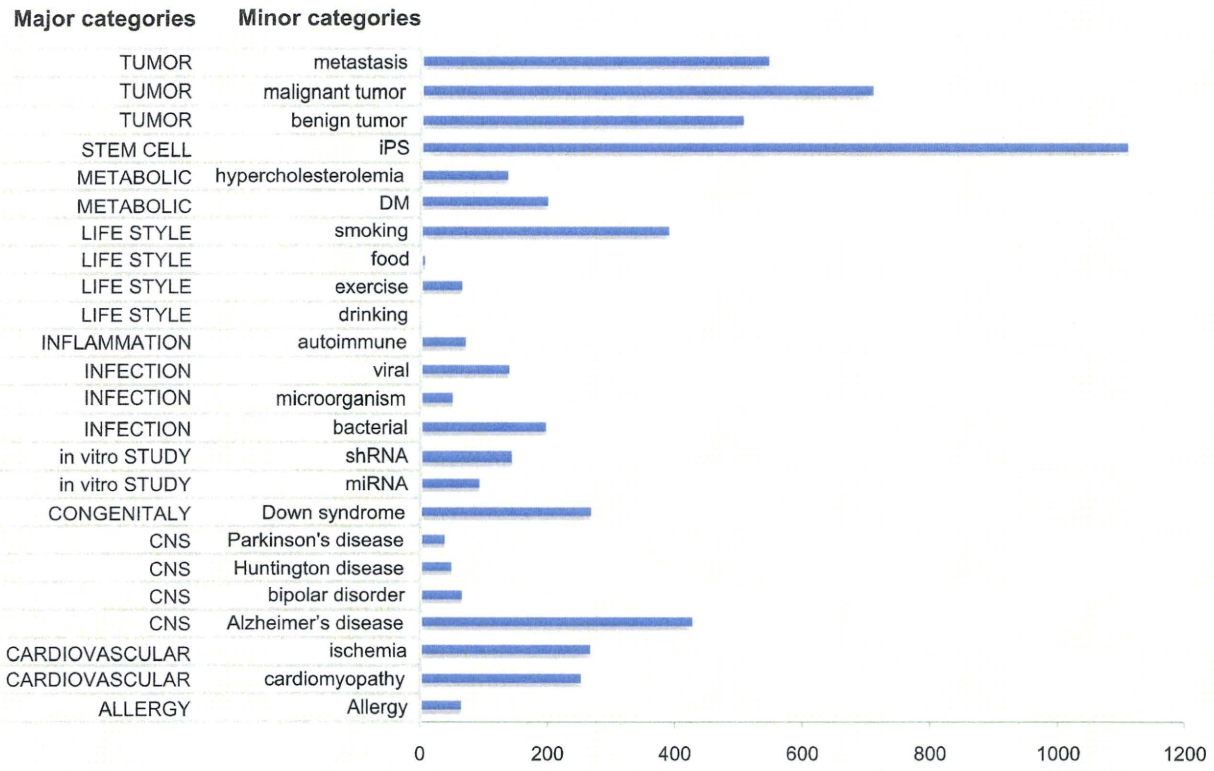


Figure 4. ArrayExpress experimental categories for microarray datasets and mean number of BTS pairs with significant ON/OFF change. There were few BTS pairs with significant changes for "lifestyle" and many with significant changes for "cancer." Note the higher number of BTS pairs for iPS cells than for donor cells.
doi:10.1371/journal.pcbi.1000851.g004

prostaglandin metabolism, is closely correlated with the differentiation grade of HCC [52]. Nonsteroidal anti-inflammatory drugs targeting PTGS2 have been shown to inhibit the proliferation of cultured hepatocellular cancer cells by inducing cell-cycle arrest [53].

When HCC tissue was compared with healthy liver tissue, toggling was most evident for CCNA2(cyclin A2)–related BTSs (Fig. 7) We therefore analyzed how the toggling of CCNA2-related BTSs rippled out to other BTS pairs during the malignant transition of HCC (Fig. 8).

CCNA2 activates CDC2 or CDK2 kinases and regulates the cell cycle positively by promoting G1/S and G2/M transitions in both the G1 and G2 phases of the cell cycle [54], while EGR1 (early growth response gene 1) has suppresses transformation [55]. The upregulation of CCNA2 and downregulation of EGR1 might thus play a key role in the dysregulation of normal growth in HCC carcinogenesis [56]. The downregulation of IL6 (interleukin 6) is involved in dysregulation of the immune response in early carcinogenesis.

After the toggling of CCNA2-related BTSs but still in the early stage of carcinogenesis, the OFF state of IL6 is related to the ON states of PTK2 and SMAD3 (SMAD family member 3). PTK2 and SMAD3 play important roles in cell growth and the activation of intracellular signal transduction pathways, suggesting that cell proliferation might accelerate during this stage.

Toggling of PTK2(ON)-BCL2(OFF) was observed in advanced and very advanced stages. BCL2 (B-cell CLL/lymphoma 2) suppresses apoptosis, and the downregulation of BCL2 might be involved in the acceleration of apoptosis in cancer cells.

Notably, the ON/OFF state of the TP53-IGF1 BTS was changed from "OFF-OFF" to "ON(TP53)–OFF(IGF1)" in advanced HCC. And in very advanced HCC, almost all IGF1-related BTS pairs demonstrated "ON(other)–OFF(IGF1)" patterns.

In the very advanced stage, many IGF1(insulin-like growth factor-1)-related BTS pairs demonstrated significant ON/OFF changes. The liver is the main source of IGF1, and the development of HCC is accompanied by significantly reduced serum IGF1 levels [57]. The downregulation of IGF1 and upregulation of a set of another pair of genes might affect a wide variety of cellular functions.

