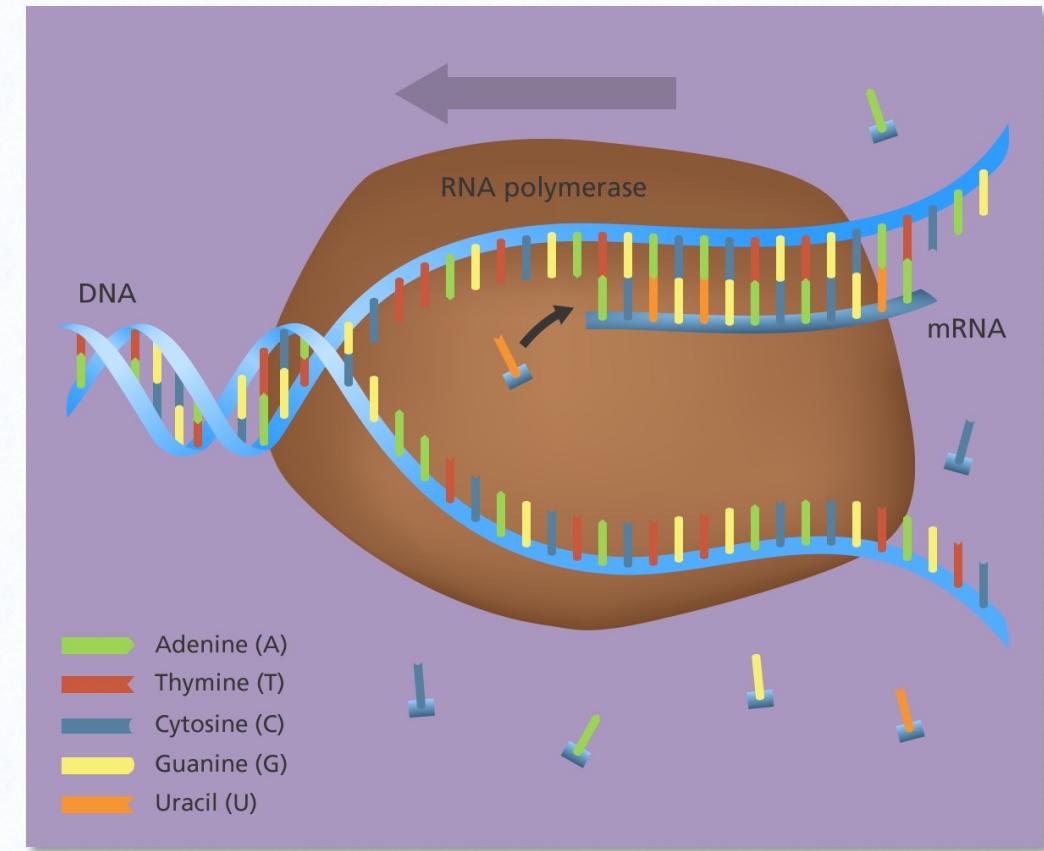
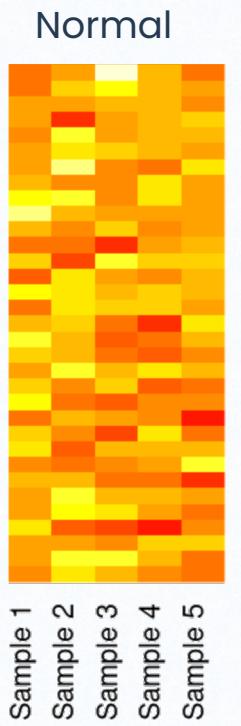


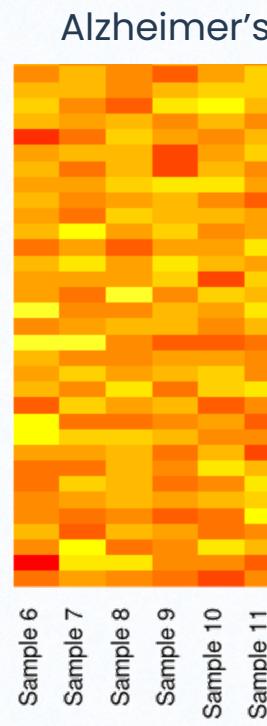
Image source: <https://www.genome.jp/pathway/hsa05010>

Background

Differential expression analysis

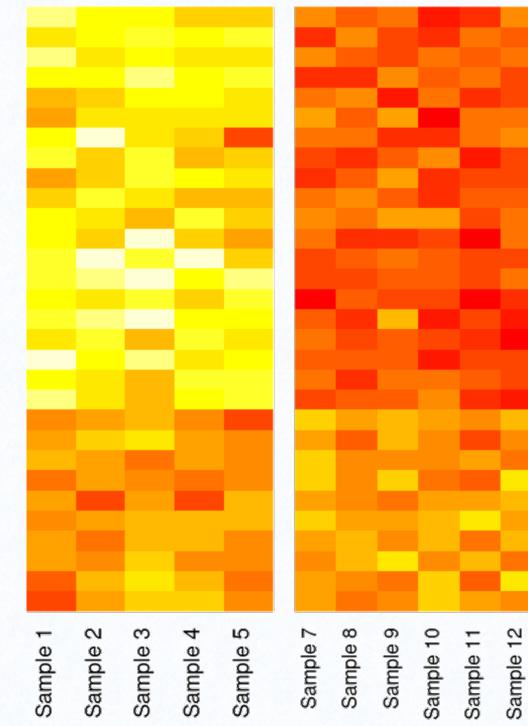


VS



LOC100420386
MTND5P23
TBX2
TNNT2
LOC642381
RNU6-884P
LINC00501
RSU1P2
LOC105373418
GAP43
TPT1P7
TMEM161B
PAF1
ZNF678
OTC
LOC100507387
OR10AH1P
RPL7P47
RNA5SP218
RN7SL363P
LOC108281164
POLR1B
LOC112694701
ERC2-IT1
LOC646708
LINC01504
LOC101060042
CHCHD2P8
LINC00638
LOC100506321
LOC101927191
IGHV1-69D
LOC101928692

Statistical test



IDDM17
LOC101927141
THSD1P1
OR4C2P
PPP1R32
GRAMD4P6
RPL12P4
PRR20P
TRG-CTG1-2
ANGPTL6
LOC110121407
LOC102723341
LOC100420778
TCAM1P
LOC105374018
LINC02775
SAG
LOC100420386
MTND5P23
TBX2
TNNT2
LOC642381
RNU6-884P
LINC00501
RSU1P2
LOC105373418
GAP43
TPT1P7
TMEM161B
PAF1

Differentially
Expressed genes

Background

Why pathway analysis?

Differential expression analysis typically yields a list of differentially expressed genes or proteins. This list:

- Is useful in identifying genes that may have roles in a given phenomenon or phenotype
- **However**, fails to provide mechanistic insights into the underlying biology of the condition being studied.

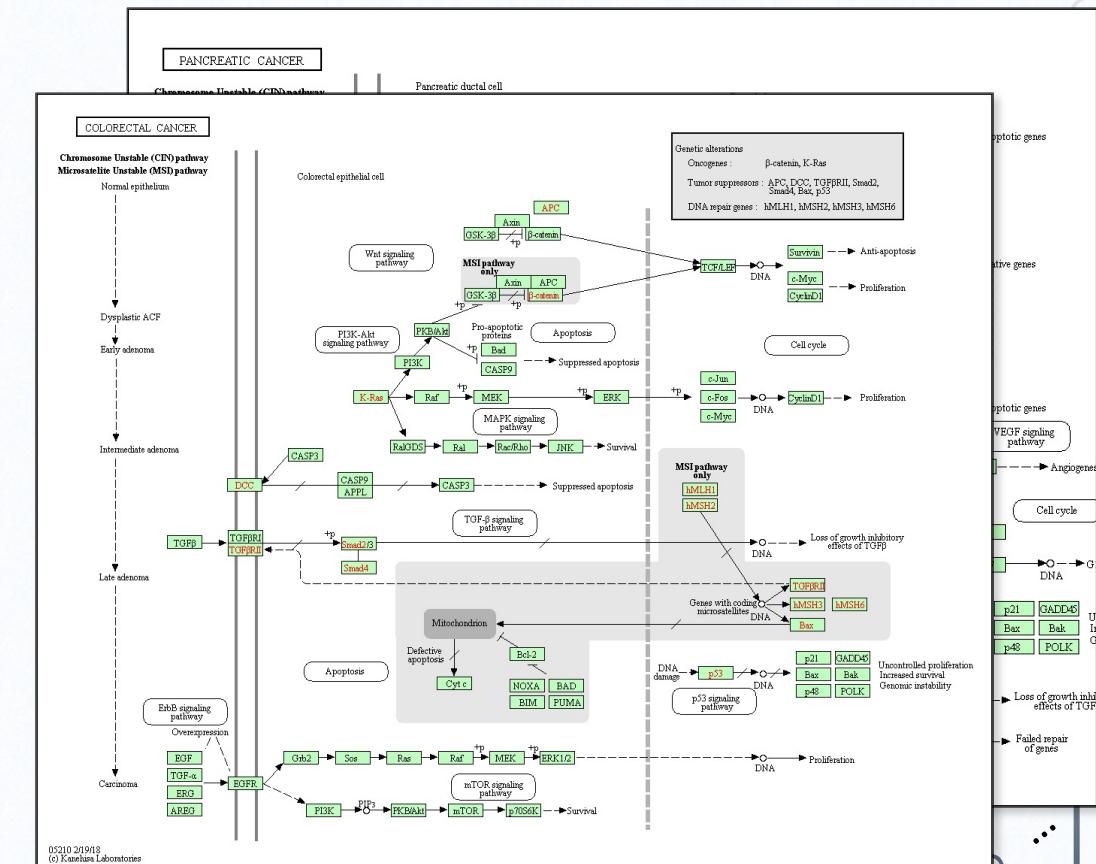
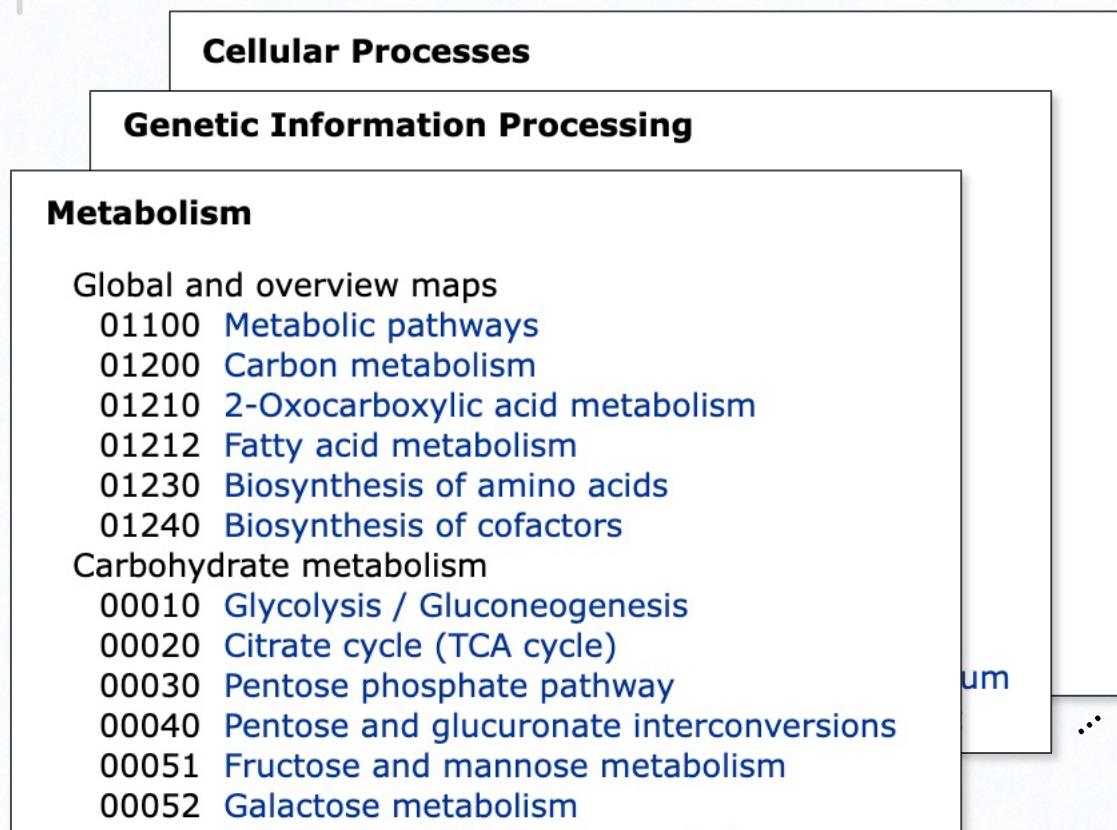
Pathway analysis:

