

Identification of the Proliferation/Differentiation Switch in the Cellular Network of Multicellular Organisms

Kai Xia¹, Huiling Xue¹, Dong Dong¹, Shanshan Zhu¹, Jiamu Wang¹, Qingpeng Zhang¹, Lei Hou¹, Hua Chen², Ran Tao², Zheng Huang¹, Zheng Fu¹, Ye-Guang Chen², Jing-Dong J. Han^{1*}

1 The Chinese Academy of Sciences Key Laboratory of Developmental Biology, Center for Molecular Systems Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, People's Republic of China, **2** The State Key Laboratory of Biomembrane and Membrane Biotechnology, Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing, People's Republic of China

The protein-protein interaction networks, or interactome networks, have been shown to have dynamic modular structures, yet the functional connections between and among the modules are less well understood. Here, using a new pipeline to integrate the interactome and the transcriptome, we identified a pair of transcriptionally anticorrelated modules, each consisting of hundreds of genes in multicellular interactome networks across different individuals and populations. The two modules are associated with cellular proliferation and differentiation, respectively. The proliferation module is conserved among eukaryotic organisms, whereas the differentiation module is specific to multicellular organisms. Upon differentiation of various tissues and cell lines from different organisms, the expression of the proliferation module is more uniformly suppressed, while the differentiation module is upregulated in a tissue- and species-specific manner. Our results indicate that even at the tissue and organism levels, proliferation and differentiation modules may correspond to two alternative states of the molecular network and may reflect a universal symbiotic relationship in a multicellular organism. Our analyses further predict that the proteins mediating the interactions between these modules may serve as modulators at the proliferation/differentiation switch.

Citation: Xia K, Xue H, Dong D, Zhu S, Wang J, et al. (2006) Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms. *PLoS Comput Biol* 2(11): e145. doi:10.1371/journal.pcbi.0020145

Introduction

How cells coordinate proliferation and differentiation has been one of the most important questions in developmental biology, cell biology, and cancer biology. The idea that growth and proliferation are poorly compatible with differentiation has wide currency, and explicit proliferation/differentiation switches have been demonstrated for many different cell types [1–4], but no general mechanism has been apparent. Due to the multifactor nature of this coordination process and the recent advances in gene networks, Waddington's theory of development as a canalization of the epigenetic landscape shaped by gene networks [5] has gained more popularity. As predicted by this theory, a breakthrough may be achieved through a systems approach. Recent production of various “-omics” data has probed the gene networks from various aspects. Integrating the static interactome together with the expression and phenotypic profiles during a certain biological process can frequently reveal the dynamics of the gene network [6,7].

The protein-protein interaction networks (PPI, or interactome networks) have been shown to have dynamic modular structures [7,8], yet the functional connections between and among the modules are less well-understood. Through examining the dynamics of the interactome network, we found that two major network modules, the “P” (for proliferation) and “D” (for differentiation) modules are anticorrelated transcriptionally over adulthood in both the human brain and the fruit fly. These modules are enriched in proliferation and differentiation genes, respectively, and

display alternatively lower and higher expressions at the cellular proliferation/differentiation switch. Most P module genes are conserved between higher organisms and unicellular organisms such as yeast, but most D module genes are absent from unicellular organisms. Thus, these modules may correspond to alternative cellular states characteristic of higher organisms.

Results

Transcriptionally Anti-Correlated Modules in the Interactome Network

To investigate the dynamic features of the human interactome network through changes in gene expression, we used as surrogates the expression profiles on 30 postmortem human brains from subjects ranging from 26 to 106 years old. This dataset was originally generated to

Editor: Diana Murray, Weill Medical College of Cornell University, United States of America

Received May 8, 2006; **Accepted** September 21, 2006; **Published** November 24, 2006

Copyright: © 2006 Xia et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: GO, Gene Ontology; HPRD, Human Protein Reference Database; NP, negative-positive; PCC, Pearson correlation coefficient; P-D, proliferation-differentiation; PPI, protein-protein interaction

* To whom correspondence should be addressed. E-mail: jdhan@genetics.ac.cn

© These authors contributed equally to this work.

Synopsis

Coordination of proliferation and differentiation is a fundamental process of multicellular organisms. Although at the cellular level proliferation and differentiation seem to correspond to different cellular states that can sometimes be seen separated by the proliferation/differentiation temporal switch, it is unclear whether such switch-like property exists at the tissue or organism level or whether it exists in postmitotic tissues in adult animals. Through integrating protein–protein interaction networks with gene expression profiles, Xia, Xue, Dong, Zhu, and colleagues found that a switch temporally separating proliferation- and differentiation-associated modules can also be detected in the adult human brain and the adult whole fruit fly. The expressions of the two modules are well coordinated at the system level. The evolutionary origins of the proliferation and differentiation modules further implicate a symbiotic relationship between the two modules. Network topologies and gene annotations support a regulatory role of the protein–protein interaction interface between the two modules.

examine the age-related changes in gene expression and the biological functions related to aging [9]. In this study, we only focus on the gene expression patterns across different individuals and try to dissect the modular structure of the interactome network by similar or opposite expression profiles between a pair of genes.

As both transcriptional correlation and anti-correlation between a pair of genes are biologically relevant under specific conditions [10,11], we focused on the subnetwork that consists of only interactions between gene pairs that are transcriptionally correlated and anti-correlated (abbreviated as correlated and anti-correlated interactions, respectively) to examine the dynamic modular structure of the interactome. Such networks will be referred to as the NP network, where NP stands for negative and positive correlations. The expression correlations and anti-correlations between a pair of genes are commonly measured by a correlation coefficient. The Pearson correlation coefficient (PCC) is known to focus on the “shape” of changes rather than the intensity or amplitude of signals, and hence does not have bias for strong signals and has been shown to be best suited for oligo arrays [12,13]. It has a value of 1 for perfect correlation or -1 for perfect anti-correlation.

Our analysis pipeline includes the following steps (Figures 1 and S1): (1) obtain all the PPIs between genes that have similar expression profiles or opposite expression profiles determined by PCCs as correlated or anti-correlated interactions to arrive at the 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 manually dissecting the largest anti-correlated clusters or automatically scanning from the top of the hierarchical tree for clusters that have $<1\%$ intracluster anti-correlated interactions and finding the largest anti-correlated clusters with an average expression of $PCC < -0.7$.

Extracted from PPIs in the Human Protein Reference Database (HPRD) [14], the NP network across the human brain frontal cortex expression profiles [9] comprises 1,055

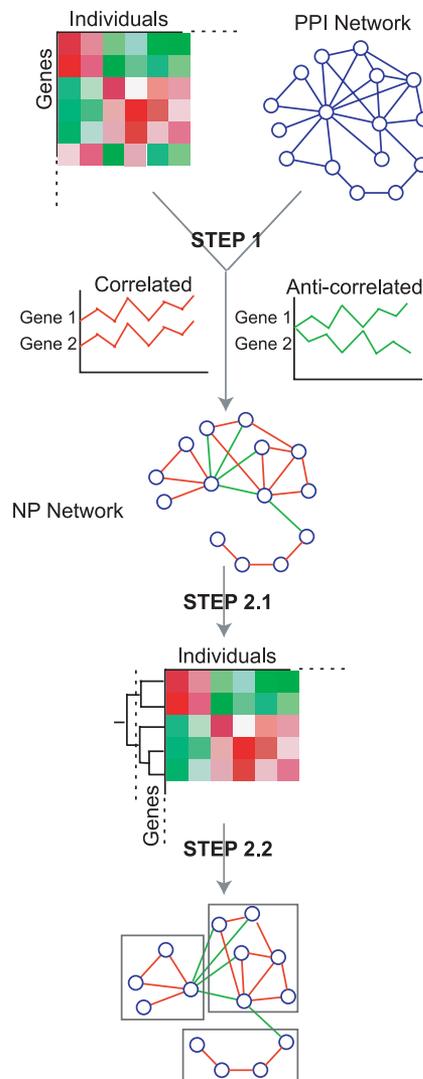


Figure 1. The Analysis Pipeline Used to Reveal the Anti-Correlated Modules

The analysis includes two major steps: step 1, calculating pairwise PCC for the transcriptional profiles of each pair of genes engaged in a PPI to extract the NP network from the PPI network; and step 2, applying hierarchical clustering to the genes in the NP network (step 2.1), then manually dissecting the largest anti-correlated clusters, or scanning from the top of the hierarchical tree for clusters that have $<1\%$ intracluster anti-correlated interactions (step 2.2) to approximate the goal of obtaining modules within which the expression profiles of genes are similar and correlated interactions are maximally enclosed, and in between which anti-correlated interactions are optimally distributed. A more detailed textual flowchart is available in Figure S1. doi:10.1371/journal.pcbi.0020145.g001

correlated and 395 anti-correlated interactions among 1,260 genes/proteins. We used PCC values of 0.4 and -0.4 as cutoffs for positive and negative correlations, respectively. These cutoffs have been established in previous studies. However, as described later, the identities of the clusters are not dependent on PCC cutoffs.

Using the hierarchical clustering algorithm implemented by Cluster [15,16] and visualizing the clusters with Tree View [15,17], we found that most of the anti-correlated interactions in the NP network bridge between two anti-correlated expression clusters among the genes (nodes) within the NP

network. This is apparent if the samples are clustered on the other dimension (Figure 2A). We named the two anti-correlated clusters “P” and “D” based on their associations with proliferation and differentiation functions, respectively (see below). Another large cluster that is largely correlated with P, but not obviously anti-correlated with D is named the “N” cluster for “not determined” function. The smallest cluster is named the “S” cluster for “small” (Figure 2A). There are 457, 435, 260, and 108 nodes, and 220, 316, 111, and 27 interactions within the D, P, N, and S clusters, respectively (Figure 2B and 2C). Genes in each module are listed in Table S1.

To examine the biological functions of each module, we first searched for overrepresented Gene Ontology (GO) categories among the genes within a module. To fully illustrate the preference of certain biological processes in one module versus others, we further grouped the related GO terms into a few broad categories, and performed a more comprehensive keyword search for genes potentially sharing the same process but not annotated by GO; these are listed at the end within each category in Table 1. The genes associated with the overrepresented GO terms and those found by the keyword search were also listed. According to Table 1, the D cluster is enriched in circulation/angiogenesis, apoptosis machinery, and ion and neurotransmitter channels, which are hallmarks of neural differentiation, cell cycle regulators, cell surface receptors, and steroid receptors. The P cluster is enriched in transcription, nuclear and intracellular transport, cell cycle, and cell motility genes. These enriched GO terms suggest that the P and D modules might be associated with cell proliferation and differentiation processes, respectively. The N cluster is enriched in genes involved in proteolysis, translation activity, intracellular transport, and energy metabolism. The S cluster is related to immunity (Table 1 and Figure 2B). A subset of the D genes (253 genes from D and one gene from N), the “SD” cluster, also anti-correlates with a subset of the N genes (250 genes from N and 40 genes from D), the “SN” cluster, across different subgroups of human subjects (Figure S2A and S2B).

The anti-correlated expressions of the D and P clusters are even more evident when the average gene expression levels of these clusters in each sample are compared across different human subjects, with a PCC reaching -0.867 ; that is, in almost each case when the expression of D is up, that of P is down, and vice versa.