Discussion

We constructed bistable switch networks, compared their ON/OFF states with those of control (healthy) samples, and found that their states changed with disease progression and differed between patient subtypes. Since most disease states exhibit a certain level of resilience against therapeutic intervention, each can be considered to be homeostatic to some extent. This homeostasis implies the robust status of a dynamical network and could not be maintained without mechanisms that drive a network to maintain a certain state. One such mechanism is a bistable switch, so we should look for sets of bistable switch circuits in large-scale protein interaction networks.

Our analysis revealed that BTS states change with disease progression, and the implications of this are far reaching. For example, it might be possible to prevent or delay disease

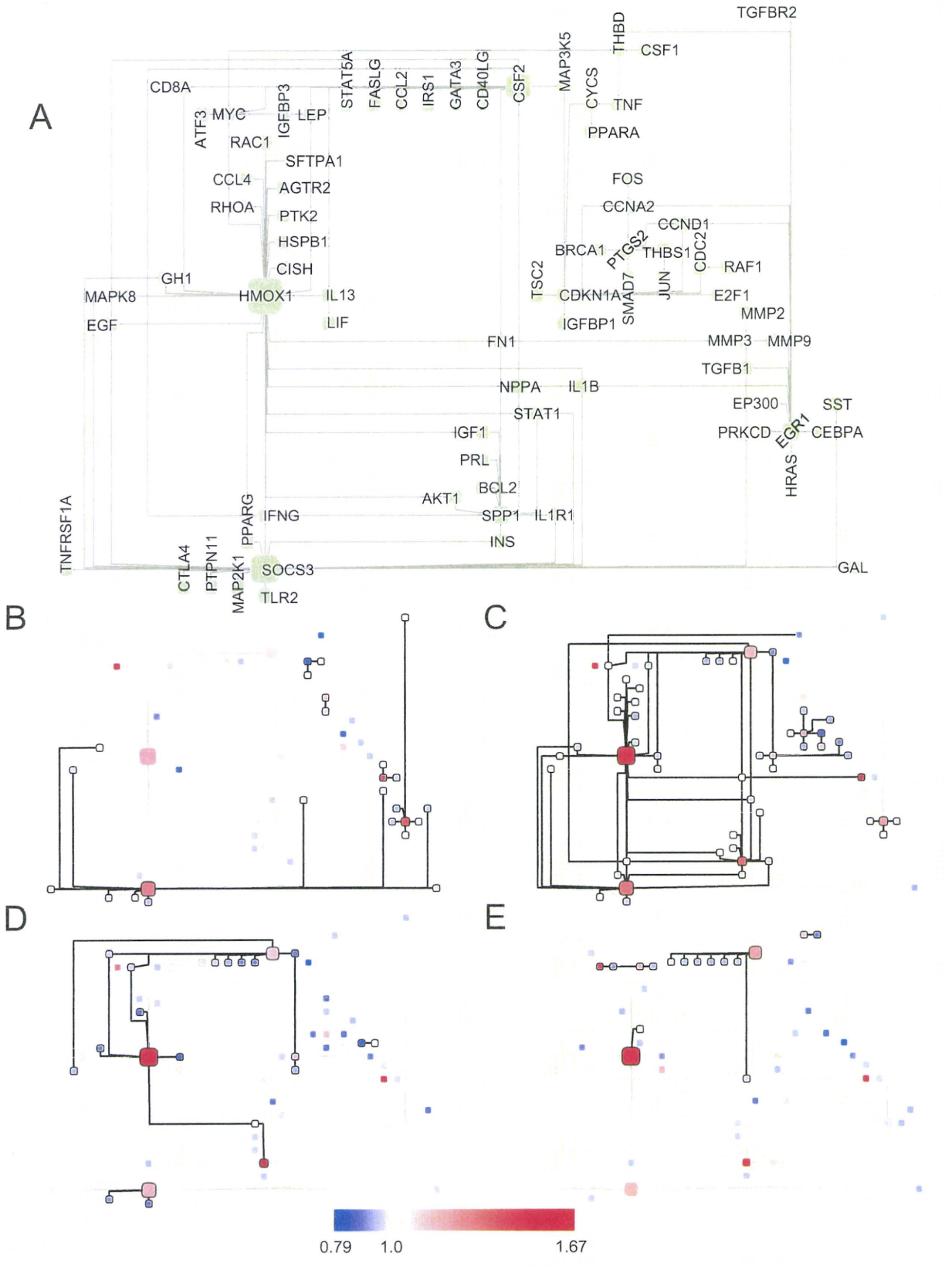

A
TGFBR2
THBD
CSF1
MAP3K5
CYCS
TNF
PPARA
STAT5A
FASLG
CCL2
IRS1
GATA3
CD40LG
CSF2
CD8A
ATF3
MYC
IGFBP3
LEP
RAC1
SFTPA1
CCL4
AGTR2
RHOA
PTK2
HSPB1
CISH
GH1
MAPK8
HMOX1
IL13
EGF
LIF
FOS
CCNA2
CCND1
BRCA1
PTGS2
THBS1
CDC2
RAF1
TSC2
CDKN1A
SMAD7
JUN
E2F1
MMP2
IGFBP1
FN1
MMP3
MMP9
TGFB1
NPPA
IL1B
STAT1
EP300
SST
IGF1
PRKCD
EGR1
CEBPA
PRL
HRAS
BCL2
AKT1
TNFRSF1A
PPARG
IFNG
SPP1
IL1R1
INS
CTLA4
PTPN11
MAP2K1
SOCS3
GAL
TLR2
B
C
D
E
0.79
1.0
1.67