- Groups genes and protein (and their interactions) that are known to be involved in specific biological process, components, or structures in the same pathway.
- Identifies active pathways that differ between two conditions. This can have more explanatory power than a simple list of different genes or proteins.

Pathway analysis

Input – Pathway list

KEGG pathway database: <https://www.genome.jp/keg/pathway.html>

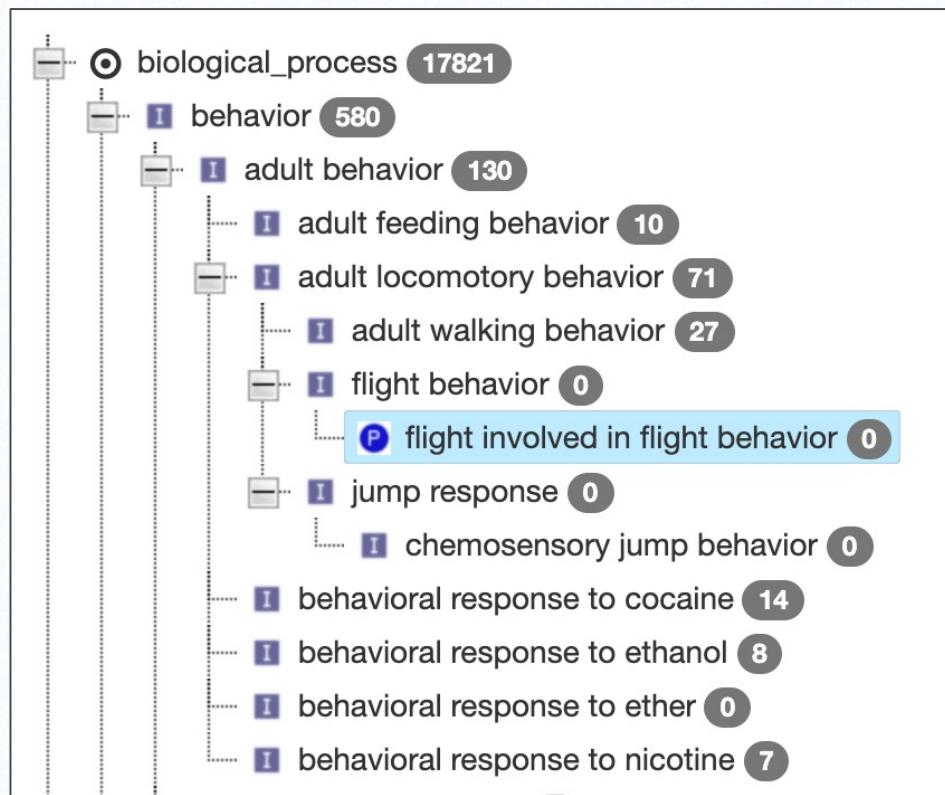


https://www.genome.jp/keg-bin/show_organism?menu_type=pathway_maps&org=hsa

Pathway analysis

Input – Pathway list (cont.)

Gene Ontology (GO) database: <http://amigo.geneontology.org/amigo/dd Browse>



Gene/product	Gene/product name	Annotation qualifier	GO class (direct)	Annotation extension	Contributor	Organism
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrial respiratory chain complex III assembly		UniProt	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrion		HPA	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrion	part of skeletal muscle tissue	UniProt	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrion		UniProt	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrial inner membrane		Reactome	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrial respiratory chain complex III		UniProt	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrial respiratory chain complex III		UniProt	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrial respiratory chain complex IV		UniProt	Homo sapiens
UQCRLS1	Cytochrome b-c1 complex subunit Rieske, mitochondrial		mitochondrial respiratory chain complex III		GO_Central	Homo sapiens
FIBP	Acidic fibroblast growth factor intracellular-binding protein		mitochondrion		PINC	Homo sapiens

Pathway analysis

Input & Output

Input:

A list of pathways, and genes measurements/statistics:

- A matrix of gene expression measurement of all samples, or
- A list of genes and their statistics/ranking, or
- A list of differential expressed genes and background genes

↓

**Pathway
analysis
method**

→

Output:

A list of significantly impacted pathways

Pathway analysis

Input – Gene measurements/statistics

DE Genes

Genes
ABCC9
CTNNB1
TSPAN7
RPN2
PRPS1
KCNAB1
FGFR3
HCFC1
SPDYA
TFEC
PDLIM5
CAPG
CUX1
ATM
...

Genes & statistics

Genes	Fold change	p-value
ABCC9	0.44	8.98E-02
CTNNB1	1.62	8.01E-01
TSPAN7	0.56	9.65E-02
RPN2	-0.38	2.62E-02
PRPS1	-0.10	3.88E-01
KCNAB1	1.89	1.25E-01
FGFR3	0.50	1.16E-02
HCFC1	1.54	3.49E-01
SPDYA	0.04	1.86E-01
TFEC	0.55	6.04E-01
PDLIM5	0.02	4.80E-09
CAPG	0.28	6.24E-10
CUX1	1.03	5.25E-01
ATM	0.77	2.02E-04
...

or

or

Full expression matrix

	S1	S2	S3	...	Sx	Sy	Sz
ABCC9	5.87	6.20	5.94	...	5.84	5.43	5.96
CTNNB1	5.66	5.61	5.57	...	5.64	5.82	5.53
TSPAN7	7.95	7.65	7.84	...	7.87	7.76	7.98
RPN2	7.17	6.67	7.25	...	7.47	6.80	7.40
PRPS1	4.38	4.58	4.41	...	4.23	4.28	4.10
KCNAB1	4.84	4.91	4.87	...	4.55	5.20	5.01
FGFR3	4.52	4.53	4.46	...	4.39	4.41	4.46
HCFC1	4.01	3.74	3.99	...	3.89	3.80	3.93
SPDYA	4.64	5.10	4.88	...	4.85	4.38	4.84
TFEC	4.64	5.10	4.88	...	4.85	4.38	4.84
PDLIM5	4.64	5.10	4.88	...	4.85	4.38	4.84
CAPG	4.64	5.10	4.88	...	4.85	4.38	4.84
CUX1	4.64	5.10	4.88	...	4.85	4.38	4.84
ATM	4.64	5.10	4.88	...	4.85	4.38	4.84
...