Since genes and the functional relationships among the genes in a coexpressed cluster are frequently called a coexpressed module in a network [18], we will refer to these clusters as modules in the context of a network. The D module can be divided into four smaller submodules, DS1 to DS4, by reclustering only the D genes so that no anti-correlated interactions are within any submodules (Figure 2B). These submodules are, however, connected by more correlated than anti-correlated interactions (Figure S2C).

The P and D Modules Reflect the Dynamics of the Cellular Network

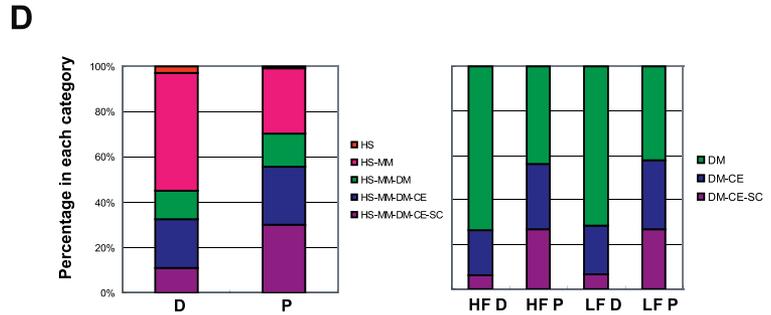
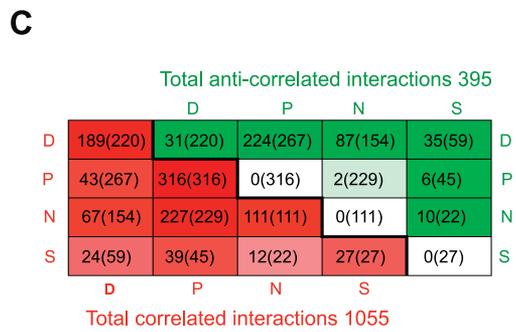
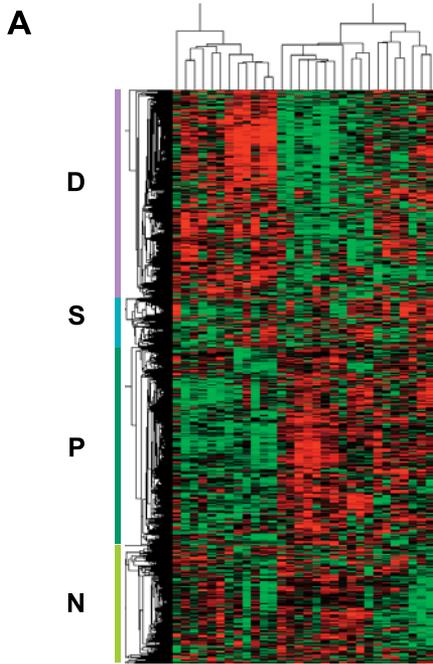
If the P–D partition is a feature of the dynamic transcriptional regulation in the adult brain, we would expect that the partition should not be dependent on whether or not the PPI network is integrated and should also be independent of the PCC cutoffs used to extract the NP network.

To test whether the physiological transcriptome is necessary for the P–D partition, or if network topology alone may give rise to such partitions, we permuted the expression values of each gene among different samples in the HPRD network, calculated the PCC of each HPRD interaction, and identified coexpressed modules using the automated pipeline (Figure 1). Among 100 such permutations, none of them gives rise to a pair of anti-correlated modules of more than 100 nodes per module (empirical $p < 0.01$; Control 1 in Table 2). Permuting gene expression intensities within each sample (Control 2 in Table 2), or permuting PCC values among different PPIs (Control 3 in Table 2) also renders the anti-correlated modules undetectable or barely detectable ($p < 0.01$ and $p = 0.08$, respectively). These randomization controls verify that the P–D partition is a true nature of the expression patterns, and cannot be derived by randomized expression patterns or pairwise relationship of expression profiles. As the networks in these controls have exactly the same network topology as the HPRD network, they also demonstrate that HPRD topology alone is not sufficient to give rise to the P–D partition. In other words, the P–D partition is not an artifact of network topology. In fact, it does not depend on any particular network at all.

We created a nonoverlapping PPI dataset to the original early version of HPRD interactions that consists of the incrementally updated interactions to HPRD since the earlier version (the interactions added to HPRD between November 22, 2004, and September 13, 2005) and two recently generated

Figure 2. The Transcriptionally Anti-Correlated Modules in the NP Network

(A) Two of the four major gene clusters, the D and P clusters, among genes in the NP network are anti-correlated. Human brain samples were clustered on the horizontal dimension according to their expression similarity. Genes were clustered on the vertical dimension. When the D gene cluster is upregulated, the P gene cluster is downregulated, and vice versa.
 (B) The NP network is reorganized to display the extensive anti-correlated interactions from D towards both the P and N modules, which are themselves linked by only correlated interactions, and the few anti-correlated interactions within the D module that separate the module into subgroups DS1 to DS4. Red edges represent correlated interactions; green edges, anti-correlated interactions. Nodes in the largest component of the network are shown in larger size. The most enriched biological functions of each cluster are listed beside the module labels.
 (C) The number of anti-correlated and correlated interactions within and between the D, P, N, and S modules. Red denotes correlated interactions; green, anti-correlated interactions. Color intensity indicates the proportion of interactions among the total correlated or anti-correlated interactions.
 (D) Evolutionary origins of P and D modules. Both the human and fruit fly P modules are more conserved from the single-cellular organism yeast to the multicellular organisms, whereas the D modules are more specific for multicellular organisms, with the human D module especially specific for human and mouse. Human genes in each module are divided into five different categories: those that are conserved among yeast, worm, fly, mouse, and human (purple); those that are conserved among worm, fly, mouse, and human (blue); and so on (green and pink); those that are only found in human are red. Fly genes are divided into those that are conserved among yeast, worm, and fly; among worm and fly; and those that are fly-specific. SC, CE, DM, RN, MM, and HS stand for yeast *Saccharomyces cerevisiae*, fly *Drosophila melanogaster*, worm *Caenorhabditis elegans*, rat *Rattus norvegicus*, mouse *Mus musculus*, and human *Homo sapiens*, respectively. HF and LF stand for “high food” and “low food”, respectively.
 doi:10.1371/journal.pcbi.0020145.g002