Figure 5. Changes in ON/OFF states of BTSs for time series data for human normal bronchial epithelial cells exposed to smoke. A: Toggling inferred from time-dependent data (ArrayExpress ID: E-GEOD-10700 and E-GEOD-10718) for the mRNA expression of normal human bronchial epithelial cells exposed to cigarette smoke for 24 hours. B: 2 hrs after exposure start, C: 4 hrs after exposure start, D: 8 hrs after exposure start, E: 24 hrs after exposure start. The nodes represent genes and the edges between them represent the pairing of bistable toggle switches. The colors of nodes were automatically assigned as a continuous color gradient from red for ON (upregulated) to blue for OFF (downregulated) according to relative gene-expression levels of the nodes. In Figs. 4B–E, the BTS pairs framed by thick lines are pairs with significant toggling scores at that time.
doi:10.1371/journal.pcbi.1000851.g005

progression by perturbing one or more such switches. Such switches may be novel drug-target candidates for controlling disease progression. Analysis of the ON/OFF states of genes constituting bistable circuits revealed similarities between disease subtypes.

While our analysis has provided insightful information, it has shortcomings. First, the network topologies were based on commercial databases created using a text-mining system. This means that the details of the molecular interactions were not verified. The development of a more accurate interaction database would enable more precise and accurate analysis of bistable network behaviors and of the contributions of switch circuits to those behaviors. Second, the analysis was based solely on network topologies—no parametric features were considered. Although topological analysis enabled us to identify circuits exhibiting bistable behavior, whether circuits exhibiting bistable behavior apparently exhibit bistable behavior depends on the kinetic parameters associated with each interaction [58].

Using microarray data, we determined that the pairs of genes in the circuits we identified are polarized into ON and OFF states. Two mutually inhibitory nodes polarized into ON and OFF states do not function as a bistable switch if both genes are ON or OFF. This is why we focused on BTSs, which demonstrated "ON/OFF" or "OFF/ON" states. We should, however, note that the "ON/ON" states of some BTSs play important roles in mammalian embryogenesis [59], T-cell differentiation [60], and visual-system specification [61].

Cluster analysis of transcriptome data in microarrays is useful for classifying disease characteristics according to differences in expression patterns. Although several disease types that are difficult to classify morphologically have been classified using this approach, the rules underlying the cluster structure of the data are unclear, and the importance of each of the molecules in a cluster cannot be determined with a reasonable degree of certainty. The analysis of changes in gene-expression levels can also be used to create a list of molecules whose levels increase or decrease significantly over time or whose levels differ significantly between healthy and diseased tissues. Although examinations of gene interrelations using gene-ontology classification and analysis of the classification results using network diagrams have led to a greater degree of understanding of the changes in molecular networks, it is difficult to infer the meanings of biological interactions between molecules.

Our proposed method (i.e., focusing on BTS ON/OFF changes) takes as the starting point the interactions between molecules. This makes it easy to infer biological meaning and makes it possible to analyze time-dependent data for time periods corresponding to that of disease progression (from hours to years). In addition, while conventional methods sometimes neglect molecules that are downregulated, our method places equal importance on both increases and decreases in expression.

DNA microarray technology makes it possible to study the expression of thousands of genes at the same time, but much of the microarray data consists of low signal intensities that can produce erroneous gene expression ratios between control and experimental samples [62]. The distribution of the ratio of two random variables approaches a Cauchy, or Lorentzian, distribution, which has longer tails than Gaussian distributions [63,64]. In our results, far more BTS pairs had significant toggling scores than did random gene pairs, but a considerable number of random gene pairs did show significant ON/OFF changes. We should therefore consider the possibility of random error in the analysis of BTS pairs.

We used the transcriptome of normal tissue as the control in our analyses. This means that the identification of the molecular ON/OFF states inherent to normal tissue was unclear. Even if the ON/OFF state of a molecular pair for a certain switch is important for a particular tissue, if this state is retained in the diseased tissue, we would be unable to detect it in the present study because the ON and OFF states are not mutually exclusive. Therefore, molecules exhibiting even the slightest change are emphasized while those showing no change are ignored. We aim to overcome this drawback by identifying what types of ON/OFF changes occur in switches when embryonic stem (ES) cells or iPS cells undergo differentiation.

Since proteins are responsible for cell function, the ON/OFF state of a molecule must be determined at the protein level when searching for molecular-network structures mediating cell functions. Because there are more than 20 control steps along the way from mRNA to functional proteins [65], the reported expression levels of mRNA do not always agree with those of proteins—their translated products [66]. And even if there were a quantitative correlation between the levels of mRNA and functional protein, the efficiency of the translation process would be greatly affected by factors such as structural change and protein localization. Proteomics data for proteins in different cellular contexts is useful but is available for only some proteins. Transcriptome data analysis is the only method currently available for examining molecular networks on a large scale, but when testing the quality of BTS pairs in the future we will use all the relevant available data for the target proteins. Furthermore, to ensure bistability, the hysteresis phenomena must be confirmed when a perturbation has vanished. By conducting time-scale experiments in both directions when applying and removing perturbations, we should be able to further test the quality of BTS pairs.

Despite its shortcomings, the approach presented here provides useful insights into the states of biological networks, insights that may lead to discovery of novel drug targets and therapeutic interventions.

Materials and Methods

Preparation of basic interaction datasets

The lists of molecular interactions were constructed using the Ariadne Genomics ResNet human protein interaction database (ver. 3.0) compiled, using MedScan [67] natural language processing technology, from more than 13,000,000 PubMed abstracts and 43 publicly available full-text journals. The database contains data on over 200,000 objects (proteins and small molecules) and over 100,000 interactions.

