

COXPRESdb: a database to compare gene coexpression in seven model animals

Takeshi Obayashi and Kengo Kinoshita*

Graduate School of Information Science, Tohoku University, 6-3-09, Aramaki-Aza-Aoba, Aoba-ku, Sendai 980-8679, Japan

Received September 15, 2010; Revised October 23, 2010; Accepted October 25, 2010

ABSTRACT

Publicly available databases of coexpressed gene sets are a valuable resource for a wide variety of experimental studies, including gene targeting for functional identification, and for investigations of regulatory mechanisms or protein–protein interaction networks. Although coexpressed gene databases are becoming more and more popular in the field of plant biology, those with animal data are rather limited, possibly due to the lower reliability of the coexpression data. The original COXPRESdb (coexpressed gene database) (<http://coexpresdb.jp>) represented the coexpression relationship for human and mouse. Here, we report updates of this database that especially focus on the enhancement of the reliability of gene coexpression data in animals. For this purpose, we implemented a new comparable coexpression measure, Mutual Rank, included five other animal species, rat, chicken, zebrafish, fly and nematoda, to assess the conservation of coexpression, and added different layers of omics data into the integrated network of genes. Comparison of coexpression is a key concept to enhance the reliability of gene coexpression, and the integration of different information can reduce the noise inherent in the information. With the functions for gene network representation, COXPRESdb can help researchers to clarify the functional and regulatory networks of genes in a broad array of animal species.

INTRODUCTION

Genes involved in related biological pathways are usually expressed cooperatively for their functions, and thus information on their coexpression is a key to understand the biological systems at the molecular level (1). Coexpression data have been utilized in a wide variety of experimental

designs, such as gene targeting, regulatory investigations and/or identification of potential partners in protein–protein interactions (PPIs) (2). A reliable estimation of coexpressed gene relationships requires large amounts of gene expression data from DNA microarray experiments, which are now available in various public repositories (3,4) for several species. Using these large public data sources, a number of coexpression databases have been constructed, and are widely used. However, most of the databases have been constructed for plant researchers (2), and thus the use of coexpression data in animal research is rather limited.

One possible reason for the limited use of coexpression data in animals is that such coexpression data showed lower performance in identifying the functionally related genes for mammals than for Arabidopsis (5). Among the many possible reasons for this, one reason is the lower variety of microarray samples in animals. Although more data are usually available for animals than for plants, the expression data for animals are quite biased, due to the inclusion of medical samples, such as cancer cells. On the contrary, in the field of Arabidopsis, systematic collections were promoted by the AtGenExpress international collaboration (6–8), and a much broader variety of samples were obtained. The tissue organization and the regulatory mechanisms in animals are far more complicated than those in plants, and thus the number of samples required for the estimation of reliable coexpression will be much larger. In addition, the higher frequency of alternative splicing in mammals hinders a precise evaluation of the strength of gene coexpression.

Our coexpression database, COXPRESdb, was originally constructed in 2007 to use coexpression data for human and mouse, where network representations of gene coexpression data were provided to show the relationships among coexpression modules, in addition to the gene-to-gene relationships (9). In this article, we describe new features of the COXPRESdb. The main improvement is the enhancement of the relevance of coexpression with the limited variety of expression data available in public databases for animals. For this purpose, the evolutionary

*To whom correspondence should be addressed. Tel: +81 22 795 7179; Fax: +81 22 795 7179; Email: kengo@ecei.tohoku.ac.jp

conservation was analyzed, and different data were integrated. The former method is one of the most powerful approaches to evaluate the relevance of gene coexpression (10,11), since if similar coexpression relations are observed for the orthologous genes in different species, then the possibility of experimental/technical artifacts is considered to be reduced. To implement this idea, we first developed a new coexpression measure to directly compare coexpression strengths in different species. Then, five other species, rat, chicken, zebrafish, fly and nematoda, were added. These coexpression data are provided in a comparative gene list representation.