B

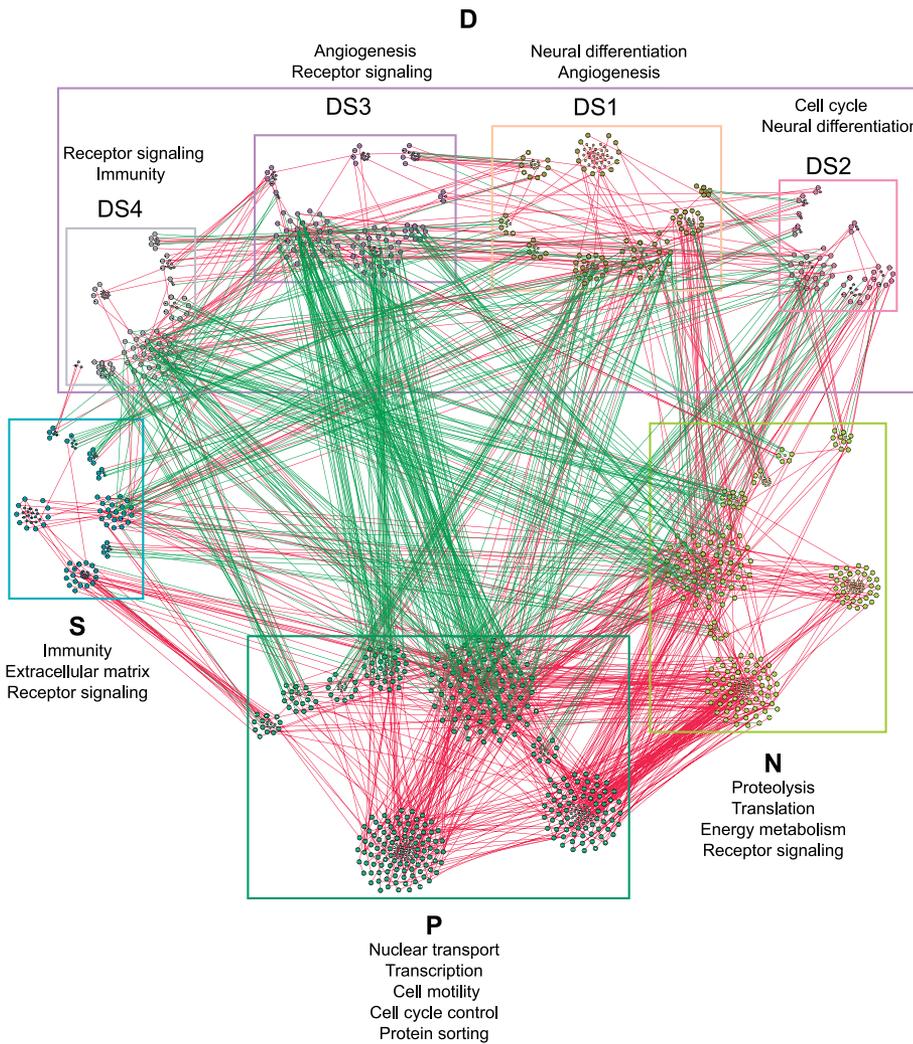


Table 1. GO Terms Overrepresented in Each of the Gene Expression Clusters

Cluster	Category	GO ID	GO Term	p-Value	Fold	Genes
D	Apoptosis	0008637	Apoptotic mitochondrial changes	1.36×10^{-3}	26.02	<i>BAK1, BAX, BCL2L1, BID</i>
		Other genes with keywords	Apoptosis, apoptotic, or cell death			<i>BAD, BCL10, BCL2, BRCA1, CASP2, CASP8, COL4A3, CRADD, DNASE1, ERCC2, FAS, FASLG, GZMB, IGF1R, IKBKKG, IL1B, KNG1, LTBR, MALT1, MAP3K10, MAPK1, MDM4, NOL3, PAX3, PIK3R2, PPP1R15A, PRKCA, PRKCE, PRLR, RSS25, RAF1, RHOB, SEMA4D, SGK, TNFRSF10C, TNFRSF25, TNFRSF7, TNFSF10, TNFSF8, TP53, TRAF1, TRAF3</i>
	Circulation/angiogenesis	0008015	Circulation	8.05×10^{-4}	6.92	<i>ACVRL1, AGTR1, APOB, AVPR1A, COL4A3, CXCL12, ELN, PLN, PPYR1, RHAG</i>
		0005021	Vascular endothelial growth factor receptor activity	5.94×10^{-3}	12.51	<i>FLT4, KDR, KIT, NRP2, PDGFRB</i>
	Other genes with keywords	Circulation, angiogenesis			<i>BAI2, EPAS1, FGF2, PGF, RHOB</i>	
	Cell-cycle control	0000075	Cell-cycle checkpoint	1.41×10^{-2}	11.83	<i>BRCA1, CDKN2A, RAD1, TP53</i>
		0007089	Traversing start control point of mitotic cell cycle	2.27×10^{-2}	16.26	<i>CDC2, CDC25C, CDC6</i>
	Other genes with keywords	Cell cycle			<i>AURKB, BAX, BCL10, BCL2, BRCA2, CCND1, CCND3, CCNE1, CDC20, CDC34, CDK3, CDK5, CDK6, CDK8, CHAF1A, CHAF1B, DMC1, DUSP1, DUSP4, E2F5, FGF2, FGF5, FRAP1, IGF1R, IL1B, JAG2, KIF23, LCK, MAPK1, MCMS, MDM2, MLH1, MNAT1, MYC, PCAF, PDGFB, PGF, PLK1, PPP1R15A, PRKCA, PTCH, RAD9A, RBBP4, RBL1, RHOB, SKP2, TFDP1, WT1, ZMYND11</i>	
	Neuronal differentiation	0007215	Glutamate signaling pathway	1.41×10^{-2}	11.83	<i>GRIA3, GRIK1, GRIK2, GRIN2A</i>
		0005911	Intercellular junction	1.49×10^{-2}	9.04	<i>CTNNB1, DLG1, DSC3, DSG2, PKP3</i>
		0005234	Glutamate-gated ion channel activity	1.53×10^{-2}	8.56	<i>GRIA3, GRIK1, GRIK2, GRIN2A, GRIN2D</i>
		0005230	Extracellular ligand-gated ion channel activity	2.27×10^{-2}	4.84	<i>CHRNA4, CHRNA7, CHRN4, GABRA3, GABRB2, GABRG3, HTR3A</i>
	Other genes with keywords	Neural, neuro, or ion channel			<i>APBA1, BAI2, CAMK1, CASP2, DLG3, DRD2, EFN3, FGF2, FGF5, FLNA, GRP, HAP1, HCRT, HCRTR1, KCNK3, NEDD4, NR2C2, NRP2, PAX3, PPYR1, RAB3A, SCNN1A, SCNN1B, SEMA4D, SIM2, STX1A, TRPC6</i>	
	Receptor signaling	0007229	Integrin-mediated signaling pathway	1.04×10^{-2}	5.54	<i>ITGA10, ITGA4, ITGA6, ITGAV, ITGB1, ITGB6, ITGB7, SYK</i>
		0008305	Integrin complex	1.04×10^{-2}	6.51	<i>ITGA10, ITGA4, ITGA6, ITGAV, ITGB1, ITGB6, ITGB7</i>
0003707		Steroid hormone receptor activity	2.28×10^{-2}	4.95	<i>ESR1, NR2C2, NR4A1, RXRB, RXRG, THRA, THRB</i>	
0007265		Ras protein signal transduction	1.17×10^{-2}	7.51	<i>CRKL, FGF2, GRAP, GRAP2, GRB2, LCK</i>	
0005070		SH3/SH2 adaptor protein activity	1.61×10^{-2}	5.30	<i>CRKL, GRAP, GRAP2, GRB2, IRS4, SHB, SIT</i>	
Other genes with keywords		Signaling			<i>ACVRL1, ADRA1B, AGTR1, AVPR1A, BAI2, BCR, CCBP2, CCL11, CCL13, CCL27, CCR3, CD4, CD8A, CHAF1B, CRSP2, CSF2RB, CSNK1D, CTNNB1, CXCL12, CXCR4, DHH, EFN3, EGF, EPHB3, ESR2, FASLG, FGF5, FGFR1, FGR, FLT4, FURIN, FYN, GABRA3, GABRB2, GABRG3, GCG, GLP1R, GNA13, GRIA3, GRIK1, GRIK2, GRIN2A, GRM7, GRP, GRPR, HCRT, HCRTR1, HTR1D, IGF1R, IL1B, JAG2, JAK3, KDR, KIT, LEF1, LNK, LTK, MUSK, PDGFRB, PGF, PIK3CG, PIK3R2, PLCB2, PPARBP, PPYR1, PRKAR2A, PRKCA, PRKCE, PRKCO, PTK6, PTPRD, PTPRR, RAF1, RGS12, RGS4, SMAD3, SMAD6, SMO, SNX1, SNX13, SNX4, SOCS1, STAT5A, STAT6, TGFB1, TGFB2, TIAM1, TNFRSF11A, TNFSF10, TNFSF8, TSHR, VAV1</i>	
P	Nuclear transport	0005643	Nuclear pore	2.02×10^{-4}	7.88	<i>CSE1L, IPO7, NUP133, NUP153, NUP98, RANBP2L1, RANBP3, RANBP5, TNPO1, XPO1</i>
		0000059	Protein-nucleus import, docking	1.43×10^{-3}	12.23	<i>CSE1L, IPO7, NUP98, RANBP5, TNPO1, XPO1</i>
	0005654	Nucleoplasm	2.44×10^{-3}	6.60	<i>CBX1, FMR1, KPNA2, NUP98, PTBP1, SMARCA5, TP53BP1, XPO1</i>	
	0008536	RAN protein binding	2.44×10^{-3}	19.81	<i>IPO7, RANBP2L1, RANBP3, RANBP5</i>	
	0008139	Nuclear localization sequence binding	2.03×10^{-2}	17.33	<i>KPNA2, RANBP5, TNPO1</i>	
	0030530	Heterogeneous nuclear ribonucleoprotein complex	2.11×10^{-2}	9.90	<i>HNRPA2B1, HNRPR, HNRPU, PTBP1</i>	
	Other genes with keywords	Nuclear translocation, nuclear translocator, nuclear transport, nuclear import, nucleus import, nuclear export, or nucleus export			<i>G3BP, GRP58, MAGOH, NXF1, RHOA, THOC1, XPOT</i>	
	Transcription	0016251	General RNA polymerase II transcription factor activity	3.03×10^{-3}	9.90	<i>GTF2E1, GTF2E2, GTF2H1, TAF1A, TAF1C, TCEA1</i>
		0008134	Transcription factor binding	1.51×10^{-2}	6.71	<i>DNMT1, HDAC1, HDAC4, HDAC9, HMGB1, NFKBIA</i>
		0006367	Transcription initiation from Pol II promoter	1.86×10^{-2}	6.30	<i>E2F3, GTF2A2, GTF2E1, GTF2E2, MED12, THRAP1</i>
0000118		Histone deacetylase complex	2.11×10^{-2}	9.90	<i>HDAC1, HDAC4, HDAC9, SAP18</i>	

Table 1. Continued.

Cluster	Category	GO ID	GO Term	p-Value	Fold	Genes
		Other genes with keywords	Transcription			ACTL6A,ATRX,BCL6,CBFA2T3,CCNT2, CEBPZ,COPS2,CREB1, CREBBP,CREG1, CTCF,DEDD,DEK,EED,FOXP1B,FOXP3A, GTF2H4,GTF3A,GTF3C1,HIF1A,HMGB2, HNRPD,HOMER1, HSF2,ILF2,ING1,JARID1A, LMO4,MBD2,MCM7,MECP2, MEF2C,MEIS1, MLLT7,MNT,NCOA4,NFE2L2,NR1D1,NR2C1, NR2F1,NR5A1,NSEPT1,OIP106,ORC2L, PARP1,PIAS3,PLAG1, POLR2A,PPARD,PRKAR1A,PTMA,RB1,RBL2,RCOR1,RFC1, RNF4,RNPC2,RNPS1,RPL6,RUNX1T1,RXRA,SAFB,SAFB2, SATB1,SMAD1,SMAD2, SMAD7,SMARCA5,SOX10,SOX9, SP3,SPEN, SQSTM1,SRF,STAT3,SUPT5H,SURB7, TADA3L, TCF3,TIP120A,TLE1,TNFAIP3, TP53BP1,TRIM28, TSG101, WWTR1, ZNF198
	Cell motility	0007266	Rho protein signal transduction	1.51×10^{-2}	8.67	ARHGAP1,PARG1,RHOA,ROCK1,TSC1
		Other genes with keywords	Motility			ACTN4,ACTR3,AMFR,CD9,CSPG3,CTGF, MAPK14, PAFAH1B1,PECAM1,STAT3,TLN1
	Cell-cycle control	0000075	Cell-cycle checkpoint	1.64×10^{-2}	12.60	ATR,FANCG,RB1,SMC1L1
		0000079	Regulation of cyclin-dependent protein kinase activity	2.03×10^{-2}	5.94	CCNT2,CDC37,CDKN1A,CDKN1B,CDKN1C, GTF2H1
		Other genes with keywords	Cell cycle			BIN1,CCT2,CCT7,CDC23,CLK1,CSPG6, CTCF,CUL2,E2F3, HDAC4,HDAC9,ING1, KPNA2,KRAS,MAPK6,MAPRE1,MCM7, MLLT7,MNT,MSH2,PARC,PCNA,PPM1D,PPP6C,PTMA, RAD17,RAD21,RBL2,SEPT2, SUPT5H, TADA3L,TSC1
	Protein sorting	0006891	Intra-Golgi transport	1.86×10^{-2}	7.88	COG2,COG5,COPB,COPB2,GGA2
		Other genes with keywords	Sorting			SNX2,VPS26
N	Proteolysis	0005839	Proteasome core complex (<i>sensu Eukaryota</i>)	5.01×10^{-6}	21.64	PSMA1,PSMA2,PSMA3,PSMA6,PSMB4, PSMB6,PSMB7
		0004175	Endopeptidase activity	5.01×10^{-6}	20.56	PSMA1,PSMA2,PSMA3,PSMA6,PSMB4, PSMB6,PSMB7
		Other genes with keywords	Proteolysis or proteasome			ADAM17,CASP3,CTSL,MBTPS1,NEDD8, NSF,PA2G4,PSEN2, PSMC4,TFRC,UBE3A
	Translation	0005852	Eukaryotic translation initiation factor 3 complex	2.23×10^{-5}	32.63	COP55,EIF3S10,EIF3S2,EIF3S3,EIF3S9
		0006446	Regulation of translational initiation	4.20×10^{-3}	11.30	DDX1,EIF2B5,EIF3S10,EIF3S2,EIF3S3
		Other genes with keywords	Translation			EIF2S1,GSPT1,SRP19,SRPR,SSR2,TFE3
	Energy metabolism	0006099	Tricarboxylic acid cycle	2.34×10^{-3}	13.35	FH,IDH3A,IDH3G,MDH1,MDH2
		0030060	L-malate dehydrogenase activity	1.31×10^{-2}	58.73	MDH1,MDH2
		Other genes with keywords	Metabolism			EIF2B5,GOT2,GSK3B,HDLBP,MBTPS1, NEU1,NME1,PAM, PFKM,PHYH,PRPS1,RAN, SORL1, TOPBP1,TYR,UNG2
	Receptor signaling	0004702	Receptor-signaling protein serine/threonine kinase activity	1.31×10^{-2}	58.73	AKT1,STK39
		0004705	JUN kinase activity	4.99×10^{-4}	58.73	MAPK10,MAPK8,MAPK9
		0004707	MAP kinase activity	4.20×10^{-3}	16.78	MAPK10,MAPK3,MAPK8,MAPK9
		0007254	JNK cascade	1.31×10^{-2}	11.75	MAP2K4,MAPK10,MAPK8,MAPK9
		Other genes with keywords	Signaling			ADAM17,APBB3,APLP2,AR,ARF1,BAG1, CAMKK2,CD3Z, DLGAP1,DLGAP2,FIBP, GNAI1,GNAI3,GNAS,GNB1, GSK3B,JAK1, PIK3R1,PRKCB1,PRKCI,PSEN2,RGS14, RGS2, RGS7,RNF14,SHANK2
	Protein sorting	0030125	Clathrin vesicle coat	1.75×10^{-2}	10.68	AP1S2,AP2M1,AP4S1,ARCN1
S	Immunity	0009596	Detection of pest, pathogen, or parasite	1.35×10^{-3}	69.16	HLA-DMA,HLA-DPB1,HLA-G
		0007259	JAK-STAT cascade	1.22×10^{-2}	21.84	FGFR3,JAK2,NMI
		0045012	MHC class II receptor activity	1.29×10^{-2}	23.05	HLA-DMA,HLA-DPA1,HLA-DPB1
		0019886	Antigen processing, exogenous antigen via MHC class II	1.29×10^{-2}	23.05	HLA-DMA,HLA-DPA1,HLA-DPB1
		0016010	Antigen presentation, exogenous antigen	1.72×10^{-2}	24.41	HLA-DMA,HLA-DPA1,HLA-DPB1
	Extracellular matrix	0019884	Dystrophin-associated glycoprotein complex	1.58×10^{-2}	55.33	SGCA,SSPN
		0005587	Collagen type IV	1.97×10^{-2}	46.11	COL4A1,COL4A2
	Receptor signaling	0008277	Regulation of G-protein-coupled receptor protein signaling pathway	1.97×10^{-2}	15.96	GIT2,RAMP2,RAMP3
		0006940	Regulation of smooth muscle contraction	2.19×10^{-2}	39.52	CNN1,PRKG1