Pathway analysis

Input – Gene measurements/statistics

DE Genes

Genes
ABCC9
CTNNB1
TSPAN7
RPN2
PRPS1
KCNAB1
FGFR3
HCFC1
SPDYA
TFEC
PDLIM5
CAPG

Genes & statistics

Genes	Fold change	p-value
ABCC9	0.44	8.98E-02
CTNNB1	1.62	8.01E-01
TSPAN7	0.56	9.65E-02
RPN2	-0.38	2.62E-02
PRPS1	-0.10	3.88E-01
KCNAB1	1.89	1.25E-01
FGFR3	0.50	1.16E-02
HCFC1	1.54	3.49E-01
SPDYA	0.04	1.86E-01
TFEC	0.55	6.04E-01
PDLIM5	0.02	4.80E-09
CAPG	0.28	6.24E-10

Full expression matrix

	S1	S2	S3	...	Sx	Sy	Sz
ABCC9	5.87	6.20	5.94	...	5.84	5.43	5.96
CTNNB1	5.66	5.61	5.57	...	5.64	5.82	5.53
TSPAN7	7.95	7.65	7.84	...	7.87	7.76	7.98
RPN2	7.17	6.67	7.25	...	7.47	6.80	7.40
PRPS1	4.38	4.58	4.41	...	4.23	4.28	4.10
KCNAB1	4.84	4.91	4.87	...	4.55	5.20	5.01
FGFR3	4.52	4.53	4.46	...	4.39	4.41	4.46
HCFC1	4.01	3.74	3.99	...	3.89	3.80	3.93
SPDYA	4.64	5.10	4.88	...	4.85	4.38	4.84
TFEC	4.64	5.10	4.88	...	4.85	4.38	4.84
PDLIM5	4.64	5.10	4.88	...	4.85	4.38	4.84
CAPG	4.64	5.10	4.88	...	4.85	4.38	4.84

Over representation analysis
(ORA)

1. FGSEA
2. KS test
3. Wilcoxon test

1. Impact Analysis (topology-based)
- 2-4. GSEA, GSA, PADOG
- 5-8. FGSEA, KS test, Wilcox test, ORA

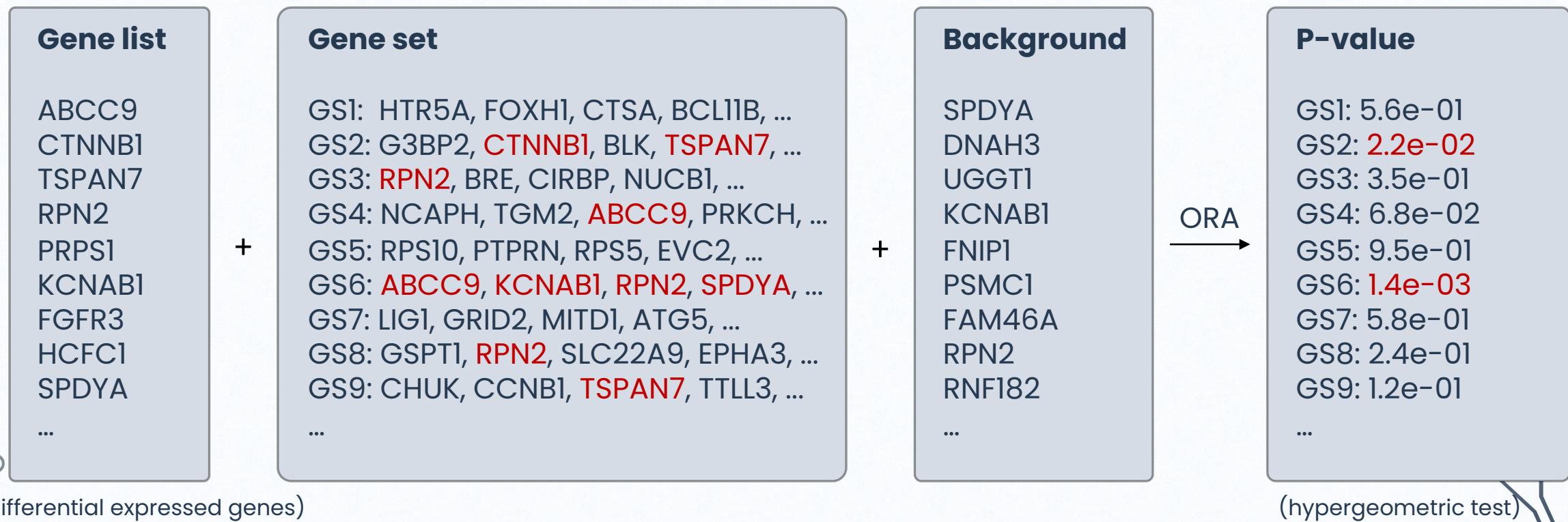
Pathway analysis

Methods (cont.)

Over representation analysis:

Required input: A list of differentially expressed genes

Optional input: A list of background genes



(Differential expressed genes)

(hypergeometric test)

Pathway analysis

Methods (cont.)

Pre-rank gene set/pathway analysis:

Required input: A list of genes and their statistic



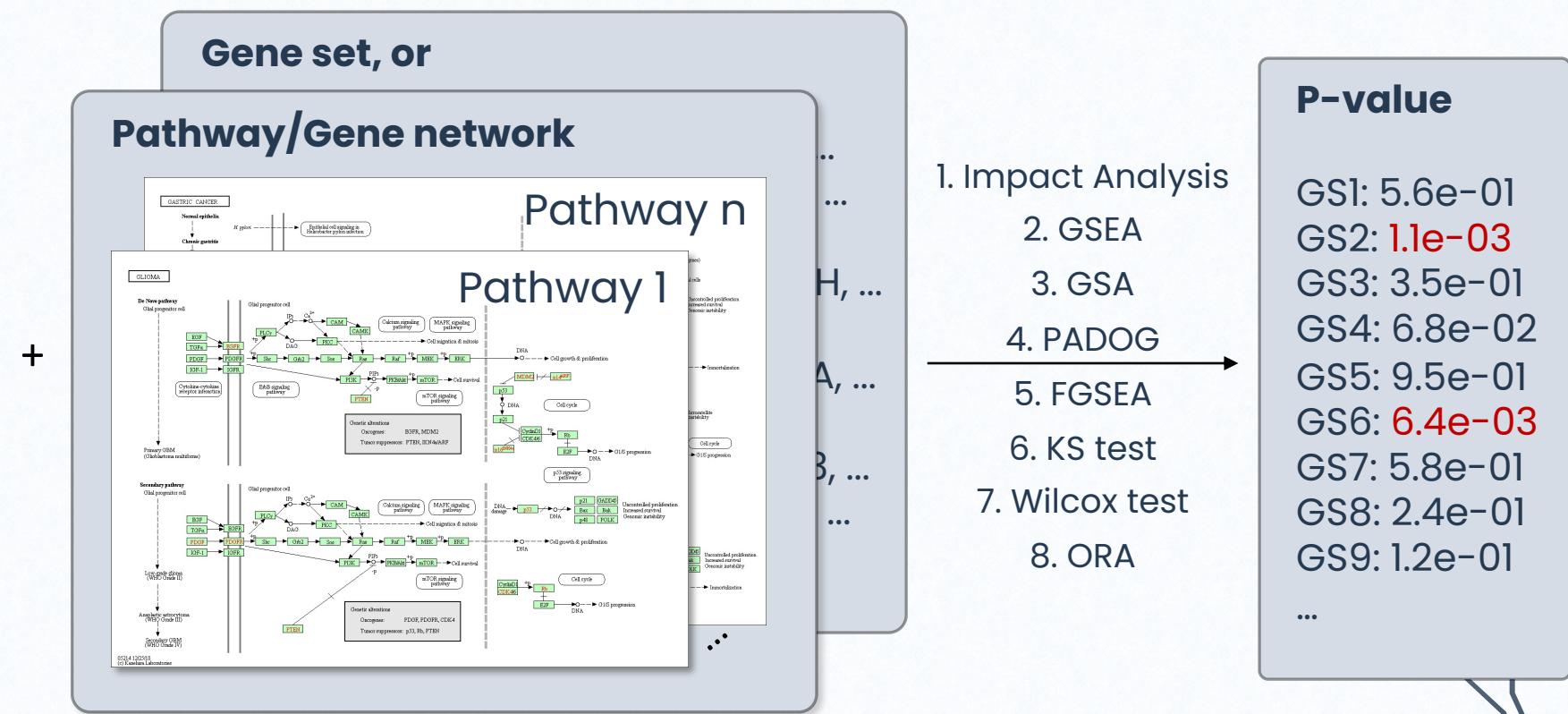
Pathway analysis

Methods (cont.)

Pathway analysis with gene expression data:

Required input: A gene expression matrix and sample grouping

	S1	S2	...	Sx	Sy
ABCC9	5.87	6.20	...	5.84	5.43
CTNNB1	5.66	5.61	...	5.64	5.82
TSPAN7	7.95	7.65	...	7.87	7.76
RPN2	7.17	6.67	...	7.47	6.80
PRPS1	4.38	4.58	...	4.23	4.28
KCNAB1	4.84	4.91	...	4.55	5.20
FGFR3	4.52	4.53	...	4.39	4.41
HCFC1	4.01	3.74	...	3.89	3.80
SPDYA	4.64	5.10	...	4.85	4.38
TFEC	4.64	5.10	...	4.85	4.38
PDLIM5	4.64	5.10	...	4.85	4.38
CAPG	4.64	5.10	...	4.85	4.38
CUX1	4.64	5.10	...	4.85	4.38
ATM	4.64	5.10	...	4.85	4.38
...



Pathway analysis

Output

Typical output:

1. P-value for each gene set/pathway (all methods)
2. Enrichment score (GSEA, FGSEA, GSA, PADOG)
3. Perturbation factor (Impact analysis)

	Name	#Genes	P-value
1	Amyotrophic lateral sclerosis	364	2.12E-07
2	Parkinson disease	249	4.00E-07
3	Pathways of neurodegeneration - multiple diseases	475	3.97E-06
4	Huntington disease	306	2.11E-05
5	Oxidative phosphorylation	133	3.91E-05
6	Alzheimer disease	369	3.91E-05
7	Prion disease	273	8.18E-05
8	Synaptic vesicle cycle	78	8.56E-05
9	Pyruvate metabolism	39	7.66E-04
10	Spinocerebellar ataxia	143	3.27E-03
11	Endocytosis	252	3.87E-03
12	regulation of macroautophagy	41	1.59E-02
13	Proteasome	46	2.23E-02
14	Endocrine and other factor-regulated calcium reabsorption	53	3.41E-02
15	protein targeting	27	5.48E-02
16	protein K48-linked ubiquitination	40	5.92E-02

Pathway analysis

Frequently asked questions

Do I really have to learn R before I can do any analysis? No.