In addition, the protein–protein interaction (PPI) and KEGG pathways in the coexpressed gene networks were added to the interaction networks, because integrating

different layers of information is considered to enhance the reliability of the regulatory relationship. Details of the new features are described in the following sections, along with examples of the comparative coexpressed gene list (Figures 2 and 3) and the integrated gene network (Figure 4). The history of COXPRESdb, as well as other miscellaneous updates, is summarized in Table 1.

COMPARATIVE COEXPRESSED GENE LIST

Gene coexpression for different species

A comparison of the coexpression data for different species is one of the approaches to enhance the relevance of the gene coexpression data. Figure 1 shows a schematic

Table 1. Evolution of COXPRESdb versions over the last 3 years

Version	Date	Human	Mouse	Rat	Chicken	Zebrafish	Fly	Nematoda	Other annotations
ver.4.1	2010-09-12	↑	↑	↑	↑	↑	↑	↑	Updated
ver.4.0	2009-08-25	↑	↑	c2.0	c1.0	c1.0	c1.0	c1.0	Updated
ver.3.3	2008-11-12	↑	↑	↑	–	–	–	–	Updated
ver.3.2	2008-05-29	↑	↑	↑	–	–	–	–	Updated
ver.3.1	2008-08-04	c3.1	c2.1	c1.1	–	–	–	–	Updated
ver.3.0	2008-03-17	c3.0	↑	c1.0	–	–	–	–	Updated
ver.2.0	2007-06-22	c2.0	c2.0	–	–	–	–	–	Updated

↑, same condition as previous version. Details of the updates of ‘Other annotations’ are available at: <http://coexpresdb.jp/versions.shtml>. Date are written as year-month-day format.