The enrichment *p*-value cutoff is set to a significance value of 0.025 after Benjamini-Hochberg correction for the number of GO terms tested in each cluster. Genes having the overrepresented GO terms are listed to the right of the corresponding GO terms. The GO terms denoting similar biological processes are manually grouped into larger functional categories. The genes possessing keywords related to the enriched functional categories in their gene descriptions or GO annotations are also listed at the bottom of each category. doi:10.1371/journal.pcbi.0020145.t001

Table 2. Probability of Obtaining P–D Partition

Experiment/Control	Number	Description	Aim of Test
Experiment	1	HPRD	
	1'	HPRD autodissection	Fair comparison to other experiments and controls
	2	Incremental update of HPRD and Y2H	Dependence on a particular map of the interactome network
	3	All genes on the microarray	Sufficiency of expression pattern
	4	Partial extended PPI network with same number of nodes as HPRD	Dependence on a particular map of the interactome network
	5	Partial extended PPI network with same number of edges as HPRD	Dependence on a particular map of the interactome network
Control	6	Extend PPI network with different PCC cutoffs	Dependence on a particular PCC cutoff
	1	HPRD with shuffled expression among samples	Compare with Experiment 1' sufficiency of HPRD topology alone and necessity of expression pattern
	2	HPRD with shuffled expression among genes	Compare with Experiment 1' sufficiency of HPRD topology alone and necessity of expression pattern
	3	HPRD with shuffled expression PCCs	Compare with Experiment 1' sufficiency of HPRD topology alone and necessity of correct expression PCCs
	4	Partial extended PPI network with same number of nodes as HPRD without extracting NP	Compare with Experiment 5, effect of extracting NP
	5	Partial extended PPI network with same number of edges as HPRD without extracting NP	Compare with Experiment 6, effect of extracting NP
	6	Random with same degree distribution as HPRD	Compare to Experiment 1' effect of replacing PPI network with artificial interaction networks
	7	Random with same degree distributions as the sampled networks in Experiment 4	Compare with Experiment 4, effect of replacing PPI network with artificial interaction networks
8	Random with same degree distributions as the sampled networks in Experiment 5	Compare with Experiment 5, effect of replacing PPI network with artificial interaction networks	

$|PCC| > 0.4$ was used to filter the initially constructed networks to extract the NP network prior to clustering and module identification unless indicated by "without NP extraction," where the step is omitted. Modules in Experiment 1 are based on manual dissection of the clusters, whereas those in Experiment 1' are derived from automated dissection procedure. The GO annotations and expression changes at cellular proliferation to differentiation switch of P and D in the two experiments are similar, except the autoprotocedure spun off many neuronal genes from the D cluster to a smaller subcluster. Column 5 and 8 are the same; the former does not have a restriction on cluster size, the latter requires a minimum of 100 nodes in each cluster. Each test of Controls and Experiments 4 and 5 consisted of 100 trials. For each trial an automated module-finding algorithm was used to detect all coexpression clusters with $<1\%$ edges of $PCC < 0$; pairwise PCCs were calculated between the average expression levels among the largest three clusters; the two modules were defined as anti-correlated if $PCC < -0.7$. doi:10.1371/journal.pcbi.0020145.t002

Yeast Two-Hybrid (Y2H) datasets [19,20]. With this different PPI dataset, similar P, D, and N modules can be extracted from the NP network (Experiment 2 in Table 2 and Figure S3) that indicate the presence of P, D, and N modules is not a bias introduced by the original collection of HPRD interactions. Loosely aggregated coexpression clusters can be also derived from all genes on the microarray to significantly overlap with P (783 nodes) and D (1,276 nodes) modules (Experiment 3 in Table 2). The P–D partitions can actually be detected in almost any network of the same number of nodes (Experiment 4 in Table 2) or the same number of edges (Experiment 5 in Table 2) as the HPRD network but randomly sampled from an extended PPI network. The extended PPI network contains the updated HPRD plus two yeast two-hybrid interactome maps [19,20] and covers 7,568 proteins, 3,973 of which have expression profiles on the Affymetrix U95 array.

The P–D partition also does not depend on any particular PCC cutoff used to extract the NP network. In the extended interactome, using the automated module dissection pipeline at various PCC cutoffs or even without a PCC cutoff ($|PCC| > 0$), we could identify large clusters of genes that share significant overlaps with the P, D, and N modules. Although they correspond to smaller and smaller fractions of the total genes available on the microarray when the $|PCC|$ cutoff increases, the fraction of genes corresponding to P and D modules maximizes around $|PCC|$ cutoffs of 0.45 and 0.5 (Text S1 and Figure S4). Furthermore, relationships among the P, D,

and N clusters did not change; only the number of genes and interactions varied to some extent without altering most of the enriched functions in each clusters (unpublished data). Therefore, although P and D modules are identified as predominate modules in the NP network of $|PCC| > 0.4$, the two modules reflect the dynamics of the whole cellular network, which is not limited to those covered by the NP network.

Stable Module Detection by Integrating the Interactome with the Transcriptome

As demonstrated above, the presence of P, D, and N modules are not dependent on a particular PPI network being examined or a particular PCC cutoff. Why then do we need to integrate the PPI network and extract the NP network? The answer lies in the difference in the stability of detecting these network modules.

Unlike the gene clusters in the NP network, where 71% of the genes fall into P and D clusters, when all the genes in the full HPRD network are clustered, a pair of loosely aggregating anti-correlated gene clusters covers only 24% of the HPRD genes (Experiment 3 in Table 2, Text S1, and Figure S4). Similarly, the anti-correlation between the SD and SN was not visually clear when all the genes in the NP network were clustered (Figure 2A). Only after we examined the distribution of the anti-correlated interactions, of which a large number between the D and the N modules are evident (see below), we decided to further cluster only the genes in the D and N clusters. Then, an obvious anti-correlation between the

Table 2. Extended.

Number of Trials that Found Anti-Correlated Clusters	Average Number of Nodes per Cluster	Average Maximum Number of Nodes per Cluster	Percent of Trials That Found Anti-Correlated Clusters with >100 Nodes	Average Fraction of Overlap to P and D Genes	Average <i>k</i>
1/1	446	457	100	0.71	3.44
1/1	342.5	435	100	0.54	3.87
1/1	231.5	458	100	0.13	3.56
1/1	1531.5	2,237	100	0.07	NA
99/100	244.47	347.7	94	0.29	3.15
98/100	404.68	607.2	98	0.29	2.83
7/8	492.75	678.9	75	0.23	5.66
0/100	0	53.47	0	0.00	NA
0/100	0	169.8	0	0.00	NA
26/100	128.94	118.1	8	0.10	4.40
66/100	256.77	343.5	66	0.12	7.16
62/100	410.55	560.2	62	0.12	6.31
98/100	256.28	349.5	92	0.18	2.77
100/100	205.5	274.1	81	0.19	2.43
100/100	301.99	416.6	100	0.18	2.21

SD and SN clusters became visible (compare Figure 2A and Figure S2A). These suggest that by concentrating on only the correlated and anti-correlated interactions, we enriched for the genes of the P, D, and N clusters.

To more rigorously test the stability of finding the P–D partition, we compared the chance of finding anti-correlated network modules with 100 or more nodes in each module and their overlap to the P and D modules with or without extracting NP networks prior to clustering the genes in the network. The chance of detecting anti-correlated modules is 99%, or 98% when an NP network with a $|PCC| > 0.4$ was extracted from a randomly chosen partial PPI network with the same number of nodes or edges as the early version of the HPRD (Experiments 4 and 5 in Table 2). By omitting the step of extracting the NP network, the chances are reduced to 66% and 62%, respectively. In addition, the fraction of P and D module genes among all the input genes also reduces from an average of 29% of total input genes to 12% (compare Experiments 4 and 5 with Controls 4 and 5, respectively, in Table 2).

Choosing a different PCC cutoff also reduces the chance to 75% (Experiment 6 in Table 2, Text S1, and Figure S4), with the fraction of genes identified as P and D module genes reaching maximal levels at $|PCC| > 0.45$ or $|PCC| > 0.5$ (Text S1 and Figure S4). If the PPI network is replaced by a randomly generated network with the same degree of distribution as HPRD or a sampled PPI network, using $|PCC| > 0.4$ as a cutoff to extracting the NP network, although the chance of identifying anti-correlated modules are still high, these anti-correlated modules are of smaller average sizes and display low overlap (18%–19%) with P and D modules

(compare Controls 6–8 with Experiments 1, 4, and 5, respectively, in Table 2). These reductions in the probability of finding anti-correlated modules and further reductions in the identification of P and D modules among the input genes (or all modules) point to a role of using the appropriate PCC cutoff and integrating true PPIs on the stability of P and D module identification. However, the contribution of integrating the PPI network is not limited to the module identification, but more importantly is linked to the identification of the large PPI interface between the P and D modules that potentially coordinate the cellular proliferation and differentiation processes (see below).

These controls demonstrate that integrating the interactome, extracting the NP network, and applying an appropriate PCC cutoff ensured a high probability of stably detecting the P and D modules and improved their homogeneity, probably by filtering out most gene pairs that function in irrelevant tissue or cell types or under irrelevant physiological conditions.

Conservation of P–D Anti-Correlation in Other Species

The P–D partitions and their transcriptional anti-correlation can also be seen in the fruit fly. We used the adult whole-fly expression profiles to probe the dynamic gene relationships in the network. In the original publication [21], the expression profiles were used to study the effect of diet restriction on aging, and consisted of two sets of profiles: one for flies fed with a large amount of food, and one for those fed with a small amount of food, called “high-food” and “low-food” conditions, respectively. Here, we used these profiles to extract the network modules based on anti-correlated and correlated interactions across different fly populations using

the automated analysis pipeline described in Figure 1. The modules derived under high-food and low-food conditions are more than 50% identical. While the composition of the P module is largely conserved between the human brain and the fly, that of the D module is quite different between the two species. In particular, the apoptosis pathways are only enriched in the human brain D module. The enriched differentiation markers in D modules are also different; in the human brain, there are the neuronal markers; in the fly, there are genes involved in eye development (unpublished data), which is consistent with their tissue- and organism-specific requirements for differentiation.