The interactions can be divided into two major classes: direct physical interactions (binding, protein modifications, and promoter binding) and indirect regulatory interactions (regulation,

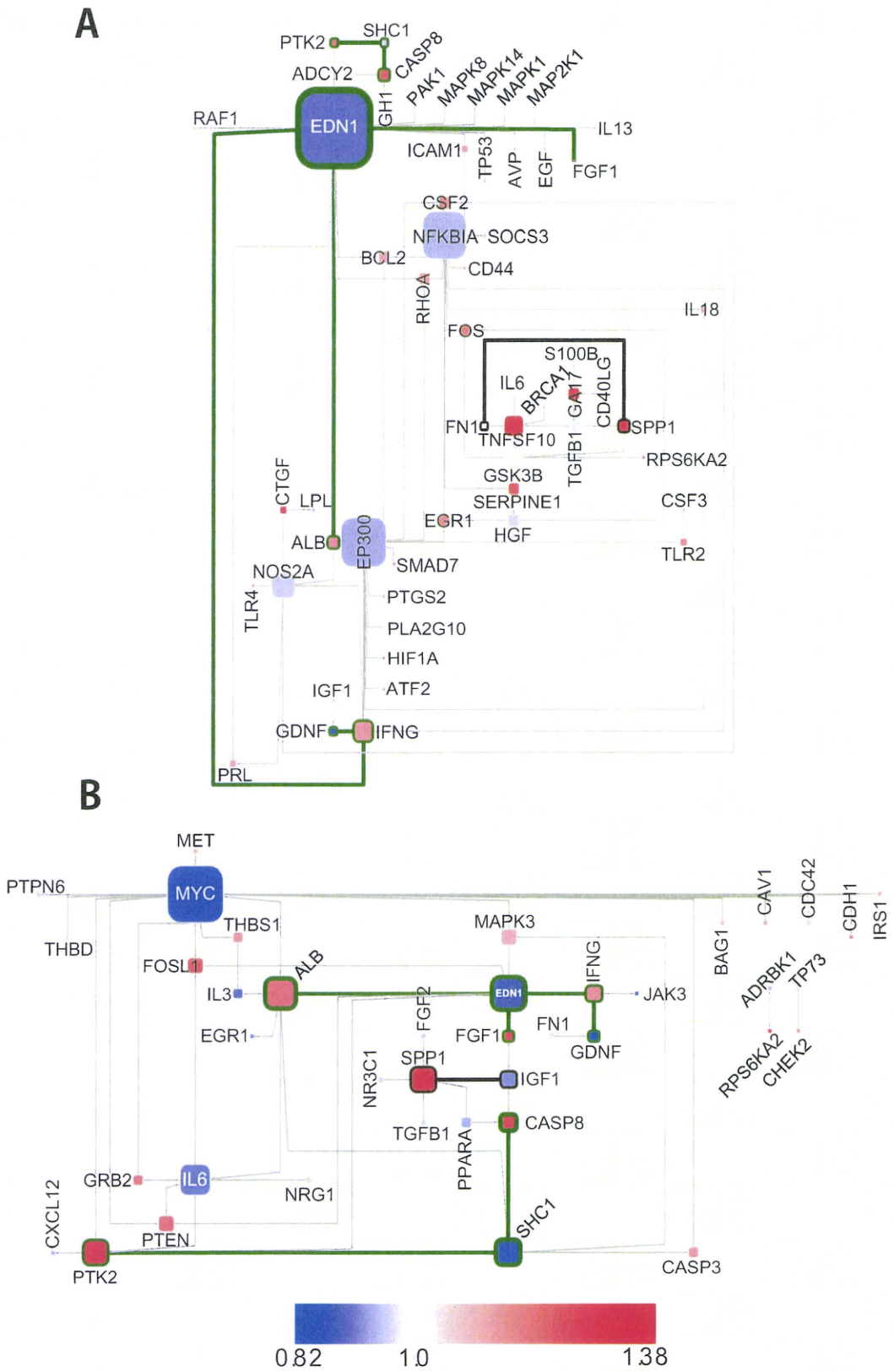

A
PTK2
SHC1
ADCY2
CASP8
PAK1
MAPK8
MAPK14
MAPK1
MAP2K1
GH1
RAF1
EDN1
IL13
ICAM1
TP53
AVP
EGF
FGF1
CSF2
NFKBIA
SOCS3
BCL2
CD44
RHOA
IL18
FOS
S100B
IL6
BRCA1
GAT7
CD40LG
FN1
TNFSF10
TGFB1
SPP1
RPS6KA2
GSK3B
SERPINE1
CSF3
CTGF
LPL
EGR1
HGF
ALB
EP300
TLR2
SMAD7
NOS2A
PTGS2
TLR4
PLA2G10
HIF1A
IGF1
ATF2
GDNF
IFNG
PRL
B
MET
PTPN6
MYC
THBS1
MAPK3
CAV1
CDC42
CDH1
IRS1
THBD
FOSL1
ALB
IFNG
BAG1
IL3
EDN1
JAK3
ADRBK1
TP73
FGF2
EGR1
FGF1
FN1
GDNF
RPS6KA2
CHEK2
NR3C1
SPP1
IGF1
TGFB1
CASP8
PPARA
GRB2
IL6
NRG1
CXCL12
PTEN
SHC1
PTK2
CASP3
0.82
1.0
1.38

