Supplementary Materials for

A Modular Network Model of Aging

Huiling Xue[1, 2#], Bo Xian[1,2#], Dong Dong[1], Kai Xia[1], Shanshan Zhu[1], Zhongnan Zhang[1], Lei Hou[1], Qingpeng Zhang[1], Yi Zhang[1] and Jing-Dong J. Han[1]*

[1] Chinese Academy of Sciences Key Laboratory of Molecular and Developmental Biology, Center for Molecular Systems Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Road, Beijing, 100101, China.

[2] Graduate School, Chinese Academy of Sciences, Beijing, China 100039

*To whom correspondence should be addressed. Email: jdhan@genetics.ac.cn

[#]These authors contributed equally to this work.

This PDF file includes:

Supplementary Note 1. 'NP analysis' method

Briefly, this analysis method includes the following steps: 1) obtain all the PPIs (Protein-Protein Interactions) between genes that have either similar expression profiles (correlated interactions) or opposite expression profiles (anti-correlated interactions) to arrive at the network of Negatively and Positively correlated interactions (NP network); 2) identify network modules so that the expression profiles of genes within a module are similar, correlated interactions are maximally enclosed within a module and anti-correlated interactions are optimally distributed between modules. The second step is approximated by first applying hierarchical clustering to the genes in the NP network, then dissecting the largest uniform clusters and anti-correlated clusters using the ratio of negative to positively correlated interaction numbers. Algorithm details are available in (Xia et al., 2006b).

Compared to conventional expression profiler clustering, the NP analysis incorporates additional biological information from PPI networks. It does not require pre-filtering the genes based on expression intensity of fold-change, which is often biased against low-level expressed regulatory genes. Instead, it is solely based on the shape of change of expression profiles between genes/proteins that potentially interact. Due to the transitive property of expression profiles, when anti-correlated interactions are included in the NP network and used to delimit the cluster boundaries, they promote the partition of anti-correlated clusters and increase the homogeneity of all expression clusters. The enriched regulatory nodes (proteins) mediating inter-module PPIs in a NP network indicates a unique advantage of this analysis method in finding regulatory nodes, edges and circuits in the cellular network (Xia et al., 2006b).

Supplementary Note 2. Getting similar modules using other interactome data

Yeast two hybrid information is unreliable when used alone, integrating with other 'omics' data can however reveal true biological information (Gunsalus et al., 2005). To confirm the biological relevance of the network modules found in the fruit fly aging network, we performed the same analysis using the subset of high-confidence Y2H dataset as defined in the original studies generated the datasets or another dataset of PPIs predicted using a probabilistic model (Xia et al., 2006a). The first dataset gives rise to the same modules except for smaller module sizes. The second dataset give rise to clear P, R and O modules, but a very small D module due to the species-specificity of the D modules. Relationships among the modules are also preserved by these other two datasets (Supplementary Figure 2 and Supplementary Table III). This indicates the identification of P, D, O and R modules are not due to the false positives in the Y2H dataset.

Supplementary Note 3. Expression of orthologous genes of fly modules in human and that of human modules in fly

To examine whether the different gene compositions of D modules between human brain and fly and the additional R-O partitions in fly are due to different coverage of the interactome or transcriptome datasets for the two species, or due to different regulation modes between fly whole body and neurons, or reflect different regulatory modes in the two species, we first examined if the P-D and R-O anti-correlations can be observed using homologous genes. The results indicate that the anti-correlations are not conserved among the homologous genes (Supplementary Figure 3A). We further examined the conservation of aging-related changes of human brain and fly modules across the two species and among various tissues, including human brain, muscle, skin, whole fly and fly heads, which consist mostly of neurons. We found 1) the age-related gene expression increase of P homologs is conserved across species and tissues; 2) the age-related decrease of D homologs is conserved among tissues within a species but different between the two species; 3) the

age-related decrease of O is observed in whole fly, fly heads and human muscle, but age-related increase of R is observed in whole fly, fly heads and human brain (Supplementary Figure 3B, Supplementary Table IV). Assuming the homologs determined by our method are conserved in their molecular functions between the two species, if the differences in D module gene composition and the lack of R-O in human are due to the different coverage of the interactome or transcriptome datasets, the homologous genes that are differentially covered by the datasets should display the same anti-correlation and age-related changes when cross-examined using the fly or human expression datasets, which is not the case here (Supplementary Figure 3A and 3B). Altogether, these suggest the lack of O and R module or the lack of overlap of D modules between fly and human might not be due to different interactome and transcriptome data coverages.