We examined the percentages of human gene orthologs that can be found in yeast, worm, fly, and mouse. We found that 60% of D genes are specific to mouse and human and only 8% have yeast origin, whereas 35% of P genes have yeast homologs, and less than 30% are mammalian-specific (Figure 2D). Similar evolutionary patterns can be seen for fly P and D modules (Figure 2D).

The above observations indicate that the P module is more conserved from the single-cellular organism yeast to the multicellular organisms *C. elegans*, *Drosophila*, mouse, and human, while the D is multicellular-specific and is subjected to species-specific and probably also tissue-specific modifications.

The conservation of the P–D partition, their relationship, and the similar evolutionary profiles between fly and human also indicate that these observations cannot be due to sample variations introduced by sample preparations or other technical factors, but instead reflect true biological features of the gene networks of different multicellular organisms.

The P/D Temporal Switch Corresponds to the Proliferation and Differentiation Switches

A switch between differentiation and proliferation has been demonstrated in myoblast C2C12 cells [1]. Inhibition of the P–D interface protein HDAC4 has been shown to promote differentiation and inhibit proliferation, whereas inhibition to another interface protein, SRF, does the reverse [1]. Both the HDAC4 and SRF proteins are downregulated upon differentiation with a concurrent increase in differentiation markers and their antagonizing microRNAs [1]. The levels of $\alpha_5\beta_1$ integrin bound to fibronectin have also been shown to control the switching between proliferation and differentiation of C2C12 cells [4]. A proliferation/differentiation switch has also been observed in neural progenitor cells, and PI3K, cyclic AMP, raf, and MAPK pathways, which are all present at the P–D protein interaction interface, have all been implicated in regulating the switch [2,3]. These findings and many

others collectively point to the existence of the switch between proliferation and differentiation at the cellular level.

If the P and D modules are indeed associated with the proliferation and differentiation processes as suggested by the enriched GO annotations, we expect they might correspond to the cellular proliferation/differentiation switch in the tissue and organism we examined. Indeed, we found a decrease of P expression and an increase of D expression when fly, rat, mouse, or human cells of various cell types are switched from the proliferation to the differentiation state upon induction by various external stimuli. In this analysis, we used previously published data on human endometrial stromal cell differentiation induced by cyclic AMP, mouse C2C12 myoblast differentiation upon shifting to differentiation medium, mouse smooth muscle cell differentiation induced by retinoic acid, the inhibition of proliferation and induction of differentiation by the FGF of rat chondrocytes, and fruit fly neural progenitor cell differentiation (detailed sample information is available in Table S2). Consistent with the conservativeness of the P module, P is more uniformly suppressed upon differentiation of various different tissues in various different organisms. For example, the expression of fly P genes, especially of those derived under diet restrictions or low-food conditions, is suppressed in all cell types (Figure 3A, and middle and bottom rows in Figure 3B). In contrast, the expression of the human brain D is strongly induced during human endometrial stromal cell differentiation, and less so during mouse and fly cell differentiation; the expression of fly D genes is only most strongly induced in fly cells, but less so in cells of other organisms (Figure 3). Furthermore, detailed time courses of the proliferation/differentiation switch revealed that the P/D transition occurs only at the exact short window of the switch and are not observed before or after the switch (human endometrial stromal cell in Figure 3B), which accounts for some weak signals when the expression levels of all timepoints before or after the switch are averaged (Figure 3A). In addition to the association to proliferation and differentiation processes suggested by the overrepresented functional annotations in the P and D modules and transcriptional anti-correlations between the two modules, the correspondence to cellular level proliferation/differentiation switch more unequivocally supports the P/D temporal switch as the switch between proliferation and differentiation.

However, except in the development of compound fly eye [22], it is not known if cellular proliferation/differentiation switches are coordinated at the tissue or individual levels,

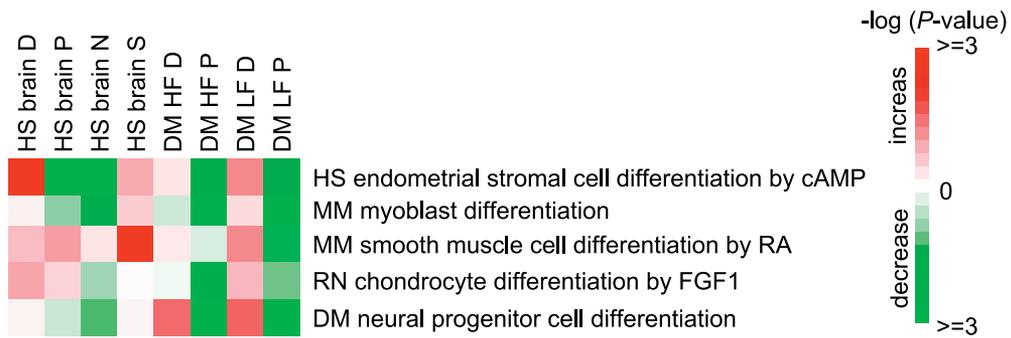
Figure 3. P and D Modules Correspond to the Proliferation/Differentiation Switch

(A) The suppression of P and induction of D expression upon cellular differentiation. The expression levels of genes in the human brain P, D, N, and S modules and fly P and D modules under high-food (HF) or low-food (LF) conditions (listed in column headers) are compared between the undifferentiated and differentiated samples (listed in row headers) by paired Student *t*-test. Red and green colors indicate an increase and a decrease in the differentiated samples, respectively. The color intensity represents the $-\log(p\text{-value})$ of the Student *t* test between the undifferentiated and differentiated samples. See text and Table S2 for details about the cell lines and differentiation conditions.

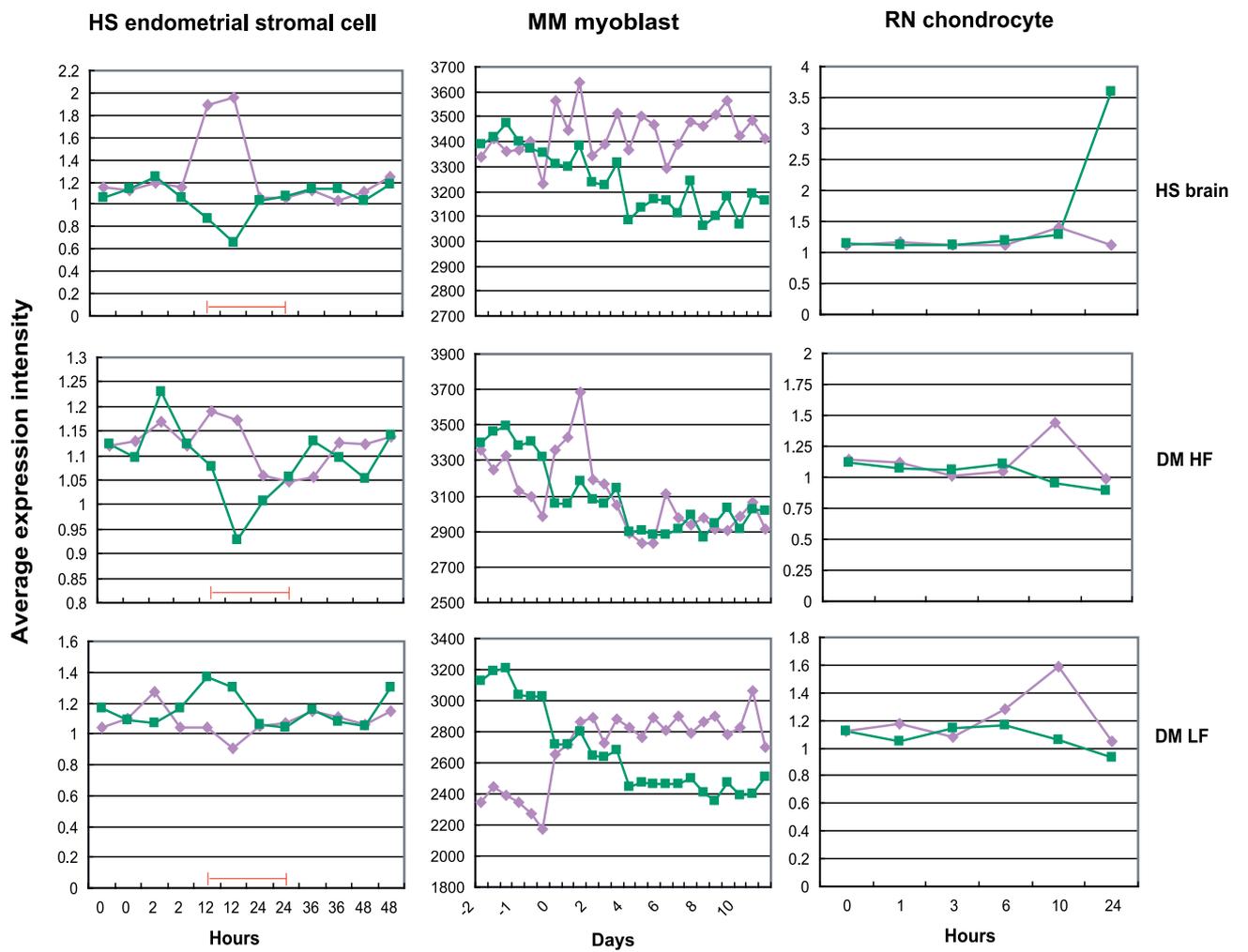
(B) The average expression level of P and D genes during the time course of the cell differentiation process of the human endometrial stroma cell (left column plots), the mouse myoblast (middle column plots), and the rat chondrocyte (right column plots). The plots in the top row are the average expression of human brain P and D genes (left column) or their mouse (middle column) or rat (right column) homologs upon induction of differentiation; plots in the middle and bottom rows are those of the human, mouse, and rat homologs of the fly genes in P and D modules derived under high- and low-food conditions, respectively. The expression levels of P and D modules are indicated by the green and lavender lines, respectively. The suppression of P and induction of D can occur during a short window. The proliferation/differentiation time window has been marked by a red line for human endometrial stroma cell differentiation (left column plots) as annotated by the original paper [33] where the experiments are published. The other experiments presented in panel A but not here include only single timepoints.

doi:10.1371/journal.pcbi.0020145.g003

A



B



especially in postmitotic tissues or among post-developmental adult animals. One way the systems level controls are achieved might be through circulating hormones and growth factors, as many of them and their downstream regulation molecules are present at the P–D interface.

Anti-Correlated Interactions Bridging the P–D Modules

In addition to facilitating the module detection, integrating the interactome and the transcriptome also revealed a large number of PPIs between a limited number of proteins forming a PPI interface between P and D modules. The high degree of interactions at the P–D interface cannot be obtained from randomly generated PPI networks of the same degree distribution as HPRD (compare Control 6 to Experiment 1' in Table 2; empirical $p < 0.01$, one-tail normal distribution $p = 1.17 \times 10^{-8}$). The degrees of the P–D interface proteins in sampled PPI networks are also significantly higher than those in the artificially created control networks of the matched degree distributions (compare Controls 7 and 8 with Experiments 4 and 5 in Table 2; one-tail Student t -test $p = 1.66 \times 10^{-41}$ and 1.93×10^{-63} , respectively).