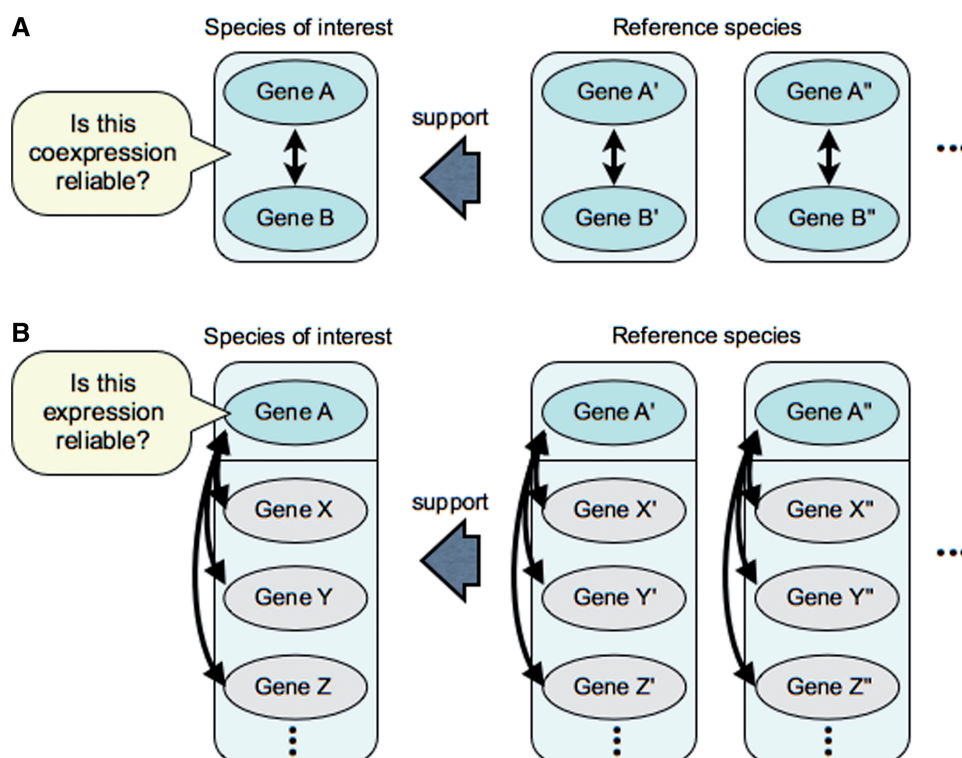


Figure 1. Schematic diagram to evaluate the reliabilities of the coexpression of a pair of genes and the expression data of a guide gene. (A) To investigate the reliability of the coexpression of a pair of genes in a species, a comparison with the coexpression of a pair of orthologous genes in other species can be used. If the coexpression relationship is also observed in other species, then the coexpression in the original species is reliable. (B) To investigate the reliability of the expression pattern of a guide gene, coexpressed gene lists are compared. If most of the coexpression in the coexpressed gene list is observed in other species, then the expression pattern of the guide gene in the original species is reliable. This reliability of the gene expression pattern can be used to investigate that of coexpression, because the coexpression between a pair of genes with a reliable expression pattern is credible.

Table 2. Coexpression data information in COXPRESdb

Species	Abbreviation	Version	Number of genes	Number of experiments	Number of microarray slides	Microarray platform	Last update
Human	Hsa	c3.1	19 777	123	4401	GPL570	2008-04-08
Mouse	Mmu	c2.1	21 036	154	2226	GPL1261	2008-04-08
Rat	Rno	c2.0	13 751	164	3526	GPL1355	2009-08-25
Chicken	Gga	c1.0	13 757	22	352	GPL3213	2009-08-25
Zebrafish	Dre	c1.0	10 112	41	590	GPL1319	2009-08-25
Fly	Dme	c1.0	12 626	52	1022	GPL1322	2009-08-25
Nematoda	Cel	c1.0	17 256	35	514	GPL200	2009-08-25

Species abbreviations in COXPRESdb are used in accordance with those in GenomeNet (<http://www.genome.jp>). Date are written as year-month-day format.

diagram of a comparison of the gene coexpression to evaluate the relevance of the data. When coexpression is observed in a pair of genes in a single species, the reliability of the coexpression may be weak, due to experimental and technical artifacts, such as cross hybridization of probes, secondary-structure formation of the probes to prevent hybridization with the target RNA, and/or inappropriate data treatment. However, if we also find coexpressed gene pairs in orthologs at the same time, then the relevance of the coexpression becomes much higher, because it is less likely that invalid expression data have been obtained multiple times for the related genes. In addition, gene coexpression found in multiple species will be more biologically relevant, because important regulatory mechanisms are conserved during evolution. Actual examples of conserved coexpression are shown in the following section.

Several reports have addressed the effectiveness of conserved coexpression to identify functionally related genes (10,11). These studies focused on orthologous gene clusters represented by the metagene or KOG gene, rather than the gene itself, and coexpression between any pair of gene clusters was calculated using order statistics (10) or 'between-species averaging' (11). These studies clearly showed improved performance in predicting gene functions. However, any gene has the possibility of becoming a key to investigate a phenomenon of interest, regardless of the existence of orthologs. Therefore, instead of making orthologous gene clusters, we simply added the coexpression data of other species to the coexpressed gene list of interest, to confirm the coexpression relationships of a pair of genes and the expression data of a guide gene.

The original version of COXPRESdb already contained coexpression data for two species: human and mouse (9). To enhance this approach, we added five more model species in the current version: rat, chicken, zebrafish, fly and nematoda, with the same calculation protocols as those used for human and mouse (9). The details for the raw data are summarized in Table 2.

Measure of gene coexpression

To evaluate the strength of coexpression, the Pearson and Spearman correlation coefficients are widely used, but we found that these measures were not suitable for direct comparisons among different species and that the correlation rank-based measure, mutual rank (MR, which is

calculated as the geometric mean of the correlation rank of gene A to gene B and of gene B to gene A), is more suitable to compare the coexpression data in multiple species on average (5). In addition, we have performed several successful case studies using MR coexpression data to identify new gene functions in the applications for Arabidopsis (12). Therefore, we adopted MR as the coexpression measure in the COXPRESdb, to compare the coexpression strengths among multiple species.