Figure 6. Changes in ON/OFF states of BTSs for lung cancer. The state of BTS toggling determined by comparing mRNA expression data (ArrayExpress ID: E-GEOD-10072) for normal lung tissue with that for lung-cancer patients with a history of smoking (former smokers) (Fig. 6A) and that for lung-cancer patients still smoking (current smokers) (Fig. 6B). The nodes and genes surrounded by bold black frames are those also shown in Fig. 5A. The nodes and edges surrounded by bold green frames are found in the former smokers as well as the current smokers. The nodes represent genes and the edges between them represent the pairing of bistable toggle switches. The colors of nodes were automatically assigned as a continuous color gradient from red for ON (upregulated) to blue for OFF (downregulated) according to relative gene-expression levels of the nodes.
doi:10.1371/journal.pcbi.1000851.g006

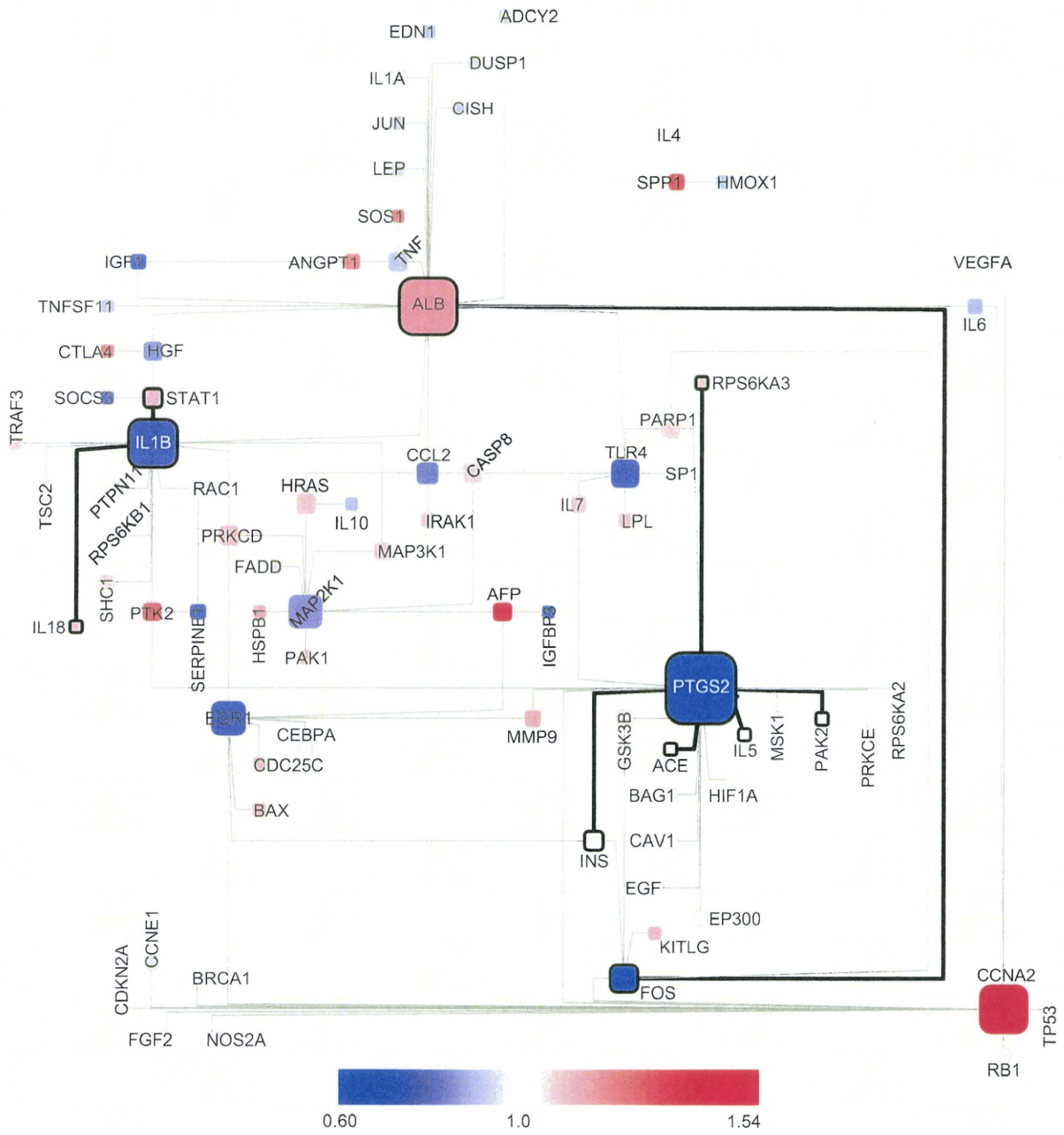


Figure 7. Changes in ON/OFF states of BTSs in dysplastic liver tissue and hepatocellular carcinoma. BTS toggling graph comparing the mRNA expression data (ArrayExpress ID: E-GEOD-6764) of normal liver tissue with that of precancerous and cancerous liver tissue. The nodes and edges surrounded by the bold lines are BTSs for which toggling was observed when comparing dysplastic liver tissue, a precursor of liver cancer, with normal liver tissue. The nodes represent genes and the edges between them represent the pairing of bistable toggle switches. The colors of nodes were automatically assigned as a continuous color gradient from red for ON (upregulated) to blue for OFF (downregulated) according to relative gene-expression levels of the nodes.
doi:10.1371/journal.pcbi.1000851.g007