As expected, anti-correlated interactions preferentially bridge between the transcriptionally anti-correlated P and D modules. More than half (58%) of the correlated interactions are within the coexpressed modules, and 22% are between the P and N modules, whereas 57% of the anti-correlated interactions bridge between the P and D modules, and 22% bridge between the D and N modules (Figure 2C). The probability of the anti-correlated interactions bridging the P and D modules compared with bridging any modules is 1.8 times that of the uncorrelated interactions (PCC between -0.4 and 0.4), which is a very significant difference (Fisher exact test $p = 2.399 \times 10^{-20}$).

In principle, the anti-correlated interactions can occur among different subjects or in different developmental stages and, as a consequence, bridge various coexpressed modules as small as one- or two-gene modules; therefore, it is surprising to see that two major coexpressed modules comprised 71% of the genes in the network, anti-correlated with each other across 77% of the samples, and were connected by the majority of the anti-correlated interactions in the network.

Regulatory Role of the P–D Interface

From the GO terms overrepresented at the interfaces, the P interface is enriched for transcription factors and the D interface is enriched in cell-cycle checkpoint, DNA repair, and receptor signaling genes (Table 3). All these processes are important regulatory mechanisms in the proliferation/differentiation switch. In particular, the D interface proteins include many of the well-known tumor suppressor genes,

such as *BRCA-1* and *p53*, and many receptors and transcription regulators known to be required for neuron differentiation, such as MYC, TOP2B, integrin, estrogen, FGF, PDGF, and TSH receptors, and many A kinase-anchoring proteins (AKAPs), etc. The P interface contains genes promoting cell proliferation, such as *K-RAS*, *HDACs*, *SRE*, *CREB*, *CREBBP*, *ILAR*, and *INSR* (the insulin receptor gene). It also contains genes that inhibit p53 and BRCA-1 functions such as *PARC* and *LMO4* (Table 3 and Figure S5). We further evaluated the potential regulatory role of the P–D interface by three well-known network and biological properties of regulatory genes: (1) genes playing crucial regulatory roles are often hubs in the network, and vice versa; (2) if the interface plays crucial regulatory roles in the proliferation/differentiation switch, malfunction of these genes may lead to cancer, and thus these genes are on average more likely to be oncogenes or tumor suppressor genes; and (3) regulatory genes function in regulatory pathways, where feedback control is a dominant network feature. We therefore compared the protein interaction degrees, the percentage of oncogenes and tumor suppressor genes, and the percentage of genes in the feedback loops between the interface and the non-interface, or core genes. The results of all three tests are consistent with a crucial regulatory role in the P/D switch: compared with the core of the P and D modules, the P and D interfaces have a much higher average interaction degree (Figure 4A; $p = 2.27 \times 10^{-12}$), percentage of known oncogenes and tumor suppressor genes (Figure 4B; $p = 3.28 \times 10^{-2}$ and 2.08×10^{-4} for P and D interfaces, respectively), and percentage of proteins/genes located in feedback loops (Figure 4C; $p = 1.07 \times 10^{-2}$ and 3.45×10^{-4} for P and D interfaces, respectively). Even though all the feedback loops are still of very limited coverage, it is already evident that most of these feedback controls are between the P and D modules and mediated by anti-correlated interactions (Figure 4D). Nearly all the proteins involved in these feedback loops are transcription regulators, and many of the loops are formed by both PPI and transcriptional regulations (Figure 4D). These special features of the P–D interface proteins make them potential key regulators for the proliferation/differentiation switch.

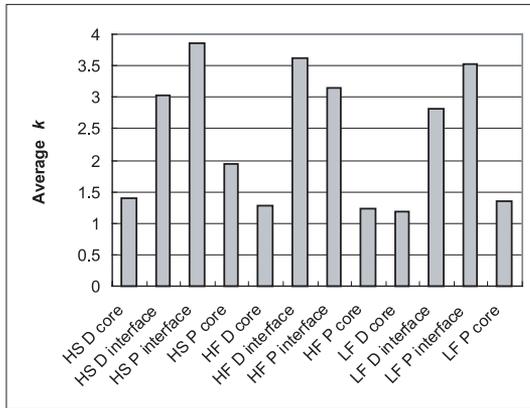
Alternating expression of genes can be brought upon by the “toggle switch” network circuit, which is a feedback loop consisting of two mutually inhibitory interactions between the nodes [23]. If we treat the P and D modules as single nodes in a module network, the P and D expression pattern can be also achieved through a simple toggle switch design between them (Xia et al., unpublished data). However, in a complex system involved in differentiation and proliferation control,

Figure 4. Regulatory Roles of the Protein Interactions between P–D Modules

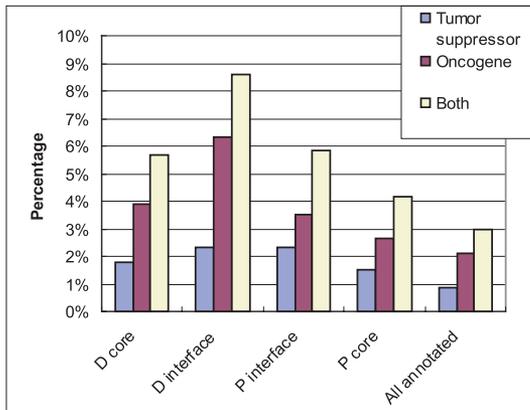
- (A) The average degree k of the nodes at the human brain and fly P–D protein interaction interfaces and inside the modules (cores).
 (B) The percentage of proto-oncogenes and tumor suppressor genes at the human brain P–D protein interaction interfaces and inside the modules (cores).
 (C) The percentage of genes located in feedback loops at the human brain P–D protein interaction interfaces and inside the modules (cores).
 (D) Network consisting of only the feedback loops traversing protein interaction interfaces and inside the modules (cores). Solid edges represent directional protein interactions; dashed edges, transcriptional regulations. Red and green edges represent transcriptional correlations and anti-correlations, respectively.
 (E) Feedback pathways potentially controlling P/D switch in the adult human brain. Feedback loops of three or more nodes that traverse the P–D interface are listed. Pathways 5, 8, 9, and 10 are related; 6 and 7 are also related, and they are between all three modules. The font colors for D, P, and N genes are lavender, green, and light green, respectively. Correlated and anti-correlated interactions are shown as red and green arrows, respectively; the arrow points from an upstream to a downstream gene.

doi:10.1371/journal.pcbi.0020145.g004

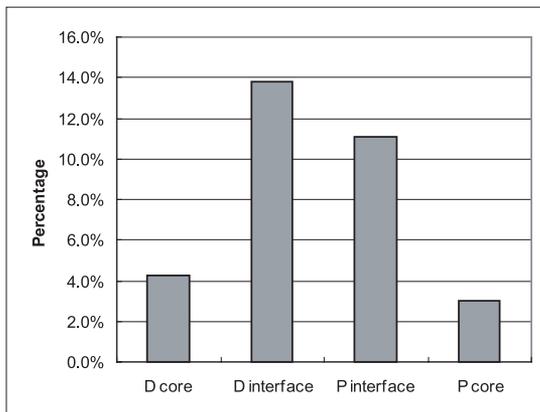
A



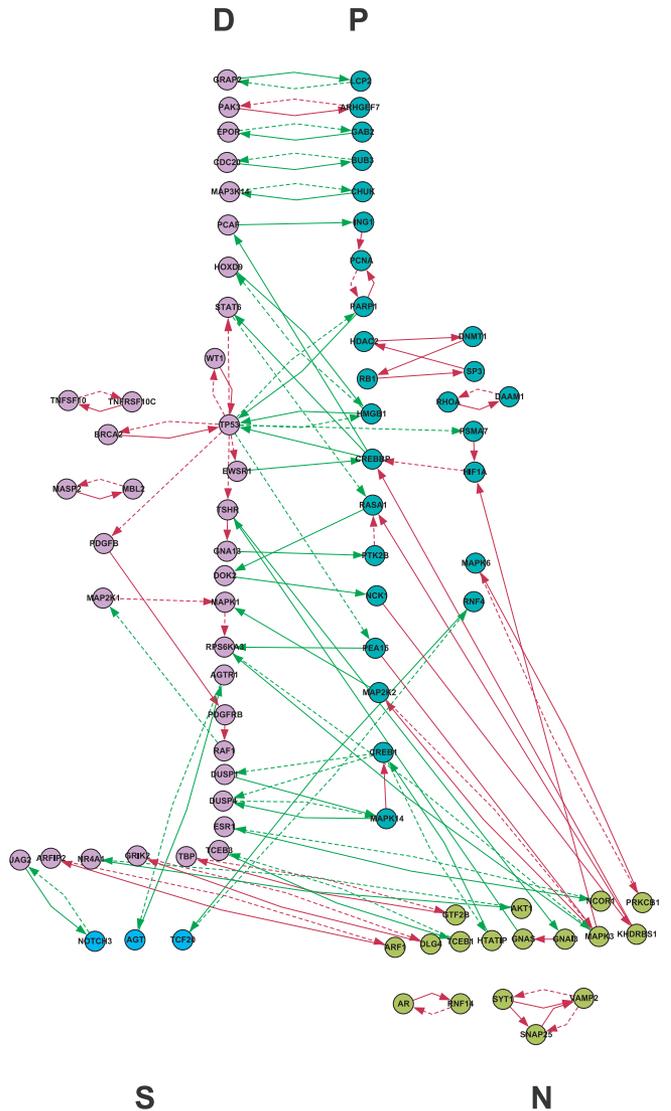
B



C



D



E

1. TP53 → PSMA7 → HIF1A → CREBBP → TP53
2. TP53 → EWSR1 → CREBBP → TP53
3. CREBBP → DUSP1 → MAPK4 → CREBBP
4. CREBBP → DUSP4 → MAPK4 → CREBBP
5. TP53 → PEA15 → MAPK3 → HIF1A → CREBBP → TP53
6. STAT6 → RASA1 → DOK2 → NCK1 → KHDRBS1 → CREBBP → STAT6
7. RASA1 → DOK2 → NCK1 → KHDRBS1 → RASA1
8. TP53 → PEA15 → RPS6KA3 → MAPK3 → RPS6KA3
9. TP53 → PEA15 → MAPK3 → RPS6KA3 → MAPK3
10. MAPK1 → RPS6KA3 → MAPK3 → MAP2K2 → MAPK1

much more redundancy and fine-tuning than one feedback control might be implemented. As examples of potential feedback controls of the differentiation and proliferation in the adult human brain, we can list at least ten interesting signaling pathways of three or more nodes that traverse the P–D interface (Figure 4E). Highlighting a unique advantage of such integrative systems analysis, these pathways are not just a collection of known pathways preexisting in the literature. Although all the interactions are derived from the literature, and most genes in these pathways have been shown to affect differentiation, proliferation, or growth, the pathways themselves have not been reported previously and are not known as control circuits for coordinating differentiation and proliferation processes.

Altogether, the P and D modules are not only transcriptionally anti-correlated across different individuals, they are also functionally associated with basic cellular proliferation and differentiation functions, evolutionarily represent cell-autonomous and multicellular-specific modules, and are respectively suppressed and induced at the cellular proliferation/differentiation switch in various cells of various multicellular organisms. The functional interdependence, antiphase temporal compartmentalization, and different evolutionary origins of the two modules suggest that the P and D modules are two counterparts in a symbiotic relationship that need to be tightly controlled and coordinated at the cellular, tissue, and organism levels by switching temporally between the two phases—proliferation and differentiation.