Supplementary Note 4. Enrichment of cell cycle commitment genes in human and fruitfly P modules

Consistent with their roles in cell-autonomous proliferation process, both the human and fly P modules have the highest percentage of G1/S and G2/M genes among all modules (enrichment P=0.08 for human brain P module, 3.65×10^{-4} and 3.51×10^{-4} for fly P modules under normal or CR condition, respectively, Supplementary Figure 6). Although we have found that 1) P module genes are enriched in proliferation-related GO terms 2) at cellular level, its expression switch from high to low expression upon induction of cellular proliferation to differentiation switch (Xia et al., 2006b). Enrichment in genes that assume high expression at G1/S and G2/M cell cycle phase is independent evidence that the P module is related to the cellular proliferation process.

Supplementary Note 5. Statistically evaluate the chance of getting anti-correlated modules corresponding to reductive and oxidative phase respectively

We generated 100 artificially constructed module pairs (gene set pairs) of the same number of genes as in the R and O modules, respectively, by randomly selecting fruit fly genes in the NP or Y2H network. We then counted the number of times when a pair of modules displayed transcriptional anti-correlation during the yeast metabolic cycle based on the expression profiles of their yeast orthologs. None was found to display transcriptional anti-correlations that are equal to or less than that between R and O modules (e.g. PCC=-0.58 for the normal and CR module overlaps, empirical $P<0.01$, assuming normal distribution, $P=2.6\times10^{-4}$ and 8×10^{-5} when modules are constructed from random nodes in the NP and Y2H network, respectively). This demonstrates that it is unlikely to observe a concerted expression changes such as that between the R and O modules during the metabolic cycle among randomly constructed modules. We also randomly selected 20 pairs for visual examination. None of the pairs display alternative high expression at the oxidative and reductive phases, respectively (data not shown).

Supplementary Note 6. Literature annotation of the genes that extend worm lifespan upon RNAi

Pak3 positively regulates Raf-1 activity (King et al., 1998) and is associated with nonsyndromic X-linked mental retardation (Allen et al., 1998). TCEB3 is a subunit of the Elongin (SIII) complex that activates elongation by mammalian RNA polymerase II (Aso et al., 1995). CDC20 is a cell cycle check point protein. Its up-regulation has been associated with tumorigenesis and poor prognosis (Ouellet et al., 2006). While genome-wide RNAi worm lifespan screens usually identify genes that extend the maximal lifespan, genes found through our

computational prediction mostly extend the mean or average lifespan of the worms, which are actually the most desirable results of anti-aging agents.

Supplementary Note 7. Literature annotation of the genes that shorten worm lifespan upon RNAi

We also found many genes that shorten worm lifespan upon RNAi. Because reduced lifespan can be caused by diseases that unnecessarily affect aging *per se*, large-scale RNAi screens generally ignore these genes. However, a literature search reveals that most of these genes we found already have evidences to support their roles in aging. For example, although MAPK1 is also required for vulva development, and the early death of MAPK1 RNAi worms is due to a vulva-less phenotype, impaired MAPK1 signaling might mediate SIRT1 inhibition induced human cell senescence (Ota et al., 2006). Studies have also indicated that MAPK1 signaling pathway is impaired in the aged mouse brain and that these impairments can be modulated by lifelong caloric restriction (Zhen et al., 1999). RNAi of the rest of these genes does not cause obvious developmental defects. The activity and fidelity of POLA decline in aged mice and the age-related decrease of POLA activity can be delayed by the caloric restriction (Srivastava and Busbee, 2002). Gain of function mutations of worm gsa-1 (G protein alpha subunit, homolog of human GNAS) can induce the neural degeneration through necrotic cell death (Korswagen et al., 1997). Sp3 is an oxidative stress-inducible, anti-death transcription factor in cortical neurons and is associated with neurodegenerative diseases, such as Huntington's disease (Ryu et al., 2003). Deactivation of TBP (TATA Binding Protein) by polyQ aggregation has been shown to contribute to neural degenerative diseases including Huntington's disease (Schaffar et al., 2004).