CPA (<http://cpa.tinnguyen-lab.com/>) is a web application with a user-friendly interface

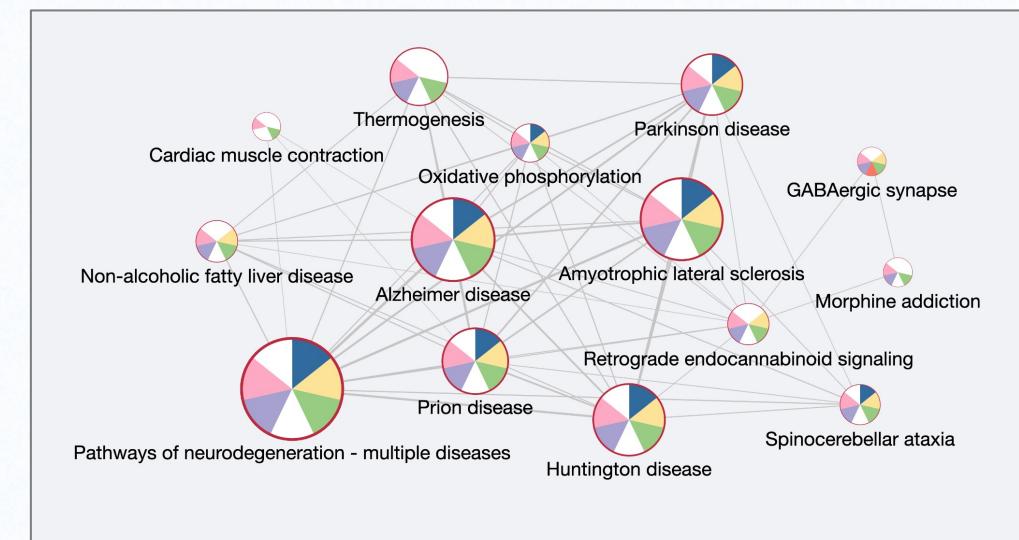
There are too many significant pathways, how to make it easier to interpret.

Users can change many thresholds (p-value, number of significant analyses, etc.)

Users can selectively hide irrelevant pathways

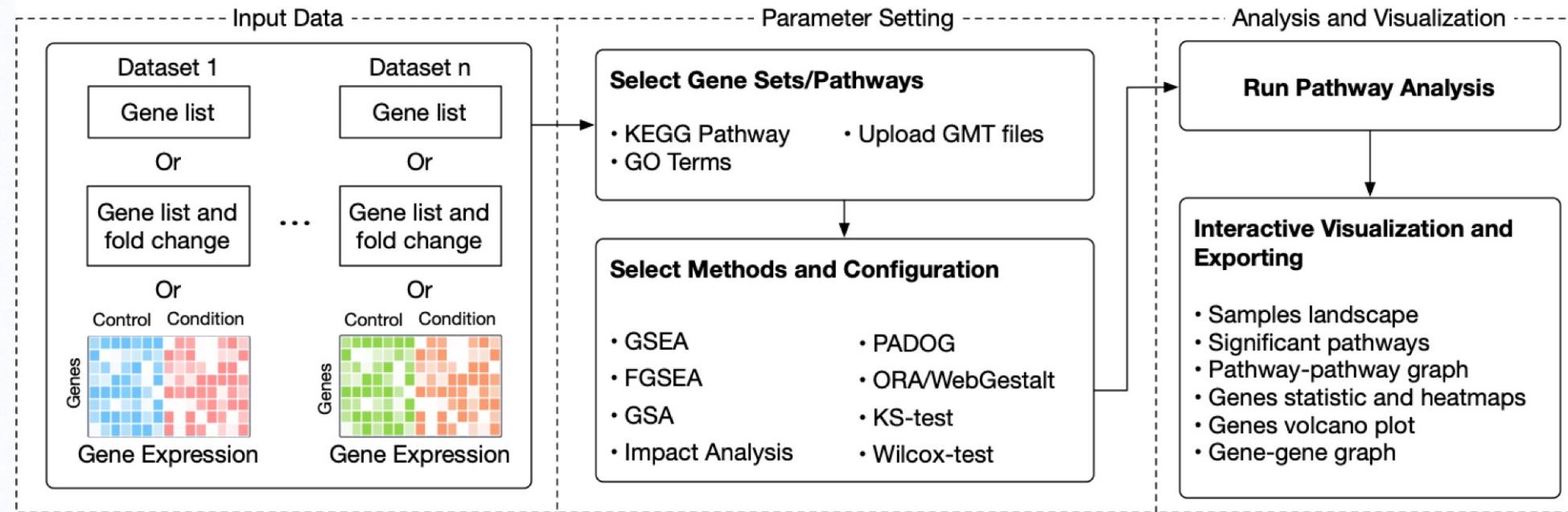
I have data obtained from multiple experiments, can I combine them? Yes.

- Combine the p-values
- Vote counting
- CPA implements both and more ...



CPA: Consensus Pathway Analysis and interactive visualization

<http://cpa.tinnguyen-lab.com>



CPA offers:

- Analyzing multiple datasets with different inputs and multiple methods **at the same time**
- Interactive visualizations for genes and pathways
- Built-in pathway/gene set from KEGG and GO that supports 1,000+ organisms
- Supports customized pathway annotations in the format of GMT file

Pathway analysis ↴

New analysis ×



Analysis name: New analysis

Input type: Gene list: ORA/WebGestalt

▼ Input

Input gene list or

 Select a file

or

 Use our example

ABCC9
CTNNB1
TSPAN7
RPN2
PRPS1
KCNAB1
FGFR3
HCFC1
SPDYA
TFEC

Input background or

 Select a file

or

 Use our example

GOLM1
POLD4
PIGA
RGL1
ZNF496
ITGA3
RARS2
CABYR
FTH1
SNORA40

Auto detected Gene ID type: Gene_Name

Click here to see all supported ID types.

Select organism and pathways

Organism: Homo sapiens (human) ▾

Enter text to search for more organisms. Click here to see all supported organisms. If the interested organism is not supported, please upload a GMT file. The ID type in the GMT file must match the ID type in the input. Download example GMT file with Gene IDs here or with Gene Symbols here.

KEGG (285)

GO (1380)

 Upload GMT file

Export table

Id	Name	#Genes	#Background genes	#Common genes
path:hsa00010	Glycolysis / Gluconeogenesis	67	47	2
path:hsa00020	Citrate cycle (TCA cycle)	30	29	1
path:hsa00030	Pentose phosphate pathway	30	25	3
path:hsa00040	Pentose and glucuronate interconversions	34	17	1

 Start analysis

Analysis name: New analysis

Input type: Gene list and fold change: KS, Wilcox, FGSEA

▼ Input

Input gene list and their statistic or

 Select a file

or

 Use our example

GOLM1 0.377827741
POLD4 0.442827998
PIGA 1.615203202
RGL1 0.558019062
ZNF496 -0.384958431

Auto detected Gene ID type:

 Gene_Name Click here to see all supported ID types.

Select organism and pathways

Organism: Homo sapiens (human)

Enter text to search for more organisms. Click [here](#) to see all supported organisms. If the interested organism is not supported, please upload a GMT file. The ID type in the GMT file must match the ID type in the input. Download example GMT file with Gene IDs [here](#) or with Gene Symbols [here](#).

KEGG (336)

GO (3206)

 Upload GMT file Export table

<input checked="" type="checkbox"/>	Id	Name	#Genes	#Common genes
<input checked="" type="checkbox"/>	path:hsa00010	Glycolysis / Gluconeogenesis	67	47
<input checked="" type="checkbox"/>	path:hsa00020	Citrate cycle (TCA cycle)	30	29
<input checked="" type="checkbox"/>	path:hsa00030	Pentose phosphate pathway	30	25
<input checked="" type="checkbox"/>	path:hsa00040	Pentose and glucuronate interconversions	34	17
—	—	—	—	—

Configure methods: Wilcox Test KS Test FGSEA

Wilcoxon signed-rank test

Enabled: Start analysis

Analysis name: New analysis

Input type: Expression matrix: 8 pathway analysis methods ▾

▼ Input

Expression file: GSE48350-expression.csv

Group file ⚠: GSE48350-group.csv

(Optional. If group file is not selected, manually selecting samples for analysis is required.)

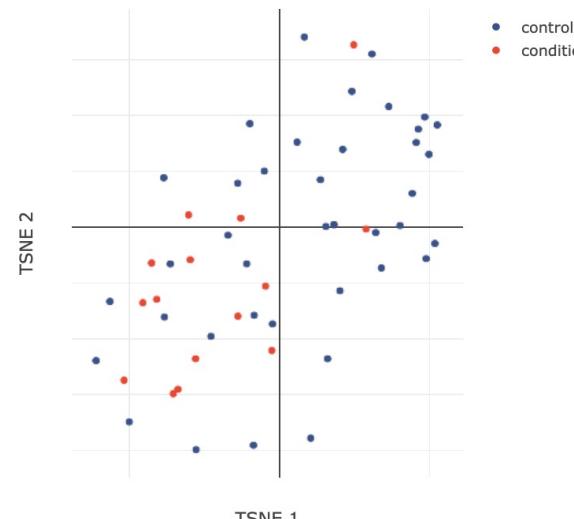
#	GSM300173	GSM300177	GSM300181	GSM300186	GSM300189	GSM300192	GSM300196	GSM300204	GSM300208	GSM300214
1	6.250802	6.002711	6.289045	6.566387	6.180764	6.504784	6.122299	6.07701	6.676641	5.981869
10	5.568055	5.324981	5.142599	5.928485	5.381474	5.14943	5.112492	5.387096	5.46842	5.251948
100	6.208615	6.510181	6.453027	7.416333	7.2724	6.838577	6.884401	6.616268	7.155568	6.70778
1000	9.146729	9.183216	9.17766	9.14898	8.982874	9.355639	8.897144	9.135813	8.433094	9.096965

1-10 of 21989 genes < 1 2 3 4 5 ... 2199 >

Auto detected Gene ID type: GenID

Click here to see all supported ID types.