To further investigate the relevance of gene coexpression between two genes of interest, we also provided a function to check the 'stability of coexpression', which is the degree of change in the coexpression when major microarray samples are subtracted (13), by using the 'coexpression viewer' tool in the 'Draw box'. More stable coexpression suggests stronger functional relationships between the genes.

An example of a comparative coexpressed gene table

The table on the left in Figure 2 shows the human coexpressed gene list from SLC39A7. The top line is the guide gene itself, and the following lines indicate the strongly coexpressed genes ordered by their coexpression strength, as measured by MR (smaller MR indicates stronger coexpression), while the five columns on the right indicate the gene coexpression strengths in other species. Note that the chicken ortholog was not available for this guide gene, and thus it was omitted from this table. On the second line, the mouse coexpression column (Mmu MR column in the table) shows the coexpression between the mouse orthologs to human SLC39A7 and that to COPG. Its coexpression strength is 38.7, which is strong enough, because the MR value distribution is 1 for the total number of genes in the species for the coexpression analysis (about 20 000 for human, mouse and rat). We previously showed that MR values less than 5000 represent significant coexpression (12), but in view of this observation, weak coexpression was judged by an MR >1000 and is shown in a faint color. The coexpressed genes from human SLC39A7 are fairly well conserved in other species, indicating that most of the coexpression in this list as well as the expression pattern of SLC39A7 are reliable. Note that in some cases, such as the 4th line for the PPIB gene in *Caenorhabditis elegans* (Cel MR column), there are multiple orthologous genes in a species, and the coexpression values are shown in parallel in a single cell (8.9 and 17.3).

	Hsa MR	Gene	Mmu MR for Slc39a7 [list]	Rno MR for Slc39a7 [list]	Dre MR for 30094 [list]	Dme MR for 18805 [list]	Cel MR for 181640 [list]
0	0.0	SLC39A7	0.0	0.0	0.0	0.0	0.0
1	3.9	COPG	38.7	125.8			
2	9.8	RPN1	11.2	22.4	68.6		3547.2
3	10.6	PPIB	529.6	13.6	24.4	861.1	8.9 17.3
4	11.5	ERGIC1	886.9	844.3			191.3
5	11.8	YIPF3	1322.0	1038.0	1047.5		11640.2
6	12.0	ST13A	75.3	1300.7	49.2	1909.6	
7	26.7	YIF1A	480.8		623.3		391.9
8	31.4	PFDN6	2028.3	3948.7	653.6	4657.3	16267.6
9	32.9	GMPPB	32.7	127.2	1191.9	214.8	1204.9
10	32.9	HM13	360.0	4.0	139.3	32.6	3836.4
11	33.0	C19orf10	157.9	65.9	8.9		
12	33.0	PDIA4	389.9	30.6	280.1		26.0
13	34.8	TMED3	433.8	36.5	349.0		
14	41.8	TMED9	131.2	52.8	667.1		
15	42.9	GMPPA	187.6	252.1	118.7 218.6	102.0	11689.1
16	46.1	ALDH18A1	257.5	641.1	3249.2	2132.9	23.2
17	47.4	SLC35B2	35.8	6.6	5502.3	5421.3	4513.9
18	48.4	PDIA3	223.6	25.0		1.7	952.6
19	54.2	DAP	4439.1	159.2	267.2		
20	55.3	MOGS	1003	17.7		5715.9	15829.4
21	58.5	SRPRB	369.4	79.7	849.4	627.2	2891.8
22	61.2	HYOU1	1100.6	880.1	939.8	157.0	9270.5 469.6
23	63.3	SEC13	251.5	354.6	122.7	272.9	2417.5
24	64.1	YIPF2	7729.2	9099.4			
25	65.0	LMAN2	244.1		2143.3	86.7	140.5