Supplementary Figure Legends

Supplementary Figure 1. The protein-protein interaction (PPI) within and among the fruitfly modules under normal and calorie restriction (CR) conditions (**A**) Under normal condition, the network forms a bipartite structure with correlated interactions (Red edges, representing PPIS between two genes having similar (PCC>0.4) expression profiles during aging) connecting between P and R as well as between D and O modules. The two partitions in the network are connected by anti-correlated interactions (Green edges, representing PPIs between two genes having opposite (PCC<-0.4) expression profiles during aging).

(**B**) Under diet/calorie restricted condition (CR), the modules are connected by mostly anti-correlated interactions, whereas those in between other module pairs are a mixture of correlated or anti-correlated interactions. Fruitfly D, P, O and R modules are represented with nodes of lavender, green, orange, light green color, respectively. The nodes in the NP network are grouped by the identified network modules to visualize the PPIs within and in between modules.

Supplementary Figure 2. Transcriptional relationships among the modules under normal or diet restricted (CR) conditions based on an independent PPI dataset.

Different from Figure 1B and C, the average expression values are based on modules derived from a Bayesian model-predicted PPI network deposited in the 'IntNetDB' database (Xia et al., 2006a). The significances of the overlaps of these modules to the modules found in Y2H PPI network are listed in Supplementary Table III.

Supplementary Figure 3. Expression of orthologous genes of fly modules in human and that of human modules in fly

(A) Average expression levels of human brain modules gene homologs in fly and average expression levels of fly modules gene homologs in human brain.

The fly homologs of human module genes and the human homologs of the fly module genes were determined by best reciprocal BLASTP hits with e-value < 10^{-6} and used to plot the average expression levels.

(B) The conservation of aging-related changes among different tissues and between human and fly. The expression levels of genes in the human brain P, D and fly P, D, R and O modules under high or low food conditions (listed in column headers) are compared between young and old samples of human brain, muscle, skin, whole fly and fly head (listed in row headers) by paired Student *t*-test. Red color indicates an increase in the old samples; green indicates a decrease in the old samples. The color intensity represents the –log(*P*-value) between the old and young samples. Old and young samples were categorized as in Supplementary Table IV.

Supplementary Figure 4. The average expression intensities of the P, D, R and O modules under normal and calorie restriction (CR) conditions

The expression levels of the genes specific for each conditions are plotted against age for each modules. Normal Specific and CR Specific represent the non-overlapping genes specific under normal or CR condition.

Supplementary Figure 5. Overlaps between the human brain and fly modules Fly modules include those under normal diet or caloric restricted diet. Color intensity in each field denotes the significance of the overlap, proportional to the –log values of the Fisher exact test *P*-values as indicated by the color legend. The numbers in parentheses are the genes in each module in the same search space between human and fly, that is, in the intersection of the fly and human interactomes and transcriptomes based on the orthologs between the two species. The numbers inside the matrix are the overlapping genes between the intersecting modules from the row and the column.

Supplementary Figure 6. The percentage of G1/S and G2/M cell cycle genes in the human and fly modules under normal and CR conditions

The averages are based on all the genes in the intersection of transcriptome and interactome probed for human or fly (lavender bar).

Supplementary Figure 7. Node-betweeness ordered node attacks.