Samples Space



Select samples for each group before starting the comparative analysis.

Control: c ✓

GSM300173 ×	GSM300177 ×	GSM300181 ×	GSM300186 ×	GSM300189 ×	GSM300192 ×
GSM300196 ×	GSM300204 ×	GSM300208 ×	GSM300214 ×	GSM300218 ×	GSM300222 ×
GSM300228 ×	GSM300230 ×	GSM300234 ×	GSM300238 ×	GSM300242 ×	GSM300248 ×
GSM300252 ×	GSM300258 ×	GSM300261 ×	GSM300267 ×	GSM300271 ×	GSM300275 ×
GSM300279 ×	GSM300285 ×	GSM300289 ×	GSM300293 ×	GSM300297 ×	GSM300300 ×
GSM300304 ×	GSM300308 ×	GSM300312 ×	GSM300316 ×	GSM300320 ×	GSM300324 ×
GSM300328 ×	GSM300332 ×	GSM300338 ×			

Condition: d ✓

GSM1176196 ×	GSM1176197 ×	GSM1176198 ×	GSM1176199 ×	GSM1176200 ×	GSM1176201 ×
GSM1176202 ×	GSM1176203 ×	GSM1176204 ×	GSM1176205 ×	GSM1176206 ×	GSM1176207 ×
GSM1176208 ×	GSM1176209 ×	GSM1176210 ×			

Analysis name: New analysis

Input type: Expression matrix: 8 pathway analysis methods

▼ Input

Expression file:

GSE48350-expression.csv

Group file ⚠:

GSE48350-group.csv

(Optional. If group file is not selected, manually selecting samples for analysis is required.)

File Manager Import from NCBI GEO

File size limit : 24MB. Maximum number of files : 10 files. Example folder is read-only. Uploaded files from guests will be automatically deleted after 24 hours. Please login to increase file size and upload more file.

#	Title	File size	Last modified
1	GSE48350-expression.csv	19.3 MB	2021-04-28 18:35:14
10	GSE48350-group.csv	663 B	2021-04-28 18:35:14
100	GSE5281-expression.csv	8.32 MB	2021-04-28 18:35:13
1000	GSE5281-group.csv	276 B	2021-04-28 18:35:13
	GSE84422-expression.csv	8.07 MB	2021-04-28 18:35:14
	GSE84422-group.csv	468 B	2021-04-28 18:35:13

SM300214
981869
251948
70778
096965
2199 >

Auto detected Group

Samples Sp

TSNE 2

TSNE 1

GSM1176170 ×
GSM1176177 ×
GSM1176178 ×
GSM1176177 ×
GSM1176200 ×
GSM1176201 ×
GSM1176202 ×
GSM1176203 ×
GSM1176204 ×
GSM1176205 ×
GSM1176206 ×
GSM1176207 ×
GSM1176208 ×
GSM1176209 ×
GSM1176210 ×

Analysis name: New analysis

Input type: Expression matrix: 8 pathway analysis methods

Input

Expression file:

GSE48350-expression.csv

Group file

GSE48350-group.csv

(Optional. If group file is not selected, manually selecting samples for analysis is required.)

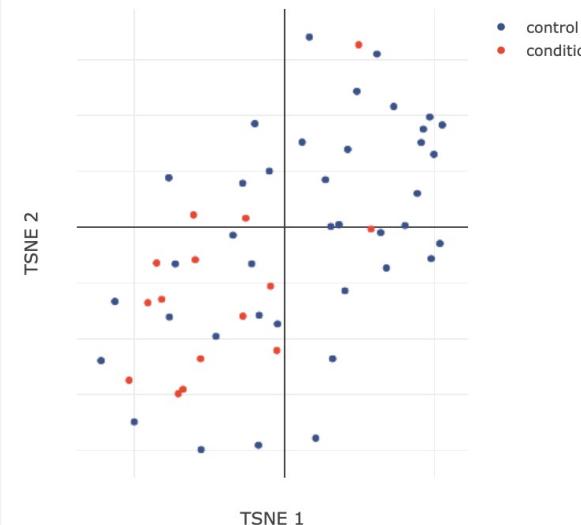
#	GSM300173	GSM300177	GSM300181	GSM300186	GSM300189	GSM300192	GSM300196	GSM300204	GSM300208	GSM300214
1	6.250802	6.002711	6.289045	6.566387	6.180764	6.504784	6.122299	6.07701	6.676641	5.981869
10	5.568055	5.324981	5.142599	5.928485	5.381474	5.14943	5.112492	5.387096	5.46842	5.251948
100	6.208615	6.510181	6.453027	7.416333	7.2724	6.838577	6.884401	6.616268	7.155568	6.70778
1000	9.146729	9.183216	9.17766	9.14898	8.982874	9.355639	8.897144	9.135813	8.433094	9.096965

1-10 of 21989 genes ...

Auto detected Gene ID type: GenID

[Click here to see all supported ID types.](#)

Samples Space



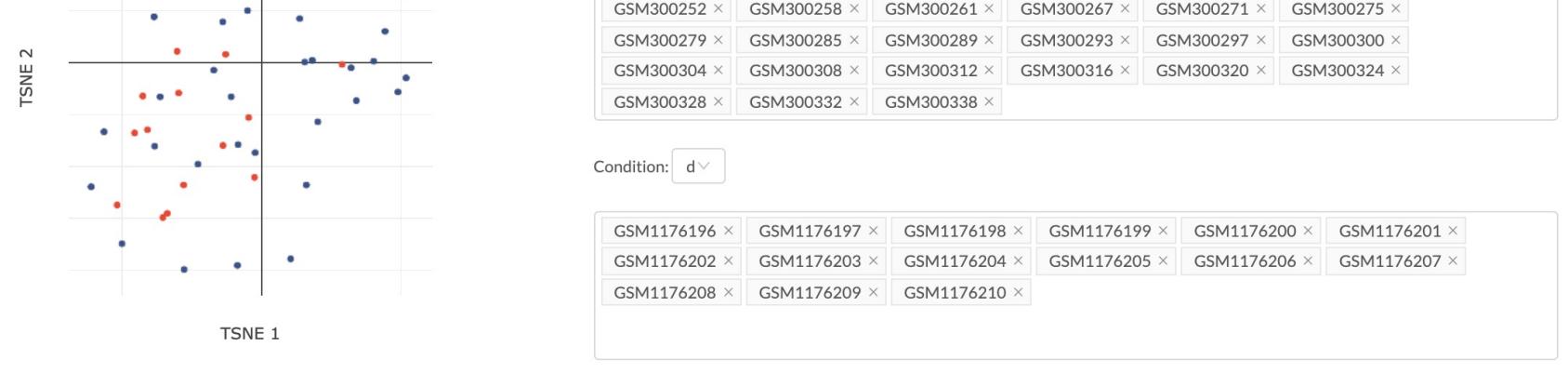
Select samples for each group before starting the comparative analysis.

Control: c

GSM300173 ×	GSM300177 ×	GSM300181 ×	GSM300186 ×	GSM300189 ×	GSM300192 ×
GSM300196 ×	GSM300204 ×	GSM300208 ×	GSM300214 ×	GSM300218 ×	GSM300222 ×
GSM300228 ×	GSM300230 ×	GSM300234 ×	GSM300238 ×	GSM300242 ×	GSM300248 ×
GSM300252 ×	GSM300258 ×	GSM300261 ×	GSM300267 ×	GSM300271 ×	GSM300275 ×
GSM300279 ×	GSM300285 ×	GSM300289 ×	GSM300293 ×	GSM300297 ×	GSM300300 ×
GSM300304 ×	GSM300308 ×	GSM300312 ×	GSM300316 ×	GSM300320 ×	GSM300324 ×
GSM300328 ×	GSM300332 ×	GSM300338 ×			

Condition: d

GSM1176196 ×	GSM1176197 ×	GSM1176198 ×	GSM1176199 ×	GSM1176200 ×	GSM1176201 ×
GSM1176202 ×	GSM1176203 ×	GSM1176204 ×	GSM1176205 ×	GSM1176206 ×	GSM1176207 ×
GSM1176208 ×	GSM1176209 ×	GSM1176210 ×			



Select organism and pathways

Organism: Homo sapiens (human) Enter text to search for more organisms. Click [here](#) to see all supported organisms. If the interested organism is not supported, please upload a GMT file. The ID type in the GMT file must match the ID type in the input. Download example GMT file with Gene IDs [here](#) or with Gene Symbols [here](#).