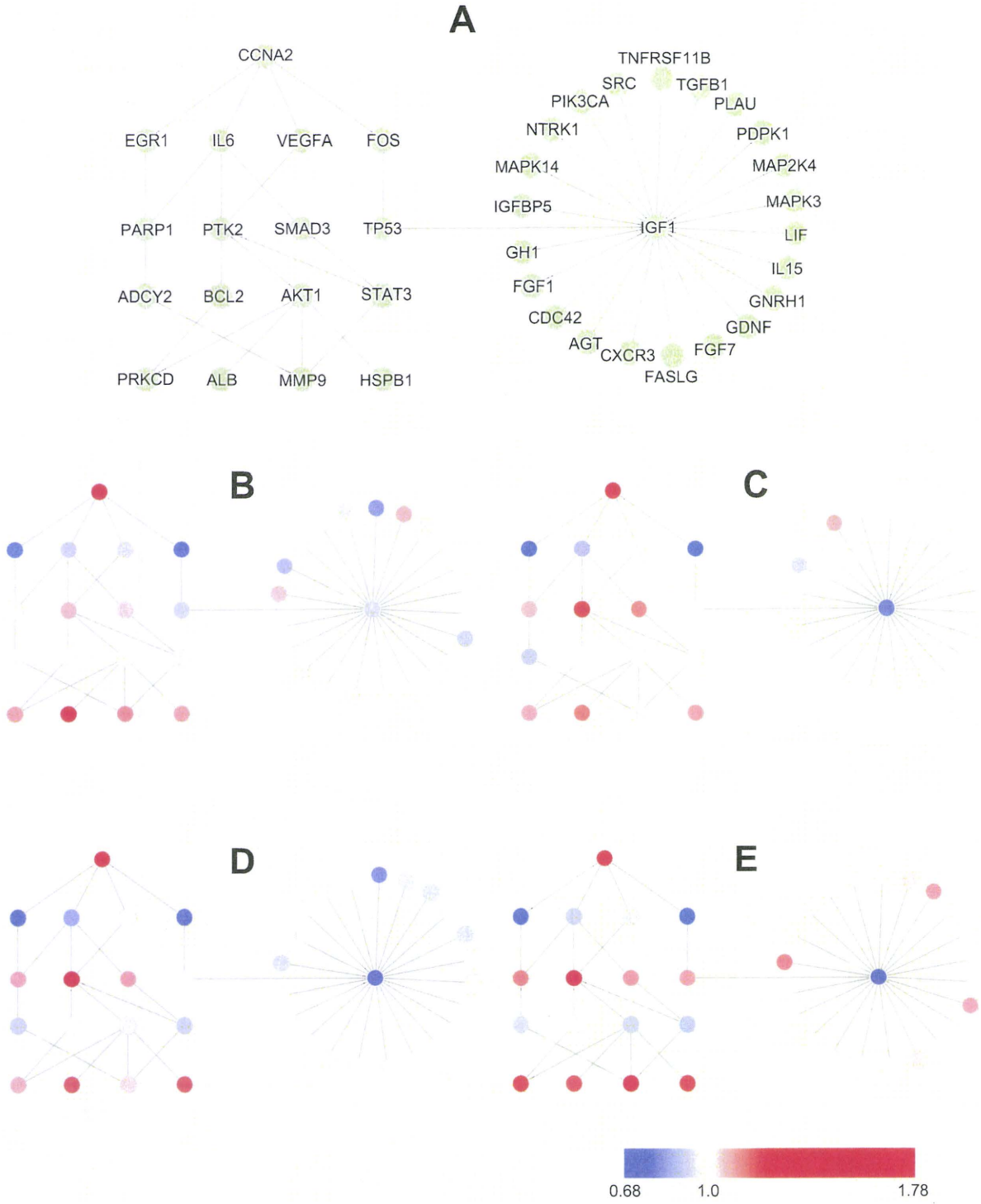


Figure 8. Rippling of toggling of CCNA2-related BTS during malignant transition of HCC. Fig. 8A: A network of CCNA2-related BTS pairs selected from the data used in Fig. 7A. Fig. 8B–E: The nodes represent genes and the edges between them represent the pairing of bistable toggle switches. The colors of nodes were automatically assigned as a continuous color gradient from red for ON (upregulated) to blue for OFF (downregulated) according to relative gene-expression levels of the nodes. B: very early HCC, C: early HCC, D: advanced HCC, E: very advanced HCC. Note that the ON/OFF status of TP53-IGF1 was changed in advanced HCC.
doi:10.1371/journal.pcbi.1000851.g008

expression regulation, direct regulation, molecular transport regulation, and molecular synthesis regulation). MedScan also extracted information on the relation direction and the effect on the target molecule. The "Effect" attribute has three possible values: "positive," "negative," and "unknown." The BTS pairs were extracted from the database on the basis of five rules.

(1) Nodes are limited to genes and proteins only.
(2) Edges are limited to "Regulation," "Expression," and "DirectRegulation."
(3) "Unknown" edges in the "Effect" attribute are omitted.
(4) Edges extracted from fewer than three references are omitted.
(5) If there is a positive and negative attribute in the same direction, the edge is extracted from additional references.

We extracted 19,178 relationships involving 3,682 genes (basic interaction datasets).

Extraction of candidate bistable toggle switches

Using basic interaction datasets, we extracted possible network motifs for toggle switches. We defined these motifs as follows.

The type-1 BTS contains two genes that have positive autoregulation and inhibit each other's expression. The type-2 BTS also contains two genes that suppress each other's expression, but each gene also has a positive or negative loop with the other gene. One of the four subtypes of type-2 BTSs (corresponding to the four possible combinations of double positive and/or negative feedback) shows the same function as the type-1 BTS. The type-3 BTS was based on a theoretical study of the modeling of genetic switches with positive feedback loops [37]. The BTS motifs are illustrated in Fig. 2, and we extracted 6585 BTSs (see supporting Table 1).