(**A**) Attacking the aging genes in the NP network increases the characteristic path length (CPL) of the network more rapidly than attacking randomly selected non-aging genes or removal of random nodes ('failure'), but slower than sequential removal of nodes of highest betweeness in the order of their betweeness values. Betweeness-matched attacks on aging and non-aging genes in the NP network are shown in the inset.

(**B**) Betweeness-ordered attacks on aging genes belonging to the NP network increase CPL of the HPRD network more rapidly than attacking aging genes not in the NP network or random removal of genes in the HPRD network. Betweeness-matched attacks on NP and non-NP aging genes in the HPRD network are shown in the inset.

(**C**) Betweeness-ordered attacks on the aging (or all) genes on the module interfaces increases CPL of the NP network more rapidly than attacking their counterparts in the cores. Betweeness-matched attacks on the core and interface genes are shown in the inset. Only the first 2% of the attacks are shown in the inset for the interface and core genes, the trend continues for the rest.

Supplementary Figure 8. The interface genes in fly NP network are more important to network topology than the core genes. Attacking the interface genes increase the CPL of the NP network more rapidly than attacking the core genes under either normal (**A**) or CR (**B**) conditions. Similar results were obtained by betweeness-ordered attacks on the interface and core genes under either normal (**C**) or CR (**D**) conditions.

Supplementary Table I. Gene list of each fly module under normal or diet-restricted (CR) condition

Provided in a separate file “Sup_2.xls”.

Supplementary Table II. GO terms enriched in each of the fly module

Provided as a separate file “Sup_3.xls”.

Supplementary Table III. The significances of overlap between the IntNetDB modules and Y2H modules

The overlap significance between modules derived from *Drosophila* interologs of a Bayesian-model predicted human PPI network (IntNetDB modules, listed as the row headers) and the modules found in Y2H PPI network (Y2H modules, listed as the column headers) are evaluated by Fisher exact test and listed in the crossing cells.

	Module	**D (557)**	**P (618)**	**O (569)**	**R (389)**
Normal	D (124)	**0.0142209**	1	0.9999190	1
	P (162)	1	**1.16E-31**	1	0.5579290
	O (411)	0.9999994	0.9999993	**2.27E-53**	1
	R (106)	1	0.9287505	1	**2.08E-31**
	Module	**D (747)**	**P (764)**	**O (416)**	**R (370)**
CR	D (127)	**0.0314082**	1	0.9683391	1
	P (272)	0.9999994	**1.27E-45**	1	0.9968089
	O (341)	0.9999993	0.9298167	**1.16E-52**	1
	R (123)	1	1	1	**1.26E-33**

Supplementary Table IV. The young and old samples used to determine the aging-related changes in expression levels of the gene modules, as shown in Supplementary Figure 3B.

	Age of young group	Age of old group
human brain	>26 yr and <40 yr	>=40 yr and <=90 yr
human male muscle	21-27 yr	69-75 yr
human female muscle	20-29 yr	20-29 yr
human fibroblast	22, 22 and 20 yr	87, 89 and 89 yr
fly normal condition	7-18 days	42-47 days
fly head	3 days	47 days

Supplementary Table V. The proteins at the human brain module interfaces have significantly higher percentage of known human 'aging genes' and transcriptional regulators than those inside the cores

	P-value				Interface percentage (aging gene/ regulatory gene)				Core percentage (aging gene/ regulatory gene)			
Module	**D**	**P**	**PP**	**I**	**D**	**P**	**PP**	**I**	**D**	**P**	**PP**	**I**
D		1.00E-04	0	0.09		12%	16%	14%		5%	6%	8%
P	0		0.01	0.01	27%		9%	13%	13%		4%	5%
PP	0.001	0.03		0.70	26%	17%		5%	16%	11%		8%
I	0.71	0.55	0.005		24%	18%	32%		21%	15%	13%	

The statistical significance of the difference between the values for the interface and those inside the two modules bridged by the interface is evaluated by the Student *t*-test.