	Hsa MR	Gene	Mmu MR for Yipf3 [list]	Rno MR for Yipf3 [list]
0	0.0	YIPF3	0.0	0.0
1	1.0	SLC35B2	387.4	937.0
2	3.0	NEU1	1499.7	81.6
3	4.2	CUL7	2481.8	4045.4
4	5.2	MEA1	1690.8	2115.4
5	6.3	TMEM115	47.2	405.2
6	6.7	CNPY3	741.4	718.0
7	11.8	SLC39A7	1322.0	1038.0
8	13.2	CHPF2		55.9
9	14.5	GLB1	742.3	3237.5
10	14.7	MRPL2	1084.5	2889.5
11	16.4	TM9SF1	3235.4	2998.7
12	21.2	CUTA	30.2	1580.8
13	21.2	TBCC	353.2	1273.7
14	21.9	WDR45	101.6	
15	23.4	BAT3	570.3	256.0
16	25.2	FAM127B		
17	25.4	YIF1A	364.8	
18	25.9	NCSTN	3672.2	18.6
19	26.1	MOGS	1698.4	832.6
20	28.8	C16orf58	791.5	713.6
21	31.0	GPR108	2075.1	203.8
22	32.1	SDF4	2065.2	1056.2
23	32.6	ATP13A1	15.9	823.8
24	34.4	FAM125A	8.1	258.3
25	34.6	COPE	235.4	248.4
26	35.6	DAP	1593.0	526.3
27	36.2	STK19	52.5	852.9
28	37.8	C1orf85	123.4	528.4
29	40.0	C6orf89	23.1	536.1

	Hsa MR	Gene	Mmu MR for Yipf2 [list]	Rno MR for Yipf2 [list]
0	0.0	YIPF2	0.0	0.0
1	5.5	CHPF	9974.4	11760.8
2	8.9	TMED1	3940.1	5898.6
3	11.9	CARM1	3016.9	12982.2
4	18.5	PRKCSH	9046.2	6543.2
5	19.6	HM13	602.2	5865.5
6	25.4	BET1L	1506.9	5208.8
7	28.5	ADAM15	9676.6	6760.3
8	29.9	ATG4D	4117.1	
9	40.6	NR2F6	319.6	10229.9
10	40.6	TMUB1	8949.4	1644.3
11	50.0	PTPRF	7887.2	11696.8
12	51.8	ZNF768	2410.9	
13	55.3	C16orf58	459.8	2716.5
14	58.3	GIPC1	2323.5	13009.0
15	61.5	KDELRL1	462.5	10812.2
16	62.5	TMEM161A	1689.6	5143.6
17	64.1	SLC39A7	7729.2	9099.4
18	65.5	FBXW9	8219.0	975.0
19	68.9	DNASE2	857.3	6477.5
20	69.3	COPE	1226.8	3433.8
21	69.6	CHPF2		5546.0
22	72.5	B4GALT7	925.1	5597.1
23	73.0	MXD4	524.9	1704.4
24	74.9	RHOD	4210.5	3818.4
25	76.0	TMEM214	2032.1	494.6
26	78.0	CC2D1A	1857.4	5358.3
27	79.4	AP1M2	7157.4	
28	82.4	CUL7	16501.9	8360.7
29	85.8	PYCR1	10642.4	11236.2

Figure 2. An example of coexpression evaluation by coexpression comparison. In the table on the left, the coexpressed gene list from human SLC39A7 is shown with coexpression in other species. Note that this gene lacks the chicken ortholog in the microarray data. In the center table, the coexpressed gene list from human YIPF3 is shown. For clarity, only mouse and rat coexpressions are shown for the coexpression comparison. The table on the right is almost the same as the center table, except for the guide gene, YIPF2. Hsa, Mmu, Rno, Dre, Dme and Cel indicate human, mouse, rat, zebrafish, fly and nematoda, respectively. These species abbreviations are also used in Table 2. This page can be obtained as follows. First, search ‘SLC39A7’ using the search box on the top page of COXPRESdb. Then click the link of ‘coexpressed gene list’.