Discussion

In this study, we describe a new integration analysis of the interactome and the transcriptome, which, even though rather straightforward, is very effective at removing the analysis noise of a conventional gene clustering process and allowed us to robustly find the P and D modules and their transcriptional anti-correlation. Moreover, the most important contribution of PPI integration is to reveal the P–D interface, which has a potential regulatory role in coordinating the proliferation and differentiation processes.

We found that anti-correlation goes beyond the individual gene pairs but between the gene populations—a pair of transcriptionally anti-correlated gene groups. The P and D modules seem to be associated with cellular proliferation and differentiation and are suppressed and induced at the cellular proliferation/differentiation switch, respectively, therefore corresponding to alternative states of the cellular network. This indicates that logical relationships also exist at the modular level in the cellular networks. A possible scenario for anti-correlation at the modular level is that it might reflect a temporal separation of the biological functions in the cellular network [24]. The metabolic cycle has been suggested to fulfill such a role of temporal compartmentalization of oxidative and reductive metabolism in eukaryotes [24]. The antiphase temporal compartmentalization of proliferation and differentiation has been demonstrated over and over for single molecules upon switching from proliferation to differentiation at the cellular level [1–4], but it is surprising that such relationships also exist at the tissue or organism level during adulthood or might be brought upon by transcriptionally anti-correlated modules through complex feedback mechanisms between the two modules. Our result is therefore

consistent with Waddington's [5] view of the development, in which differentiation and proliferation correspond to two states of the network where a balance between them is achieved at a systems level. More importantly, the tissue-level and organism-level coordination during adulthood and the evolutionary conservation level of the P and D modules imply the balance is not restricted to the single-cell level during early development, but rather exists during the whole life of an organism.

Although such an expression pattern can be achieved by a simple toggle switch between them (Xia et al., unpublished data), in a complex system, redundancy is often implemented to ensure robustness; therefore, multiple toggle switches may exist between the two modules, and the switches must be connected with each other to transfer information. The exact molecular mechanisms giving rise to the transcriptional correlated and anti-correlated modules at the systems level during adulthood remain to be determined; we expect that many signaling pathways involved in cancer formation and aging will be part of the control mechanisms. But due to methodology limitations in the past, most of these pathways have been only studied individually; a general and comprehensive mechanism is still lacking. Our identification of the P and D modules at the systems level has provided an entry point for arriving at such a general mechanism. It is possible that the genes in each module share a few common immediate upstream transcription regulators. For example, we have found that the 5' untranslated regions of the fly P module genes are clearly enriched for Dref-targeting sites among a few other less-well characterized sequence motifs (Xue et al., unpublished data). Ectopic overexpression of Dref has been shown to block the proliferation/differentiation switch in the fruit fly eye imaginal disc [22].

As the P module is concentrated at transcription-level activities and the N module is concentrated at protein-level activities, a temporal delay between transcription and translation might account for the lack of complete synchronization between the two clusters. Even though the samples are mostly from subjects of different ages, the timescale reflected by these samples may not be restricted to age differences; a delay between translation and transcription may well be reflected as individual differences.

Although the current coverage of the interactome comprising both the literature and large-scale yeast two-hybrid interactions is still limited [25], the conservation of the P–D pattern in the human brain and fruit fly across different datasets indicates that the coverage is sufficient at the current level to detect, annotate, and analyze the P and D modules. In the NP network, we only focused on the strongly correlated and strongly anti-correlated interactions, but the genes excluded this way may also play important regulating functions toward the temporal compartmentalization between P and D modules or in the proliferation/differentiation control. Nevertheless, our identification of the large interconnected P–D modules for the first time revealed a proliferation/differentiation switch and their interrelationship at a systems level. It opens a new avenue to examine differentiation and proliferation at the systems and network levels, and provides a channel to connect physiological level events, such as hormone secretions, to the underlying cellular and molecular changes. It will help to elucidate many complex biological processes.

Table 3. GO Terms Overrepresented in the Interface and Core (Noninterface) of the P and D Modules

Cluster	Category	GO ID	GO Term	p-Value	Fold Genes	
D core	Apoptosis	0008632	Apoptotic program	1.44×10^{-2}	24.07 <i>BAD,CASP2,CASP8</i>	
		Other genes with keywords	Apoptosis, apoptotic, or cell death		<i>BAX,BCL10,BCL2,BID,CRADD,ERCC2, GZMB,IKBK,IL1B,KNG1, MALT1, MAP3K10,MDM4,NOL3,PIK3R2,PRKCA,PRLR,RHOB, SEMA4D,TNFRSF10C, TNFRSF7,TNFSF10,TNFSF8, TRAF3</i>	
	Neuronal differentiation	0005911	Intercellular junction	1.50×10^{-2}	16.04 <i>DLG1,DSC3,DSG2,PKP3</i>	
		Other genes with keywords	Neural, neuro, or ion channel		<i>CASP2,CHRNA4,CHRN4,DLG3,EFNB3,FGF5,GABRA3, GABRG3,GRIK2,GRIN2A,GRP,HCRT,HCRTR1,NEDD4,NRP2, RAB3A, SCNN1B,SEMA4D,SIM2</i>	
	Circulation/ angiogenesis	0003805	Coagulation factor Xla activity	1.44×10^{-2}	72.20 <i>F11,F12</i>	
		0001525	Angiogenesis	1.50×10^{-2}	10.03 <i>EPAS1,KDR,NRP2,PGF,RHOB</i>	
		0008015	Circulation	1.53×10^{-2}	9.22 <i>ACVRL1,APOB,AVPR1A,ELN,PLN,RHAG</i>	
		Other genes with keywords	Circulation, angiogenesis			
	D interface	Cell-cycle control	0000075	Cell-cycle checkpoint	1.06×10^{-2}	23.33 <i>BRCA1,RAD1,TP53</i>
			Other genes with keywords	Cell cycle		<i>AURKB,CCND1,CCNE1,CDC2,CDC20,CDC25C,CDK3,CDK8, CHAF1A,DUSP1,DUSP4,FGF2,MAPK1,MCM5,MLH1,MNAT1, MYC,PCAF,RBBP4,TFDP1,WT1</i>
Receptor signaling		0007265	Ras protein signal transduction	1.06×10^{-2}	13.16 <i>CRKL,FGF2,GRAP2,GRB2</i>	
		0007229	Integrin-mediated signaling pathway	1.16×10^{-2}	9.10 <i>ITGA4,ITGAV,ITGB1,ITGB7,SYK</i>	
		0017017	MAP kinase phosphatase activity	1.39×10^{-2}	25.67 <i>DUSP1,DUSP2,DUSP4</i>	
		0008330	Protein tyrosine/threonine phosphatase activity	1.51×10^{-2}	57.04 <i>DUSP2,DUSP4</i>	
		0030284	Estrogen receptor activity	1.55×10^{-2}	85.55 <i>ESR1,ESR2</i>	
		0008305	Integrin complex	2.29×10^{-2}	9.78 <i>ITGA4,ITGAV,ITGB1,ITGB7</i>	
		Other genes with keywords	Signaling		<i>AGTR1,CSNK1D,CTNNB1,FASLG,FYN,GABRB2,GNA13,GRIA3, GRIK1,JAK3,KIT,LEF1,LTK,MUSK,PDGFRB,PIK3CG,PPARBP, PRKCE,RAF1,RGS12,RGS4, SMAD3,SMAD6,SNX1,SNX4,STAT6, TSHR, VAV1</i>	
		DNA repair	0006265	DNA topological change	1.78×10^{-2}	28.52 <i>TOP1,TOP2A,TOP2B</i>
0008408			3'-5' exonuclease activity	2.10×10^{-2}	16.04 <i>POLA,POLD1,RAD1</i>	
0003918			DNA topoisomerase (ATP-hydrolyzing) activity	2.10×10^{-2}	42.78 <i>TOP2A,TOP2B</i>	
0003684			Damaged DNA binding	2.32×10^{-2}	9.01 <i>BRCA1,GTF2H3,RAD1,RAD51</i>	
Other genes with keywords			DNA repair		<i>CHAF1A,CSNK1D,HUS1,MNAT1,RBBP4</i>	
P core		Nuclear transport	0005643	Nuclear pore	4.00×10^{-3}	10.30 <i>NUP133,NUP153,NUP98,RANBP2L1, RANBP3,TNPO1</i>
	0030530		Heterogeneous nuclear ribonucleoprotein complex	4.00×10^{-3}	21.58 <i>HNRPA2B1,HNRPR,HNRPU,PTBP1</i>	
	0006406		mRNA-nucleus export	1.75×10^{-2}	11.19 <i>FMR1,MAGOH,NUP133,THOC1</i>	
	Other genes with keywords	Nuclear translocation, nuclear translocator, nuclear transport, nuclear import, nucleus import, nuclear export, or nucleus export		<i>G3BP,RHOA,XPOT</i>		
	Cell-cycle control	0000079	Regulation of cyclin dependent protein kinase activity	9.00×10^{-3}	10.79 <i>CCNT2,CDC37,CDKN1A,CDKN1B,CDKN1C</i>	
		0004861	Cyclin-dependent protein kinase inhibitor activity	9.05×10^{-3}	28.32 <i>CDKN1A,CDKN1B,CDKN1C</i>	
		Other genes with keywords	Cell cycle		<i>ATR,BIN1,CCT2,CCT7,CDC23,CLK1,CSPG6, CTCF,HDAC9, MLL7,PPP6C,RAD21, SUPT5H</i>	
	Protein sorting	0006891	Intra-Golgi transport	1.07×10^{-2}	13.73 <i>COG2,COG5,COPB2,GGA2</i>	
		0005788	Endoplasmic reticulum lumen	1.73×10^{-2}	20.60 <i>HSPA5,PPIB,TRPC4AP</i>	
	Other genes with keywords	Sorting				
Translation	0005853	Eukaryotic translation elongation factor 1 complex	1.80×10^{-2}	50.34 <i>EEF1B2,EEF1D</i>		
	Other genes with keywords	Translation		<i>DEK,EIF1AX,EIF2B4,EIF3S5,EIF3S6,EIF3S7, EIF5B,PAIP1</i>		
P interface	Transcription	0006367	Transcription initiation from Pol II promoter	9.15×10^{-3}	13.27 <i>E2F3,GTF2A2,GTF2E2,MED12,THRAP1</i>	
		Other genes with keywords	Transcription		<i>ATRX,BCL6,CEBPZ,COPS2,CREB1,CREBBP,EED,FOXO3A, GTF2H1,HDAC1,HDAC4,HMGB1,HNRPD,HOMER1,ILF2,ING1, LMO4,MCM7,MEI1,MNT,NCOA4,NFKBIA,NR2F1,PARP1, POLR2A,PRKAR1A,PTMA,RB1,RBL2,RCOR1,RFC1,RNPC2, RXRA,SAFB,SATB1,SMAD2,SOX10,SPEN, SRF,STAT3, TADA3L,TAF1C,TLE1,TNFAIP3, TP53BP1,TSG101,WWTR1</i>	

Materials and Methods

Datasets. The HPRD dataset [14] was downloaded from www.hprd.org on November 22, 2004; a later version of the HPRD dataset was obtained on September 13, 2005; two human Y2H datasets were included in the extended PPI network [19,20]. Two Y2H screens were combined as the fly protein interaction dataset [26,27].

Microarray expression profiles were obtained from previously published studies on postmortem human brains of subjects between 26 and 106 years of age [9] and on adult *Drosophila melanogaster* of various ages [21].