The percentage of known 'aging genes' inside the human modules or at the module interfaces are shown at the upper right half of the matrices, while those for the transcription regulatory genes are at the lower left half.

References

Allen, K.M., Gleeson, J.G., Bagrodia, S., Partington, M.W., MacMillan, J.C., Cerione, R.A., Mulley, J.C. and Walsh, C.A. (1998) PAK3 mutation in nonsyndromic X-linked mental retardation. *Nat Genet*, **20**, 25-30.

Aso, T., Lane, W.S., Conaway, J.W. and Conaway, R.C. (1995) Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science*, **269**, 1439-1443.

Gunsalus, K.C., Ge, H., Schetter, A.J., Goldberg, D.S., Han, J.D., Hao, T., Berriz, G.F., Bertin, N., Huang, J., Chuang, L.S., Li, N., Mani, R., Hyman, A.A., Sonnichsen, B., Echeverri, C.J., Roth, F.P., Vidal, M. and Piano, F. (2005) Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. *Nature*, **436**, 861-865.

King, A.J., Sun, H., Diaz, B., Barnard, D., Miao, W., Bagrodia, S. and Marshall, M.S. (1998) The protein kinase Pak3 positively regulates Raf-1 activity through phosphorylation of serine 338. *Nature*, **396**, 180-183.

Korswagen, H.C., Park, J.H., Ohshima, Y. and Plasterk, R.H. (1997) An activating mutation in a Caenorhabditis elegans Gs protein induces neural degeneration. *Genes Dev*, **11**, 1493-1503.

Ota, H., Tokunaga, E., Chang, K., Hikasa, M., Iijima, K., Eto, M., Kozaki, K., Akishita, M., Ouchi, Y. and Kaneki, M. (2006) Sirt1 inhibitor, Sirtinol, induces senescence-like growth arrest with attenuated Ras-MAPK signaling in human cancer cells. *Oncogene*, **25**, 176-185.

Ouellet, V., Guyot, M.C., Le Page, C., Filali-Mouhim, A., Lussier, C., Tonin, P.N., Provencher, D.M. and Mes-Masson, A.M. (2006) Tissue array analysis of expression microarray candidates identifies markers associated with tumor grade and outcome in serous epithelial ovarian cancer. *Int J Cancer*, **119**, 599-607.

Ryu, H., Lee, J., Zaman, K., Kubilis, J., Ferrante, R.J., Ross, B.D., Neve, R. and Ratan, R.R. (2003) Sp1 and Sp3 are oxidative stress-inducible, antideath transcription factors in cortical neurons. *J Neurosci*, **23**, 3597-3606.

Schaffar, G., Breuer, P., Boteva, R., Behrends, C., Tzvetkov, N., Strippel, N., Sakahira, H., Siegers, K., Hayer-Hartl, M. and Hartl, F.U. (2004) Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation. *Mol Cell*, **15**, 95-105.

Srivastava, V.K. and Busbee, D.L. (2002) Replicative enzymes and ageing: importance of DNA polymerase alpha function to the events of cellular ageing. *Ageing Res Rev*, **1**, 443-463.

Xia, K., Dong, D. and Han, J.D. (2006a) IntNetDB v1.0: An integrated protein-protein interaction network database generated by a probabilistic model. *BMC Bioinformatics*, **7**, 508.

Xia, K., Dong, D., Xue, H., Zhu, S., Wang, J., Zhang, Q., Hou, L., Chen, H., Tao, R., Huang, Z., Fu, Z., Chen, Y.G. and Han, J.D. (2006b) Identification of the Proliferation/Differentiation Switch in the Cellular Network of Multicellular Organisms. *PLoS Computational Biology*, **2**, e145.

Zhen, X., Uryu, K., Cai, G., Johnson, G.P. and Friedman, E. (1999) Age-associated impairment in brain MAPK signal pathways and the effect of caloric restriction in Fischer 344 rats. *J Gerontol A Biol Sci Med Sci*, **54**, B539-548.