In this comparative gene list, two genes were found to be unsupported by other species. One is the fifth strongest coexpressed gene, YIPF3; and the other is the 24th strongest coexpressed gene, YIPF2, which are marked with red circles in Figure 2. There are two possible explanations for the unsupported coexpression: (i) these coexpressions with the guide gene are human specific, and (ii) the expression patterns of these two genes are affected by some technical problems. To assess these two possibilities, coexpressed gene lists from these two genes can be used. As shown in the center table of Figure 2, the coexpressed genes from YIPF3 were mostly conserved, while those from YIPF2 were not supported by mouse and rat (right table of Figure 2), indicating that the expression pattern of YIPF2 was less reliable than that of YIPF3, and suggesting that the coexpression between SLC38A7 and YIPF3 in the table on the left is human specific.

In addition to evaluating the reliabilities of coexpression between a pair of genes and the expression pattern of the guide gene, a coexpression comparison can be used to identify ‘functional orthologs’ from traditional sequence orthologs. Figure 3 shows a comparative coexpressed gene list from a mouse guide gene. For this mouse guide gene,

there are two human orthologs. Both of the human genes show strong coexpression with the orthologous genes of the coexpressed genes in mouse. However, for almost all of the coexpressed gene relationships, the gene on the left side (PRSS1) shows stronger coexpression, suggesting that PRSS1 is more likely to be a functional ortholog, rather than PRSS3, which might have acquired some other cellular functions after gene duplication.

INTEGRATING DIFFERENT INFORMATION INTO THE GENE COEXPRESSION NETWORK

In addition to the conserved coexpression, two types of omics data were mapped to the gene coexpression network: the known pathways and the PPIs.

The known pathway information on gene coexpression networks is shown on the nodes, which are marked with the KEGG pathway annotation (14) (Figure 4). For each gene network, up to five KEGG pathways with larger numbers of genes within the pathway are selected. For example, the gene network for human CD3D has eight genes for ‘T cell receptor signaling pathway’ (red circles in Figure 4). The same marks in the table, placed just

	Mmu MR	Gene	Hsa MR for PRSS1 [list]	Hsa MR for PRSS3 [list]
0	0.0	Prss2	0.0 64.4	64.4 0.0
1	1.0	2210010C04Rik		
2	1.4	Cel	24.7	149.1
3	2.2	Ctrb1	9.5	100.2
4	3.0	Cpa1	6.7	89.0
5	3.0	Cela2a	8.7	92.5
6	3.5	Ctrl	16.0	532.6
7	3.5	Cela3b	10.2	93.0
8	3.5	Cpb1	52.5	11184.3
9	4.0	Cips	11.6	109.2
10	4.6	11723		
11	4.7	Reg1	214.8	40.0
12	6.2	Gp2	44.0	1583.3
13	6.2	Pnliprp1	15.9	403.0
14	7.1	Rnase1	3890.9	11192.1
15	7.5	Iapp	37.1	2413.2
16	7.8	Zg16	13672.9	112.2
17	8.4	Sycn	12.7	297.7
18	8.5	Pdia2	377.8	3251.9
19	8.8	Ins2	12.4	155.4
20	9.2	Cpa2	17.5	537.0
21	9.8	1810028F09Rik		
22	9.8	Serpini2	16.9	659.7
23	9.8	Pnlip	5.3	84.2
24	11.5	1810009J06Rik		
25	11.8	Cuzd1	17.0	154.6
26	13.1	Ins1		
27	14.5	1810073G21Rik		
28	15.7	Reg2	112.9	34.6
29	18.8	Pnliprp2	24.2	46.7

Figure 3. Coexpression comparison to investigate the functional orthologs, using the sequence orthologs. There are two human orthologs for the mouse Prss2 gene. Although both genes show similar coexpression with mouse Prss2, human PRSS1 shows stronger coexpression than PRSS3, suggesting that PRSS1 is the functional ortholog of mouse Prss2, while PRSS3 adopted another function in the human cell.

below the network, represent the links to the KEGG pathway, in which the marked genes in the coexpressed gene networks are highlighted by red boxes. This allows the coexpressed gene modules to be associated with the metabolic pathways. In this case, the KEGG pathway shows that the genes coexpressed with CD3D include the genes for the T cell receptor complex and the surrounding signaling proteins.