KEGG (336) GO (3248)

Id Name #Genes #Common genes

<input checked="" type="checkbox"/>	path:hsa00010	Glycolysis / Gluconeogenesis	67	65
<input checked="" type="checkbox"/>	path:hsa00020	Citrate cycle (TCA cycle)	30	30
<input checked="" type="checkbox"/>	path:hsa00030	Pentose phosphate pathway	30	30
<input checked="" type="checkbox"/>	path:hsa00040	Pentose and glucuronate interconversions	34	30
<input checked="" type="checkbox"/>

[Upload GMT file](#) [Export table](#)

Configure methods: ORA/WebGestalt Wilcox Test KS Test GSA GSEA FGSEA PADOG Impact Analysis

Fast Gene Set Enrichment Analysis

Website: <http://bioconductor.org/packages/release/bioc/html/fgsea.html>

Manual: <http://bioconductor.org/packages/release/bioc/manuals/fgsea/man/fgsea.pdf>

Enabled: Permutation ②: 1000 Minimum size ②: 15 Maximum size ②: 500

[Start analysis](#)

Alzheimer's disease ↴

GSE5281 X GSE84422 X GSE48350 X



Analysis name: GSE5281

Input type: Expression matrix: 8 pathway analysis methods

Results:

Highlight: pValue ≤ 0.05 pValue.fdr ≤ 0.05

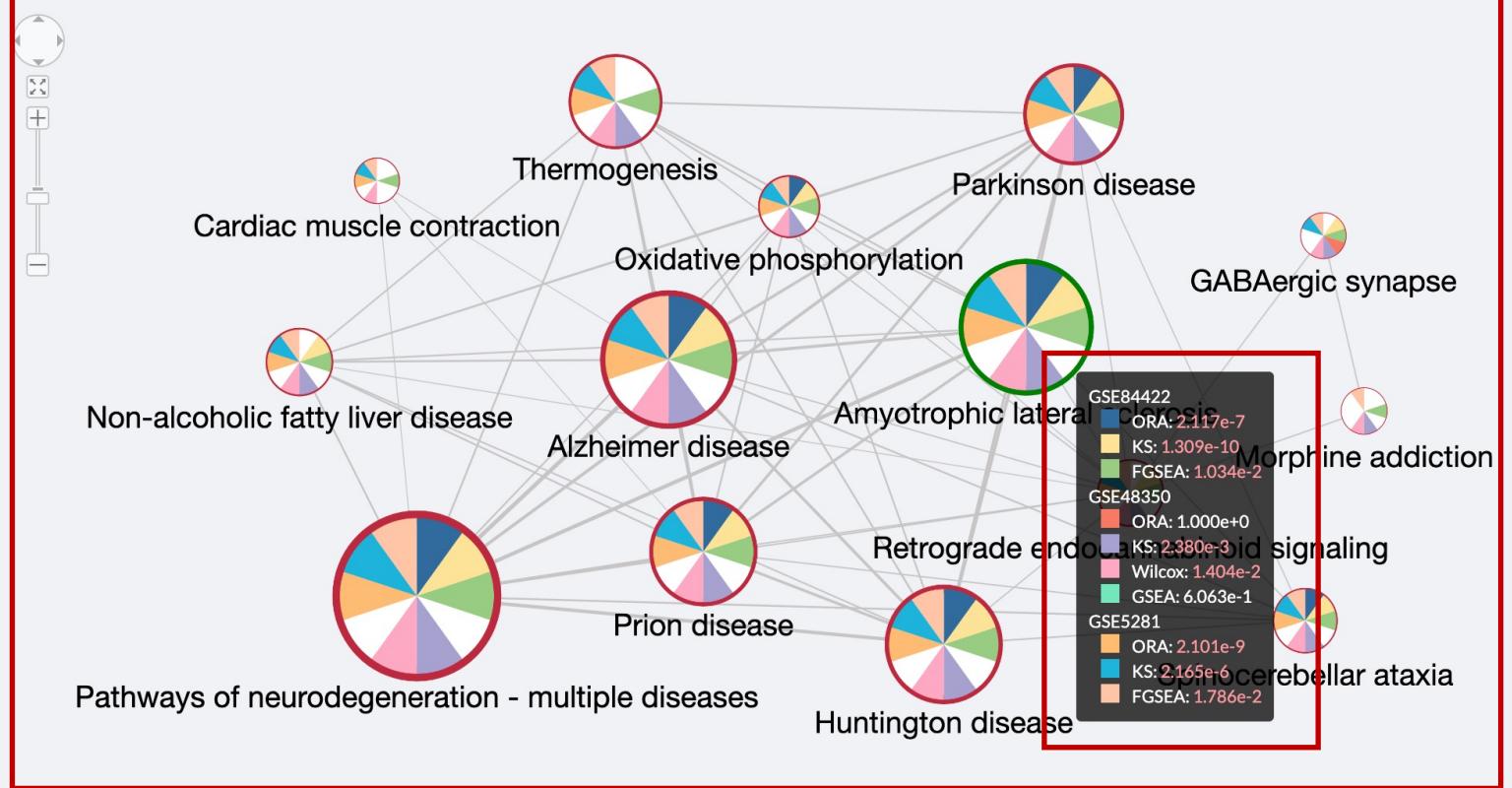
KEGG GO

Id	Name	#Genes	ORA/We bGestalt	ORA/We bGestalt- FDR	KS Test	KS Test- FDR	FGSEA	FGSEA- FDR
path:hsa00010	Glycolysis / Gluconeogenesis	67	2.383e-2	2.224e-1	1.722e-1	4.782e-1	1.126e-3	1.786e-1
path:hsa00020	Citrate cycle (TCA cycle)	30	4.442e-3	7.107e-2	5.992e-3	5.799e-2	5.102e-3	5.645e-1
path:hsa00030	Pentose phosphate pathway	30	9.963e-2	5.812e-1	2.485e-1	5.511e-1	1.276e-3	1.786e-1
path:hsa00040	Pentose and glucuronate interconversions	34	9.268e-1	1.000e+0	5.092e-2	2.676e-1	8.776e-1	1.000e+0
path:hsa00051	Fructose and mannose metabolism	33	2.683e-2	2.373e-1	6.575e-2	3.031e-1	7.576e-3	6.527e-1
path:hsa00052	Galactose metabolism	31	4.088e-1	1.000e+0	6.002e-1	8.034e-1	1.863e-1	4.437e-1
path:hsa00053	Ascorbate and aldarate metabolism	30	9.801e-1	1.000e+0	4.137e-3	4.793e-2	4.846e-1	8.265e-1
path:hsa00061	Fatty acid biosynthesis	18	4.075e-1	1.000e+0	5.949e-1	8.028e-1	2.016e-1	4.703e-1
path:hsa00062	Fatty acid elongation	27	6.186e-2	4.330e-1	4.499e-1	7.098e-1	5.548e-2	2.220e-1
path:hsa00071	Fatty acid degradation	43	3.586e-1	1.000e+0	2.241e-1	5.377e-1	2.069e-1	4.762e-1
path:hsa00072	Synthesis and degradation of ketone bodies	10	3.990e-1	1.000e+0	2.945e-1	6.033e-1	1.000e+0	1.000e+0

Visualize pathways



Heatmap


[Data](#) [Graph settings](#) [Export](#)

Highlight: pValue ≤ 0.05 pValue.fdr ≤ 0.05 Show pvalue-FDR only Hide results Quick selection Combine pvalues

<http://cpa.tinnguyen-lab.com>

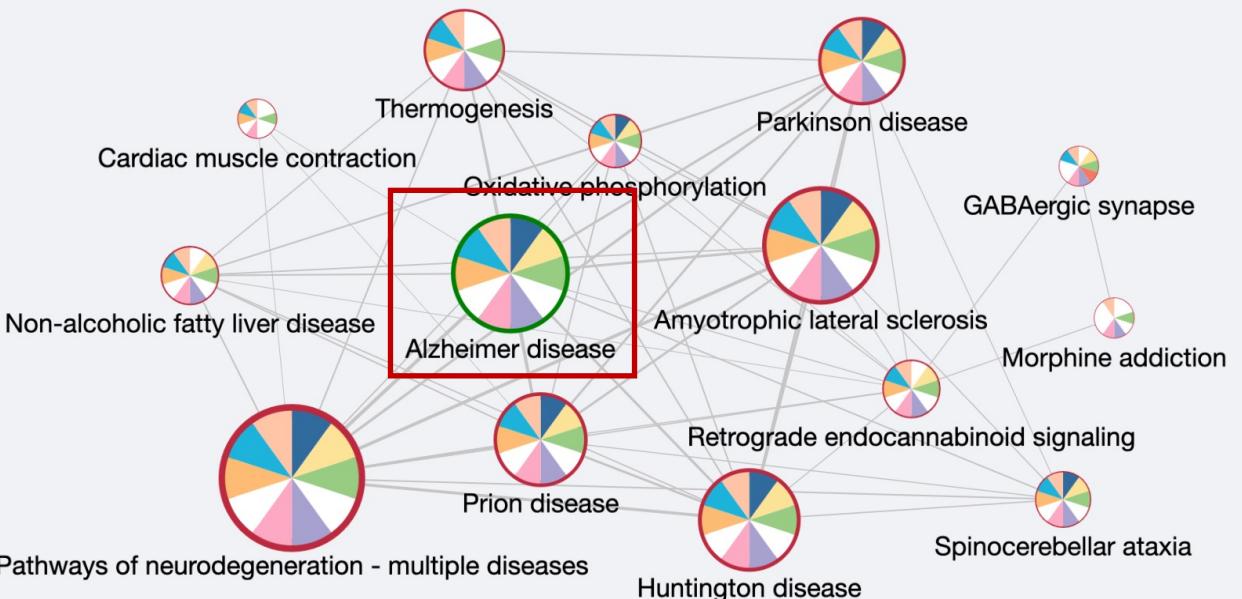
Database	Id	Name	#Genes	GSE84422			GSE48350				
				ORA-FDR	KS-FDR	FGSEA-FDR	ORA-FDR	Wilcox-FDR	KS-FDR	GSEA-FDR	ORA-FDR
KEGG	path:hsa00010	Glycolysis / Gluconeogenesis	67	6.910e-1	6.919e-2	3.146e-1	1.000e+0	6.042e-1	5.914e-1	6.063e-1	2.224e-1
KEGG	path:hsa00020	Citrate cycle (TCA cycle)	30	8.955e-2	2.676e-1	1.034e-2	1.000e+0	9.794e-2	2.601e-1	6.063e-1	7.107e-2
KEGG	path:hsa00030	Pentose phosphate pathway	30	1.000e+0	7.426e-1	1.748e-1	1.000e+0	2.234e-1	2.901e-1	6.063e-1	5.812e-1
KEGG	path:hsa00040	Pentose and glucuronate interconversions	34	1.000e+0	5.026e-1	1.000e+0	1.000e+0	3.107e-1	3.233e-1	9.798e-1	1.000e+0