Analysis of toggling

We used mRNA microarray data to examine the changes in the ON/OFF states of BTS candidates. CEL format files or tab-limited text files were downloaded via ArrayExpress (http://www.ebi.ac.uk/arrayexpress/), which is a public repository provided by the European Bioinformatics Institute [68]. We only used microarray data obtained from experiments with humans and with platforms of Affymetrix HG-U133A&B (631 sets) and HG-U133Plus2.0 (404 sets). These data were normalized and summarized using the robust multichip analysis method [69] implemented in the Affymetrix Expression Console software.

The toggling of a BTS pair was defined as instances in which the mRNA levels of a sample increased for one molecule of the pair and decreased for the other. To remove background noise, we calculated the toggling score using

$$\text{toggling score} = (\text{SW1 sample value}/\text{SW1 control value}) / (\text{SW2 sample value}/\text{SW2 control value}),$$

where SW1 and SW2 are the two molecules in alphabetical order. Changes in the ON/OFF states were considered significant when the toggling score was more than two standard deviations greater than the mean of all the toggling scores.

Network visualization

For pathway visualization, we used Cytoscape (Version 2.6.3), which is widely used open-source software for visualization and analysis of networks [70]. The nodes in the visualized BTS network represent genes, the edges between nodes represent the pairing of bistable toggle switches, and the color of nodes were automatically assigned as a continuous color gradient from red for ON (upregulated) to blue for OFF (downregulated) according to relative gene-expression levels of the nodes.

Supporting Information

Protocol S1 List of BTS pairs SW1 and SW2 are the two molecules comprising a BTS pair in alphabetical order.
Found at: doi:10.1371/journal.pcbi.1000851.s001 (0.16 MB XLS)

Protocol S2 Cytoscape session file for Figure 3.
Found at: doi:10.1371/journal.pcbi.1000851.s002 (0.09 MB ZIP)

Acknowledgments

We are grateful to Dr. Y. Hamada (Sony Corporation) for preparing the basic interaction datasets, Mr. J. Suzuki (Tokyo Institute of Technology) for assisting us with the data extraction, and Dr. S. Ueda (Otsu Municipal Hospital) for providing us with the microarray data. We also thank Dr. K. Tabuchi for his useful comments and discussion.

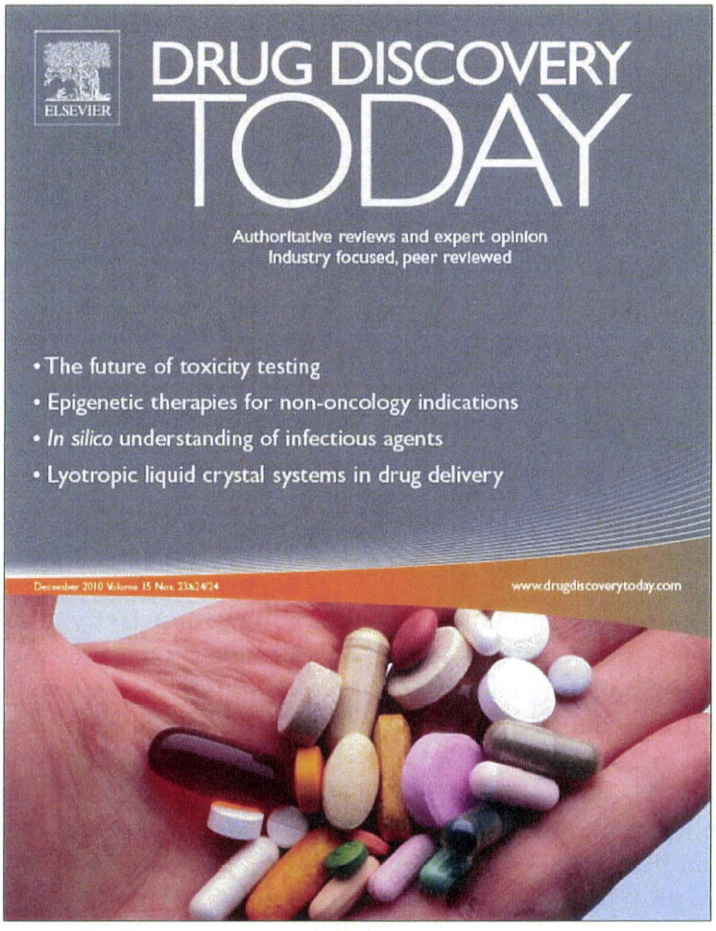
Author Contributions

Conceived and designed the experiments: TS SM HK. Performed the experiments: TS. Analyzed the data: TS SM. Contributed reagents/materials/analysis tools: TS SM. Wrote the paper: TS HK.