PPI information is also useful to predict the functions of genes, because gene coexpression and PPI reflect two different regulatory layers, mRNA and protein, respectively. In the previous COXPRESdb, we used HPRD data for the human PPI network, but to show the PPI information

in other species, we also used IntAct data for PPI in the current version of COXPRESdb. Number of networks with PPI and with KEGG annotation is summarized in Table 3.

In addition to the pre-calculated gene networks for each gene, each GO functional group and each tissue, we also provide a 'NetworkDrawer' tool, to draw the gene network for the user-defined set of genes. This tool detects genes displaying coexpression and/or PPI with the query genes, according to the user-defined criteria. The result networks are depicted in the Pajak (15) and Cytoscape (16) formats, in addition to the default PNG figure.

OTHER IMPROVEMENTS

New top page design

The top page of COXPRESdb was revised to intuitively show the four major functions of our database. COXPRESdb has eleven tools, in addition to the pre-calculated pages. To easily access these tools, we classified them into two categories, 'Search box' and 'Draw box'. The Search box is composed of tools to search genes by various aspects, such as coexpression and PPI, and it outputs a gene table. On the other hand, the tools in Draw box can draw various pictures, such as a gene network, a hierarchical gene cluster and a detailed view of gene coexpression. There are two additional boxes, 'Browse' and 'Bulk download'. The Browse box has a link to the gene networks for 49 tissues. These gene networks are huge, and therefore, these networks are provided as a Google Map interface. The Bulk download box is used for downloading coexpression data and protein subcellular localization predicted by WoLF PSORT (17). These data are now available under the Creative Commons Attribution license.

A simple gene search is available at the search window just below the scenic picture. The user can search genes and GO terms using keywords, gene aliases, Entrez gene IDs or GO IDs. The default searches for all of them. Users can access the coexpressed gene list and network for any single guide gene in human, mouse and rat.

Version numbering

COXPRESdb is composed of our own coexpression data and integrated public annotations. The coexpression data are updated yearly, and the public annotations are updated every few months (Table 1). Each data update is assigned a new version number, to prevent confusion about the data update and as a reference for users' publications. The major version number essentially corresponds to the coexpression data update, whereas the minor version number corresponds to the public data update. Version histories can be checked at: <http://coxpresdb.jp/versions.shtml>. Note that we maintain previous versions of the gene coexpression data, but we do not store any other annotations.