GO annotations were downloaded from <ftp://ftp.ncbi.nlm.nih.gov/genefunction/> on March 10, 2005.

Lists of proto-oncogene and tumor suppressors were obtained at <http://ca.expasy.org/cgi-bin/get-entries?KW=Anti-oncogene> and <http://ca.expasy.org/cgi-bin/get-entries?KW=proto-oncogene> on July 4, 2005.

Orthologs. Human orthologs in mouse, fly, worm, and yeast were identified as the best reciprocal BlastP hits with an *e*-value cutoff of 10^{-6} based on RefSeq protein sequences downloaded on December 9, 2004, from <http://www.ncbi.nlm.nih.gov/RefSeq>.

Filtering GO terms. To filter out the GO terms [28] that are broadly associated with many proteins, we calculated the number of proteins each GO term associated with and used only the GO terms that have a detection probability (*p*-value) among randomly paired genes less than the significance value of 0.05 after Bonferroni correction for multiple hypothesis testing on the total number of GO terms in each species. The *p*-value for a GO term is defined as $p = (n/t)^2$, where *n* is the number of genes associated with the GO and *t* is the total number of genes in the species; a GO term was used only if $p < (0.05/g)$, where *g* is the total number of GO terms associated with all the genes in a species.

GO-term enrichment was determined by Fisher exact test followed by Benjamini-Hochberg correction [29] for multiple hypothesis testing on all the GO terms tested in each gene set.

Feedback loops. The NP network was searched with the breadth first-search algorithm for pathways that have the same start and end nodes with lengths between two (minimum, one protein and one regulatory edge between two nodes) and ten based on all the directed protein and regulatory interactions we annotated or extracted. Protein interaction annotation was based on the PubMed abstracts of the references listed by HPRD. Regulatory relationships between PPI partners within the NP network were extracted from PubMed by the Pathway Assist text-mining function (Ariadne Genomics, <http://www.ariadnegenomics.com/products/pscentral>). Other regulatory relationships were predicted from the transcription factor motifs annotated by the TRANSFAC database or from the p53 chromatin immunoprecipitation experiment [30], then filtered with pairwise expression $|PCC| > 0.4$.

Expression clustering and visualization. Unsupervised agglomerative hierarchical clustering of genes was performed in Cluster 3.0 [15,16]. The expression values were first adjusted by the following operations: log transform, median center genes, normalize genes, median center arrays, and normalize arrays. Then, hierarchical uncentered correlation and centroid linkage were used for clustering in both dimensions. The clustering results were visualized in JavaTreeView 1.0.12 [15,17].

The layout of the reorganized NP network to visualize the intercluster interactions was created with a new visualization tool we developed in Java (Hou et al., unpublished data) and imported into Cytoscape 2.1 [31]. All other network visualization was achieved directly with Cytoscape 2.1.

Networks of a predefined degree distribution were generated by an algorithm used by Milo et al. [32].

Supporting Information

Figure S1. A More Detailed Textual Flowchart of the Pipeline for Revealing Anti-Correlated Modules

Found at doi:10.1371/journal.pcbi.0020145.sg001 (240 KB PDF).

Figure S2. The SD and SN Modules and the Interaction Distribution in the SD-SN and DS1-4 Coexpressed Modules

(A) SD and SN clusters. The gene expressions of these two clusters are also anti-correlated between two sample clusters across 63% of the samples. The two sample clusters are different from those giving rise to D and P gene differential expressions. The average expression intensities of the genes in the SD and SN modules are also anti-correlated across different individuals with a PCC of -0.744 .

(B) Number of correlated and anti-correlated interactions within and between the SD and SN modules.

(C) Number of anti-correlated and correlated interactions within and between DS1 to DS4 modules, and those between the D submodules and the other major modules (P, N, and S modules).

Found at doi:10.1371/journal.pcbi.0020145.sg002 (382 KB PDF).

Figure S3. Overlaps of HPRD Modules to Those Derived from the Extended Minus HPRD PPI Network

Even though there is not a single edge or interaction in common between this network and the HPRD network, three modules derived from it share significant overlaps to P, D, and N modules, respectively.

Found at doi:10.1371/journal.pcbi.0020145.sg003 (337 KB PDF).

Figure S4. The P-D Partition Does Not Depend on a PPI Network or PCC Cutoffs, but the Right [PCC] Cutoff Can Facilitate its Identification

(A) Hierarchical clustering followed by an automated exhaustive search (for clusters that contain less than 1% of interactions of PCC < 0) identified anti-correlated modules of sizes ranging from 100 to 1,200 genes that significantly overlap the original P and D modules when a more comprehensive PPI network and various PCC cutoffs (including no PCC cutoff) were used to extract NP network. Even without integrating a PPI network, a careful manual search can identify loosely aggregated gene expression clusters of 783 and 1,276 genes that significantly overlap with the P and D modules. When the PCC cutoff is > 0.7 , no gene cluster that has more than 100 genes can be identified as significantly overlapping with the D module. Input gene sets together with the number of genes in each set (in parentheses) are listed as the row headers on the left. The sizes of the clusters identified under various PCC cutoffs or without a PPI network are listed as the row headers on the right, and those of the original P, D, N, and S modules are listed on the column headers on the top. The number of genes overlapping between the original and latter examined clusters is indicated in each cell of the matrix; the intensity of the background color of a cell reflects the overlap significance as $-\log(p\text{-value})$ by Fisher exact test normalized by standard deviation among all the cluster overlaps examined, which equals 44. (B) The fractions of P-D genes within the NP network at various [PCC] cutoffs used to extract the NP network from the extended PPI network. A maximal fraction is achieved at [PCC] cutoffs of 0.45 and 0.5 (yellow bars). Because some N module genes are merged into the P modules at [PCC] cutoffs of 0.45 and 0.5, the fractions of the module genes that overlap with the original NP network P and D modules are also plotted to exclude the potential bias introduced by the inclusion of some N genes. The fraction of P and D module genes overlapping with the original P and D modules (blue bars) display a similar trend to the fraction of total genes in each module (yellow bars).

Found at doi:10.1371/journal.pcbi.0020145.sg004 (342 KB PDF).

Figure S5. The P-D Protein Interaction Interface

The cores of the modules are represented as big squares on either side of the interfaces.

Found at doi:10.1371/journal.pcbi.0020145.sg005 (239 KB PDF).

Table S1. Gene List of Each Module

Found at doi:10.1371/journal.pcbi.0020145.st001 (161 KB XLS).

Table S2. Sample Information of Cell Differentiation Induction Experiments

Found at doi:10.1371/journal.pcbi.0020145.st002 (137 KB PDF).

Text S1. The P-D Partition Is a Feature of the Expression Pattern in the Cellular Network

Found at doi:10.1371/journal.pcbi.0020145.sd001 (11 KB PDF).

Acknowledgments

We thank Drs. Nicholas Baker, Jiandie Lin, Siming Li, Li Cai, Xu Zhang, and Yong Liu for helpful discussions and critically reading the manuscript.

Author contributions. JDJH conceived and directed the project and designed the analyses. KX, HX, DD, SZ, QZ, and LH performed the analyses. ZF contributed materials and analysis tools. YGC and JDJH wrote the paper. KX, HX, DD, SZ, JW, HC, RT, and ZH annotated the interaction directions according to the literature.

Funding. This work was supported by grants from the China National Science Foundation (grants 30588001 and 90508006) and the Bai Ren Plan of the Chinese Academy of Sciences to JDJH and

grants from the China National Science Foundation (grants 30125021 and 30430360) and the 973 Program (2004CB720002) to YGC.

References

- Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, et al. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* 38: 228–233.
- Conti L, Sipione S, Magrassi L, Bonfanti L, Rigamonti D, et al. (2001) Shc signaling in differentiating neural progenitor cells. *Nat Neurosci* 4: 579–586.
- Dugan LL, Kim JS, Zhang Y, Bart RD, Sun Y, et al. (1999) Differential effects of cAMP in neurons and astrocytes. Role of B-raf. *J Biol Chem* 274: 25842–25848.
- Garcia AJ, Vega MD, Boettiger D (1999) Modulation of cell proliferation and differentiation through substrate-dependent changes in fibronectin conformation. *Mol Biol Cell* 10: 785–798.
- Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150: 563–565.
- Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, et al. (2005) Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* 436: 861–865.
- Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 430: 88–93.
- Stelling J, Sauer U, Szallasi Z, Doyle FJ III, Doyle J (2004) Robustness of cellular functions. *Cell* 118: 675–685.
- Lu T, Pan Y, Kao SY, Li C, Kohane I, et al. (2004) Gene regulation and DNA damage in the ageing human brain. *Nature* 429: 883–891.
- Dhillon IS, Marcotte EM, Roshan U (2003) Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics* 19: 1612–1619.
- Qian J, Lin J, Luscombe NM, Yu H, Gerstein M (2003) Prediction of regulatory networks: Genome-wide identification of transcription factor targets from gene expression data. *Bioinformatics* 19: 1917–1926.
- D'Haeseleer P (2005) How does gene expression clustering work? *Nat Biotechnol* 23: 1499–1501.
- Jansen R, Greenbaum D, Gerstein M (2002) Relating whole-genome expression data with protein–protein interactions. *Genome Res* 12: 37–46.
- Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, et al. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* 13: 2363–2371.
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863–14868.
- de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20: 1453–1454.
- Saldanha AJ (2004) Java Treeview—Extensible visualization of microarray data. *Bioinformatics* 20: 3246–3248.
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D, et al. (2003) Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34: 166–176.
- Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al. (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437: 1173–1178.
- Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al. (2005) A human protein–protein interaction network: A resource for annotating the proteome. *Cell* 122: 957–968.
- Pletcher SD, MacDonald SJ, Marguerie R, Certa U, Stearns SC, et al. (2002) Genome-wide transcript profiles in aging and calorically restricted *Drosophila melanogaster*. *Curr Biol* 12: 712–723.
- Jasper H, Benes V, Atzberger A, Sauer S, Ansorge W, et al. (2002) A genomic switch at the transition from cell proliferation to terminal differentiation in the *Drosophila* eye. *Dev Cell* 3: 511–521.
- Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339–342.
- Tu BP, Kudlicki A, Rowicka M, McKnight SL (2005) Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science* 310: 1152–1158.
- Han JD, Dupuy D, Bertin N, Cusick ME, Vidal M (2005) Effect of sampling on topology predictions of protein–protein interaction networks. *Nat Biotechnol* 23: 839–844.
- Formstecher E, Aresta S, Collura V, Hamburger A, Meil A, et al. (2005) Protein interaction mapping: A *Drosophila* case study. *Genome Res* 15: 376–384.
- Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. (2003) A protein interaction map of *Drosophila melanogaster*. *Science* 302: 1727–1736.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—A practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B Methodol* 57: 289–300.
- Wang L, Wu Q, Qiu P, Mirza A, McGuirk M, et al. (2001) Analyses of p53 target genes in the human genome by bioinformatic and microarray approaches. *J Biol Chem* 276: 43604–43610.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504.
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: Simple building blocks of complex networks. *Science* 298: 824–827.
- Tierney EP, Tulac S, Huang ST, Giudice LC (2003) Activation of the protein kinase: A pathway in human endometrial stromal cells reveals sequential categorical gene regulation. *Physiol Genomics* 16: 47–66.

Competing interests. The authors have declared that no competing interests exist.