References

1. Sontag ED (2007) Monotone and near-monotone biochemical networks. Syst Synth Biol 1: 59–87.
2. Kauffman S, Peterson C, Samuelsson B, Troein C (2004) Genetic networks with canalyzing Boolean rules are always stable. Proc Natl Acad Sci USA 101: 17102–17107.
3. 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.
4. Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet 31: 64–68.
5. Tyson JJ, Chen KC, Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol 15: 221–231.
6. Wolf DM, Arkin AP (2003) Motifs, modules and games in bacteria. Curr Opin Microbiol 6: 125–134.
7. Isaacs FJ, Hasty J, Cantor CR, Collins JJ (2003) Prediction and measurement of an autoregulatory genetic module. Proc Natl Acad Sci USA 100: 7714–7719.
8. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 94: 814–819.
9. Ptashne M (1992) A genetic switch; phage lambda and higher organisms. Blackwell Science.
10. Novick A, Weiner M (1957) Enzyme Induction as an all-or-none phenomenon. Proc Natl Acad Sci USA 43: 553–566.
11. Ozbudak EM, Thattai M, Lim HN, Shraiman BI, Van Oudenaarden A (2004) Multistability in the lactose utilization network of Escherichia coli. Nature 427: 737–740.
12. Vilar JM, Guet CC, Leibler S (2003) Modeling network dynamics: the lac operon, a case study. J Cell Biol 161: 471–476.
13. Bagowski CP, Ferrell JE, Jr. (2001) Bistability in the JNK cascade. Curr Biol 11: 1176–1182.
14. Bhalla US, Ram PT, Iyengar R (2002) MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. Science 297: 1018–1023.
15. Ferrell JE, Jr. (1996) Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends Biochem Sci 21: 460–466.
16. Ferrell JE, Jr. (1997) How responses get more switch-like as you move down a protein kinase cascade. Trends Biochem Sci 22: 288–289.
17. Ferrell JE, Jr., Machleder EM (1998) The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. Science 280: 895–898.
18. Ferrell JE, Xiong W (2001) Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible. Chaos 11: 227–236.

19. Huang CY, Ferrell JE, Jr. (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci USA 93: 10078–10083.
20. Markevich NI, Hoek JB, Kholodenko BN (2004) Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. J Cell Biol 164: 353–359.
21. Lai K, Robertson MJ, Schaffer DV (2004) The sonic hedgehog signaling system as a bistable genetic switch. Biophys J 86: 2748–2757.
22. Cross FR, Archambault V, Miller M, Klovstad M (2002) Testing a mathematical model of the yeast cell cycle. Mol Biol Cell 13: 52–70.
23. Pomerening JR, Sontag ED, Ferrell JE, Jr. (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. Nat Cell Biol 5: 346–351.
24. Sha W, Moore J, Chen K, Lassaletta AD, Yi CS, et al. (2003) Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. Proc Natl Acad Sci USA 100: 975–980.
25. Reynolds AR, Tischer C, Verveer PJ, Rocks O, Bastiaens PI (2003) EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. Nat Cell Biol 5: 447–453.
26. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403: 339–342.
27. Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, et al. (2004) Programmable cells: interfacing natural and engineered gene networks. Proc Natl Acad Sci USA 101: 8414–8419.
28. Kramer BP, Viretta AU, Daoud-El-Baba M, Aubel D, Weber W, et al. (2004) An engineered epigenetic transgene switch in mammalian cells. Nat Biotechnol 22: 867–870.
29. Sabouri-Ghomi M, Ciliberto A, Kar S, Novak B, Tyson JJ (2008) Antagonism and bistability in protein interaction networks. J Theor Biol 250: 209–218.
30. Bagowski CP, Besser J, Frey CR, Ferrell JE, Jr. (2003) The JNK cascade as a biochemical switch in mammalian cells: ultrasensitive and all-or-none responses. Curr Biol 13: 315–320.
31. Laslo P, Spooner CJ, Warmflash A, Lancki DW, Lee HJ, et al. (2006) Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell 126: 755–766.
32. Ferrell JE, Jr. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr Opin Cell Biol 14: 140–148.
33. Hasty J, McMillen D, Isaacs F, Collins JJ (2001) Computational studies of gene regulatory networks: in numero molecular biology. Nat Rev Genet 2: 268–279.
34. Laurent M, Kellershohn N (1999) Multistability: a major means of differentiation and evolution in biological systems. Trends Biochem Sci 24: 418–422.
35. Smolen P, Baxter DA, Byrne JH (2000) Mathematical modeling of gene networks. Neuron 26: 567–580.
36. Guantes R, Poyatos JF (2008) Multistable decision switches for flexible control of epigenetic differentiation. PLoS Comput Biol 4: e1000235.
37. Kobayashi T, Chen L, Aihara K (2003) Modeling genetic switches with positive feedback loops. J Theor Biol 221: 379–399.
38. Cao Y, Liang J (2008) Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. BMC Syst Biol 2: 30.
39. Chatterjee A, Kaznessis YN, Hu WS (2008) Tweaking biological switches through a better understanding of bistability behavior. Curr Opin Biotechnol 19: 475–481.
40. Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S (2008) Generation of mouse induced pluripotent stem cells without viral vectors. Science 322: 949–953.
41. Boyer LA, Mathur D, Jaenisch R (2006) Molecular control of pluripotency. Curr Opin Genet Dev 16: 455–462.
42. Chickarmane V, Troein C, Nuber UA, Sauro HM, Peterson C (2006) Transcriptional dynamics of the embryonic stem cell switch. PLoS Comput Biol 2: e123.
43. Niwa H (2007) How is pluripotency determined and maintained? Development 134: 635–646.
44. Chari R, Lonergan KM, Ng RT, MacAulay C, Lam WL, et al. (2007) Effect of active smoking on the human bronchial epithelium transcriptome. BMC Genomics 8: 297.
45. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, et al. (2008) Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics 9: 259.
46. Gao H, Ward PA (2007) STAT3 and suppressor of cytokine signaling 3: potential targets in lung inflammatory responses. Expert Opin Ther Targets 11: 869–880.
47. He B, You L, Uematsu K, Zang K, Xu Z, et al. (2003) SOCS-3 is frequently silenced by hypermethylation and suppresses cell growth in human lung cancer. Proc Natl Acad Sci USA 100: 14133–14138.
48. Boldrini L, Donati V, Dell'Omodarme M, