## **Coexpedia Tutorial**

#### Contents

- Data content of Coexpedia
- How to search Coexpedia
- How to interpret your search results

## Data content of Coexpedia

### Co-expression links inferred from GEO data

• Coexpedia contains co-expression links inferred from microarray data of the main data depository, Gene Expression Omnibus (GEO).

Gene Expression Omnibus

• Human

We analyzed 2,622 GEO series which contains at least 12 samples , comprising 181,142 samples based on 5 platforms:

- Affymetrix Human Genome U133A Array (GPL96)
- Affymetrix Human Genome U133 Plus 2.0 Array (GPL570)
- Affymetrix Human Genome U133A 2.0 Array (GPL571)
- Affymetrix HT Human Genome U133A Array (GPL3921)
- Affymetrix Human Gene 1.0 ST Array [transcript version] (GPL6244)
- Mouse

We analyzed 2,468 GEO series which contains at least 12 samples, comprising 76,002 samples based on 5 platforms:

- Affymetrix Mouse Expression 430A Array (GPL339)
- Affymetrix Mouse Genome 430 2.0 Array (GPL1261)
- Affymetrix Mouse Gene 1.0 ST Array (GPL6246)
- Affymetrix Mouse Genome 430A 2.0 Array (GPL8321)
- We used R packages 'affy' and 'oligo' to obtain and normalize raw gene expression values and normalized them using MAS5 and RMA normalization.
- Finally, we selected 384 and 248 co-expression networks from each GEO series for human and mouse.
  - 384 human GSEs(GEO Series) are from HumanNet v2 project (unpublished data).
  - 183 mouse GSEs from MouseNet v2 (Eiru Kim *et al.* Nucleic Acids Res. 2015) and additional 63 mouse GSEs are selected.
- The constructed co-expression networks of each selected GEO series are consisted of co-expressed gene pairs which have log-likelihood scores (LLS) as edge weights. (Lee *et al.*, Science 2004).

#### Associated biomedical context information

#### A GEO series(GSE) has PubMed information if it is based on publication.



PubMed articles may have several Medical subject heading (MeSH) terms.

Send to: -

☆

"co-expressed gene pairs  $\leftrightarrow$  GSE  $\leftrightarrow$  PubMed  $\leftrightarrow$  MeSH"

All co-expression links are associated with biomedical information by MeSH terms.

# How to search Coexpedia

### Search using a single gene

This is a useful option if you want to **identify co-expressed genes of a query gene** and the associated **functions** (GO biological process) and **phenotypes** (Disease Ontology for human, Mammalian Phenotype for mouse) **of the query gene**.

1 (a) Input a gene or (b) click an example gene		ick the ubmit' button
	Submit Reset   Examples: GWAS catalog Schizophrenia   Also you can input gene sets from GWAS Catalogs Or   GWAS datalog Schizophrenia GWAS datalogs Or	

#### Search using multiple genes

- This is a useful option if you want to know which genes are co-expressed among multiple query genes and what are their associated functions (GO biological process) and phenotypes (Disease Ontology for human, Mammalian Phenotype for mouse).
- Coexpedia provides the **pre-compiled gene lists** from two major databases of GWAS candidate genes: GWAS catalogs and GWASdb.

	Human
	Gene NCBI Gene ID or Symbol
	BRCA1 submit
(a) Input genes , (b) click an example genes, or	Example: #1: BRCA1 Genes NCBI Gene ID or Symbol
(c) select the gene set from GWAScatalog/ GWASdb.	
Click the 'Submit' button	Submit Reset   Examples: GWAS catalog Schizophrenia 1(D)   Also you can input gene sets from GWAS Catalogs Or

# How to interpret your search results

#### Search results (human BRCA1)



from all links) Clicking "View" will show a MeSH

## Network viewer



Network Viewer of a queried gene and its co-expressed neighbors, or Co-expressed queried genes using CytoscapeWeb (Adobe Flash Plugin needed).

Mouse-over to the genes or the MeSH headings on the left and right sidebars to highlight nodes(genes) or edges (links).

#### Relationships among information



Co-expressed gene pairs (edges) of each GSEs have LLS scores. Each gene pairs can have more than one LLS scores from multiple GSEs.



The relationship between GSEs and MeSH headings.

#### Co-expressed genes (left sidebar, first tab)



**'Co-expressed Genes'** on the left sidebar displays Co-expressed genes of the query gene. They are ordered by the sum of the scores of their queried neighbors (Neighbor sum).

Descending OrderGene AScoreGene BScoreGene Cscore

#### Candidate functions and phenotypes (left sidebar, 2<sup>nd</sup> and 3<sup>rd</sup> tab)

Co	GeneSet Analysis Genes GO-BP	et Analysis DO		o-expressed Genes	GeneSet Analysis GO-BP	GeneSet Analysis DO			
	GeneSet Analysis:			GeneSet Analysis:					
Gene Ontology - Biological Process				Disease Ontology					
Enriched terms by p-value < 0.05				Enriched terms by p-value < 0.05					
Rank	GO Acc. and Desc.	p-value	Rank	c	DO ID and Name	p-value			
1	GO:0006260 DNA replication	2.862e-53	1	DOID:162 cancer		3.065e-18			
2	GO:0007062 sister chromatid cohesion	5.072e-51	2	DOID:5683 hereditary	breast ovarian cancer	1.086e-15			
3	GO:000082 G1/S transition of mitotic cell cycle	4.255e-38	3	DOID:4241 malignant r	neoplasm of breast	2.482e-14			
4	GO:0000722 telomere maintenance via recombination	2.901e-24	4	DOID:9256 colorectal c		1.622e-9			
5	GO:000086 G2/M transition of mitotic cell cycle	3.065e-22	5	DOID:768 retinoblasto	oma	3.184e-9			

- Coexpedia also provides Gene Set Analysis of co-expressed genes using annotations of Gene Ontology - Biological Process (GO-BP) (reliable evidence codes only; IDA, IMP, IGI, IPI, IEP, TAS) and **Disease Ontology (DO)**.
- The guery gene, BRCA1, is co-expressed with the genes known to be involved in 'DNA replication', 'sister chromatid cohesion', 'G1/S transition of mitotic cell cycle', and more, which suggest that BRCA1 is likely involved in these functions as well.
- The query gene, BRCA1, is co-expressed with the genes known to be involved in diseases such as 'cancer', 'hereditary breast ovarian cancer', 'malignant neoplasm of breast' and more, which suggest that BRCA1 is involved in such diseases.
- For mouse, DO analysis is replaced with MGI Mammalian Phenotype (MP) analysis. ٠

#### Associated MeSH terms (right sidebar, 1<sup>st</sup> and 2<sup>nd</sup> tab)

excl	MeSH . Neoplasms	incl.	MeSH Neoplasms			Î			
Rank	Rank MeSH h		g links	S	core				
1	Heart		110	2	13.40 <u>View</u>				
2	Myofibroblast	<u>s</u>	90	1	48.97 <u>View</u>				
3	Epithelial Cells		99	1	41.67 <u>View</u>				
4	Th1 Cells	_	00	1	21.61.\/iow	_	_	_	_
5	Th17 Cells	ovel	MeSH		MeS				- 1
5	Immune Syste		Neoplasms	_	incl. Neo	-			
7	Lymphoprolife	Rank	MeSH		eading	lir	iks	Score	
ľ –	Disorders	1	carcinoma,	N	on-Small-	15	3	324.83	View
8	Killer Cells, Na	2	Cell Lung	_	plasms	10	c	204.02	View
9	Inflammation	2	Prostatic Ne	20	plasms			284.02	
9	Inflammatory	3	<u>Carcinoma</u>			13		273.45	
9	Diseases	4	Neuroblast	om	<u>na</u>	11	2	250.96	View
11	<u>Hematopoieti</u>	5	Adenocarci	10	ma	12	6	236.15	<u>View</u>
12	<b>Fibroblasts</b>	6	Heart			11	0	213.40	<u>View</u>
12	Mitochondrial	7	Lymphoma,	Ν	1antle-Cell	11	5	192.90	<u>View</u>
			Uterine Con	÷	<u></u>	_	-		_

MeSH terms are highly biased to cancer (Neoplasms).

Therefore, Coexpedia shows associated MeSH terms in two different ways:

'MeSH excluding Neoplasm' (1<sup>st</sup> tab) 'MeSH including Neoplasm' (2<sup>nd</sup> tab).

**For example**, BRCA1 has been known for a major risk genes for human breast cancer. A query for BRCA1 returns 'Heart' as the top associated MeSH term, and BRCA1 was recently reported as an essential regulator of heart function and survival following myocardial infarction (*Nat. Commun* 2:593, 2011).



Ranking the MeSH headings:

MeSH headings are ranked by calculating the contribution of their associated GEO series in the total sum of LLS scores.

Score of MeSH2 = Summation Scores from {GSE2 , GSE3, GSE4} = LLS2 + (LLS3 + LLS5) + (LLS4 + LLS6)

## MeSH Viewer



#### Search results (human BRCA1 $\rightarrow$ Heart)



# Search results (human Schizophrenia candidate genes from GWAS)



If we submit 16 genes associated with Alzheimer by GWASdb, 'Kidney' comes as second most associated MeSH term, implicating association between Alzheimer and kidney disorder. Indeed, multiple studies reported chronic kidney disease as a risk factor of Alzheimer diseases (reviewed in *Alzheimer's Research and Therapy* 7:29, 2015).