C
B
A

Heatmap

Id	Symbol	Description	GSE84422		GSE48350	
			pFDR	logFC	pFDR	logFC
10000	AKT3	AKT serine/threonine kinase 3	6.191e-1	0.064066	2.566e-1	0.083984
10023	FRAT1	FRAT regulator of WNT signaling pathway 1	3.939e-1	-0.056448	5.588e-2	0.156805
100506742	CASP12	caspase 12 (gene/pseudogene)	-	-	8.606e-1	0.007587
100532726	NDUFC2-KCTD14	NDUFC2-KCTD14 readthrough	9.282e-1	-0.006101	2.366e-3	-0.150066
10105	PPIF	peptidylprolyl isomerase F	3.633e-1	0.099755	2.578e-1	0.105219
102	ADAM10	ADAM metallopeptidase domain 10	9.477e-1	0.005199	7.704e-1	0.02449
1020	CDK5	cyclin dependent kinase 5	2.417e-1	0.223349	5.890e-3	-0.535189
10213	PSMD14	proteasome 26S subunit, non-ATPase 14	7.864e-2	0.32746	3.507e-2	-0.158322
102800317	TPTEP2-CSNK1E	TPTEP2-CSNK1E readthrough	5.504e-1	-0.054341	5.893e-1	0.031694
10297	APC2	APC regulator of WNT signaling pathway 2	3.695e-1	-0.044695	2.109e-1	0.052817
10313	RTN3	reticulon 3	4.530e-1	0.150486	2.730e-3	-0.363949
10376	TUBA1B	tubulin alpha 1b	8.985e-1	-0.022938	5.341e-4	-0.329838


Data
Graph settings
Export

 Highlight: pValue ≤ 0.05 pValue.fdr ≤ 0.05 Show pvalue-FDR only Hide results Quick selection Combine pvalues

X

Id	Database	GSE84422				GSE48350						
		KEGG	Name	#Genes	ORA-FDR	KS-FDR	FGSEA-FDR	ORA-FDR	Wilcox-FDR	KS-FDR	GSEA-FDR	ORA-FDR
path:hsa05010	KEGG	Alzheimer disease		369	3.908e-5	5.414e-11	1.034e-2	1.000e+0	7.317e-4	3.835e-4	6.063e-1	3.518e-9

GSE84422 GSE48350 GSE5281

P-value threshold Δ : 1

Cardiac

Non-alcoholic fa

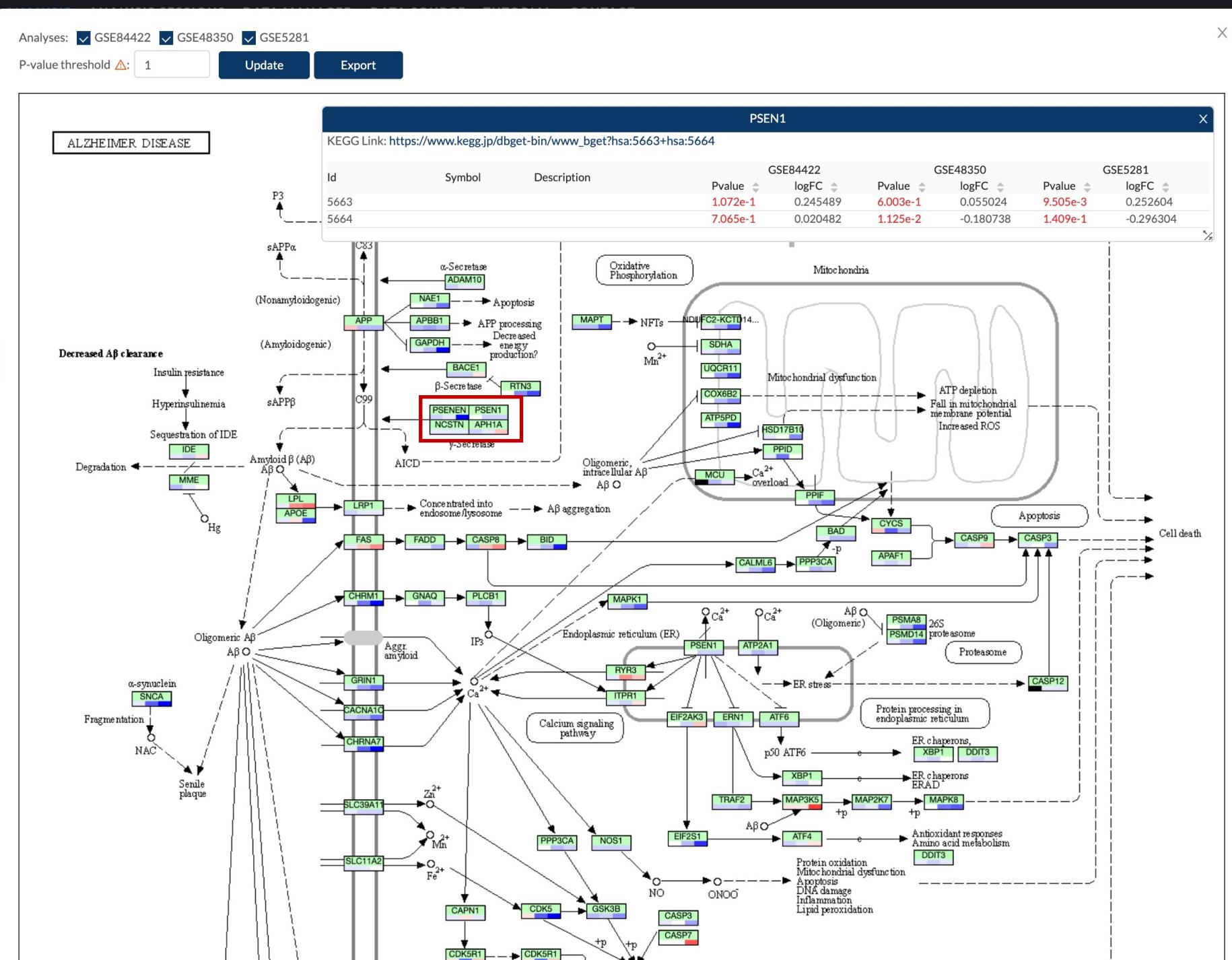
Pathways of neu

Data Graph settings Export

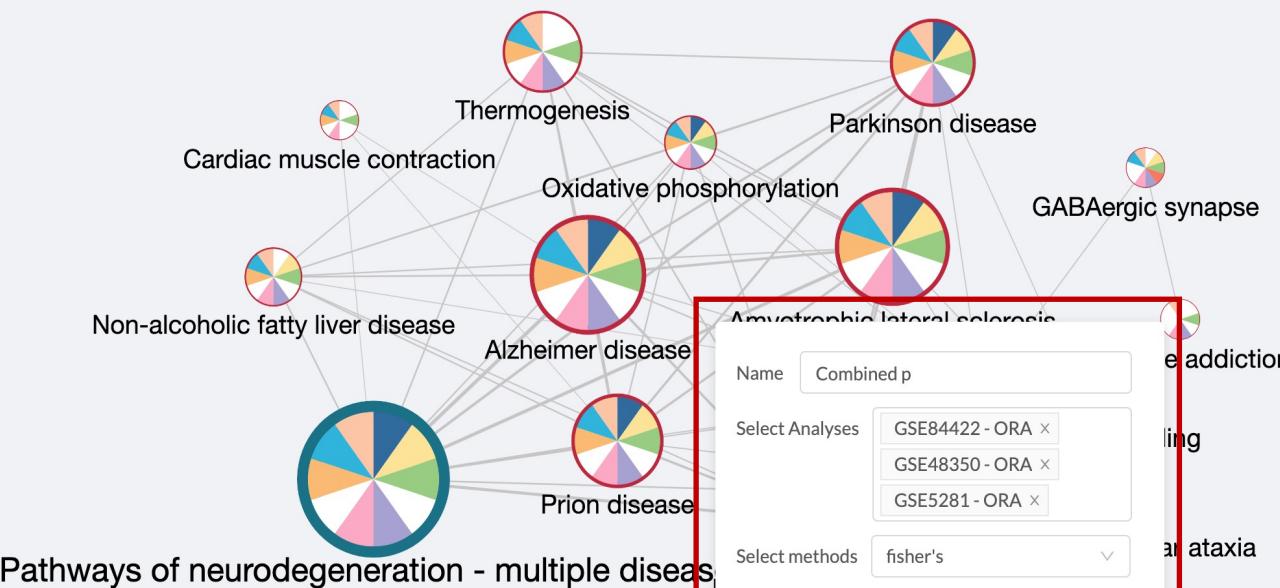
Highlight: pValue \leq 0.05

Id Database KEGG

path:hsa05010 KEGG



4422	GSE48350
logFC	pFDR
0.064066	2.566e-1
-0.056448	5.588e-2
-	0.156805
8.606e-1	0.007587
-0.006101	2.366e-3
0.099755	2.578e-1
0.005199	7.704e-1
0.223349	5.890e-3
0.32746	3.507e-2
-0.054341	5.893e-1
-0.044695	2.109e-1
0.150486	2.730e-3
-0.022938	5.341e-4
FDR	ORA-FDR
-1	3.518e-9
1	

<
□
C


Amyotrophic lateral sclerosis

Name	Combined p
Select Analyses	GSE84422 - ORA × GSE48350 - ORA × GSE5281 - ORA ×
Select methods	fisher's
Min p-value	0.0000000001

Note: combining p-values of different methods for the same dataset might lead to artificially low meta p-values.