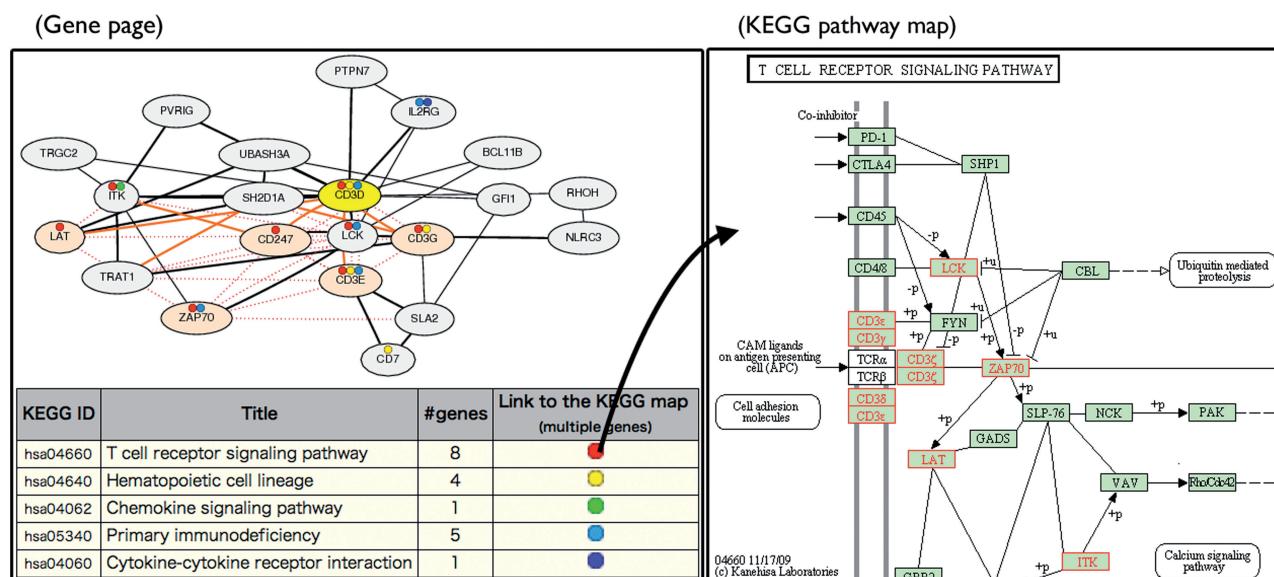


Figure 4. Screenshots using the KEGG pathway annotation on a coexpressed gene network. An example of a coexpressed gene network with KEGG pathway annotations. Common KEGG pathways in the network are denoted by the color-coded (red, yellow, green, blue and purple) dots in the nodes. This page can be obtained at: <http://coexpresdb.jp/data/locus/915.shtml>.

Table 3. Statistics of gene networks provided in COXPRESdb version 4.1

Type of page	Species	Total	With common KEGG annotation	With PPI
Locus page	Human	19 777	19 204	3584
Locus page	Mouse	21 036	20 363	232
Locus page	Rat	13 751	13 624	66
GO term	Human	1862	708	569
GO term	Mouse	1530	711	42
GO term	Rat	1927	592	17

ACKNOWLEDGEMENTS

The super-computing resource was provided by Human Genome Center, Institute of Medical Science, The University of Tokyo.

FUNDING

Grant-in-Aid for Scientific Research (No. 50397048, to T.O.); Grant-in-Aid for Innovative Areas “HD physiology” (No. 22136005, to K.K.). Funding for open access charge: MEXT of Japan.

Conflict of interest statement. None declared.

REFERENCES

- Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
- Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhäuser, D., Persson, S. and Provat, N.J. (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.*, **32**, 1633–1651.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M. and Edgar, R. (2007) NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acids Res.*, **35**, D760–D765.
- Parkinson, H., Kapushesky, M., Kolesnikov, N., Rustici, G., Shojatalab, M., Abeygunawardena, N., Berube, H., Dylag, M., Emam, I., Farne, A. *et al.* (2009) ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.*, **37**, D868–D872.
- Obayashi, T. and Kinoshita, K. (2009) Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Res.*, **16**, 249–260.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D. and Lohmann, J.U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.*, **37**, 501–506.
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D’Angelo, C., Bornberg-Bauer, E., Kudla, J. and Harter, K. (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.*, **50**, 347–363.
- Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., Ogawa, M., Yamauchi, Y., Preston, J., Aoki, K. *et al.* (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.*, **55**, 526–542.
- Obayashi, T., Hayashi, S., Shibaoka, M., Saeki, M., Ohta, H. and Kinoshita, K. (2008) COXPRESdb: a database of coexpressed gene networks in mammals. *Nucleic Acids Res.*, **36**, D77–D82.
- Stuart, J.M., Segal, E., Koller, D. and Kim, S.K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.
- Oti, M., van Reeuwijk, J., Huynen, M.A. and Brunner, H.G. (2008) Conserved co-expression for candidate disease gene prioritization. *BMC Bioinformatics*, **9**, 208.
- Obayashi, T. and Kinoshita, K. (2010) Coexpression landscape in ATTED-II: usage of gene list and gene network for various types of pathways. *J. Plant Res.*, **123**, 311–319.
- Kinoshita, K. and Obayashi, T. (2009) Multi-dimensional correlations for gene coexpression and application to the large-scale data of Arabidopsis. *Bioinformatics*, **25**, 2677–2684.
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2010) KEGG for representation and analysis of

- molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
15. Batagelj, V. and Mrvar, A. (1998) Pajek program for large network analysis. *Connections*, **21**, 47–57.
16. Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
17. Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, **35**, W585–W587.