Apply

Data
Graph settings
Export
Highlight: pValue ≤ 0.05
pValue.fdr ≤ 0.05
Show pvalue-FDR only

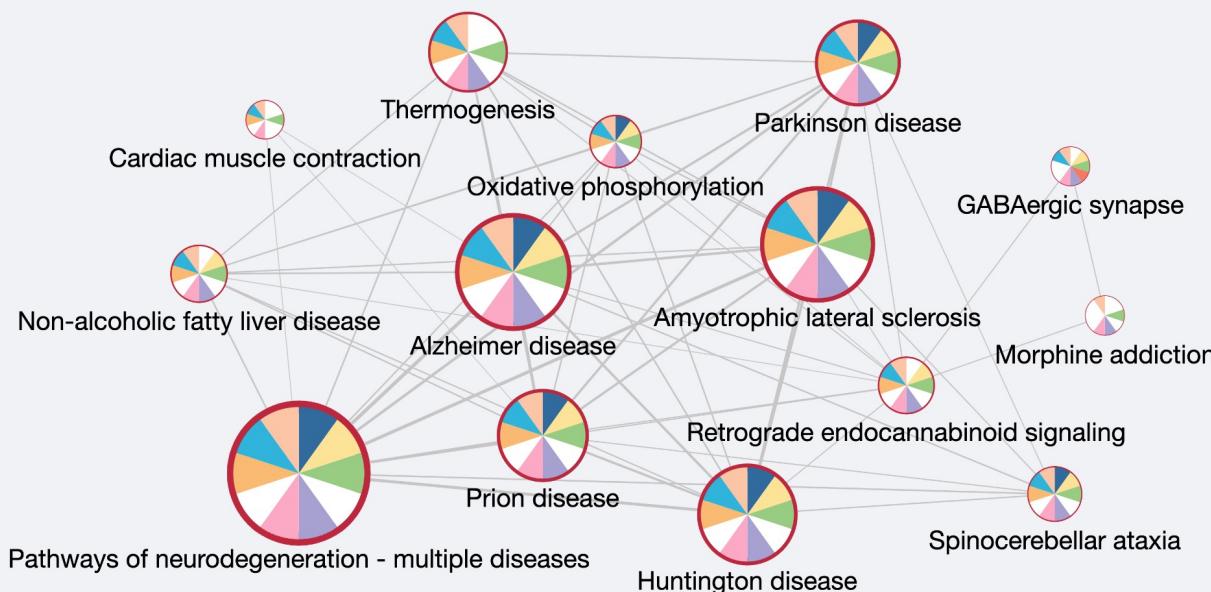
Hide results
Quick selection
Combine pvalues

GSE84422

	Id	ORA-FDR	KS-FDR	FGSEA-FDR	ORA-FDR	Wilcox-FDR	KS-FDR	GSEA-FDR	ORA-FDR	KS-FDR	FGSEA-FDR	Combined p	Combined p-FDR
<input checked="" type="checkbox"/>	path:hsa05014	1.17e-7	1.309e-10	1.034e-2	1.000e+0	1.404e-2	2.380e-3	6.063e-1	2.101e-9	2.165e-6	1.786e-2	6.171e-17	2.243e-13
<input checked="" type="checkbox"/>	path:hsa05012	1.04e-7	0.000e+0	1.034e-2	1.000e+0	4.043e-4	1.091e-4	6.063e-1	0.000e+0	4.103e-13	1.786e-2	2.296e-16	4.171e-13
<input checked="" type="checkbox"/>	path:hsa05022	1.65e-6	4.444e-10	1.034e-2	1.000e+0	1.239e-5	6.619e-6	6.063e-1	1.226e-11	1.598e-8	1.786e-2	2.657e-15	3.219e-12
<input checked="" type="checkbox"/>	path:hsa05016	1.10e-5	4.265e-12	1.034e-2	1.000e+0	2.464e-5	1.763e-5	6.063e-1	4.197e-12	8.909e-10	1.786e-2	1.775e-14	1.612e-11

GSE48350

GSE5281

[←](#) [↻](#) [C](#)

Data

Graph settings

Export

...

General settings:

Graph layout: Fcose Show geneset database name: Node color: Node size scale (%): 100 Minimum common genes: 36 Edge color: Edge size scale (%): 10 Text color: Text size: 35

Border color: Border size scale (%): 200 Selection Mode: Union Minimum edge: 1

Pie node settings:Use FDR correction: **GSE84422:**

ORA: KS: FGSEA:

GSE48350:

ORA: KS: Wilcox: GSEA:

GSE5281:

ORA: KS: FGSEA:

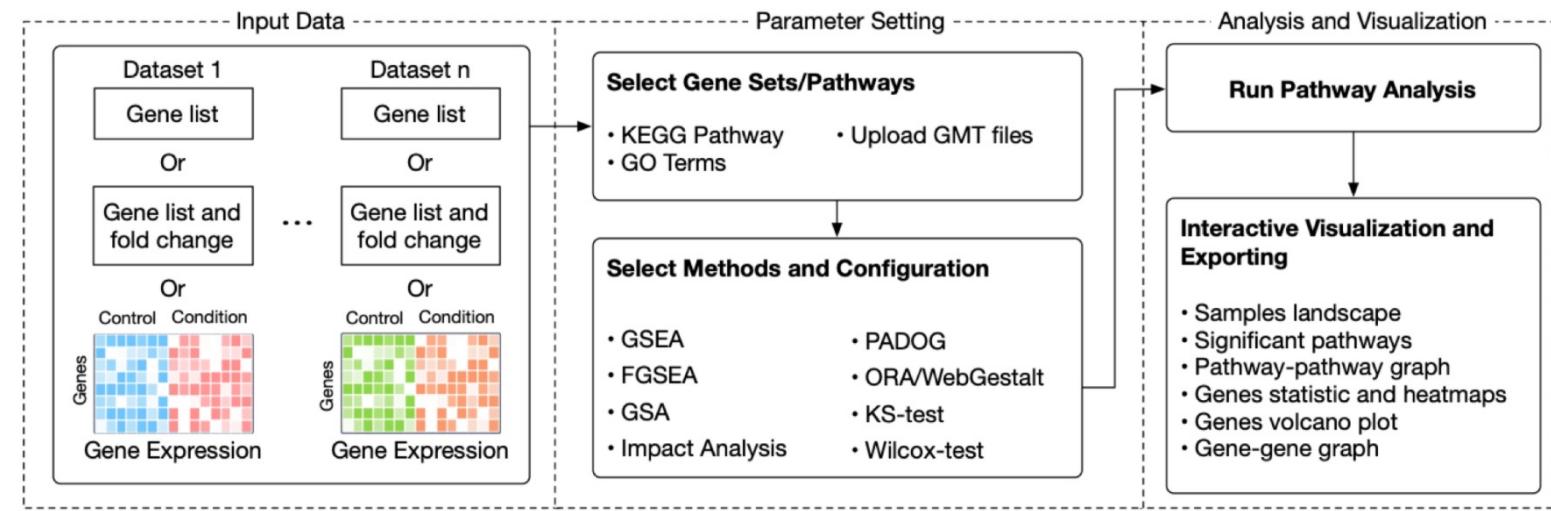
Tutorial

Overall workflow

The workflow of CPA includes three main steps:

- **Data input:** Upload a gene list, a gene list and their fold change, or expression data and its grouping file.
- **Parameter setting:** Select methods and configure their parameters.
- **Visualization:** Visualize the result once all analyses are done.

Each of the step will be described in the following sections. We also provide an example for analysis using three datasets ([GSE48350](#), [GSE5281](#), [GSE84422](#)) available [here](#) for a quick overview of the tools. This analysis is read-only (users cannot modify its settings). The visualization for its analysis results is available is available [here](#). 



Data preparation

The CPA support more than 1,000 organisms from KEGG and GO. Click [here](#) to list all supported organisms.

CPA is free to use without registration. However, data uploaded from anonymous users will be automatically deleted after 24 hours. To save your data and access your analyses across devices, please login using the [Login](#) button on the top right corner or click [here](#) to login using your Google account. The system only requests your name and email address. No other permission are requested.

CPA supports three different inputs, each input support analysis with different methods:

- Gene list: Over-representation analysis (ORA)
- Gene list and fold change: Kolmogorov-Smirnov test (KS Test), Wilcoxon signed-rank test (Wilcox Test), Fast Pre-ranked Gene Set Enrichment Analysis (FGSEA)
- Expression data matrix: ORA, KS test, Wilcox test, FGSEA, GSA (Geneset Analysis), Gene Set Enrichment Analysis (GSEA), Pathway Analysis with Down-weighting of Overlapping Genes (PADOG), and Impact

Other related products

<https://bioinformatics.cse.unr.edu/home/software>

CANCER SUBTYPING

- **SCFA**: subtyping and risk prediction
GitHub: <https://github.com/duct317/SCFA>
Bioconductor: <https://bioconductor.org/packages/SCFA/>
- **PINSPlus**: subtyping and big data analysis
CRAN: <https://CRAN.R-project.org/package=PINSPlus>
- **SMRT**: subtyping and big data analysis
Web application: <http://smrt.tinnguyen-lab.com/>

DATABASE

- **FAT-PTM**: examining post-translational modification in plants
Web application: <https://bioinformatics.cse.unr.edu/fat-ptm/>

PATHWAY ANALYSIS

- **CPA**: Consensus pathway analysis and interactive pathway visualization
Web application: <https://bioinformatics.cse.unr.edu/software/cpa/>
- **BLMA**: a suit of tools for bi-level meta-analysis
Bioconductor: <https://bioconductor.org/packages/release/bioc/html/BLMA.html>

SINGLE-CELL ANALYSIS

- **scDHA**: Fast and precise single-cell data analysis using a hierarchical autoencoder
Publication: <https://www.nature.com/articles/s41467-021-21312-2>
CRAN: <https://cran.r-project.org/web/packages/scDHA/>

CPA: A web-based platform for **C**onsensus **P**athway **A**nalysis and interactive visualization

<http://cpa.tinnguyen-lab.com>

NASA Genelab (Sylvain, Jon, Afshin)
Bioinformatics Lab @ UNR

Q & A

Hung Nguyen, Duc Tran, Jonathan M. Galazka, Sylvain V. Costes, Afshin Beheshti, Juli Petereit, Sorin Draghici, and Tin Nguyen. "CPA: a web-based platform for consensus pathway analysis and interactive visualization." *Nucleic Acids Research* (July 2021).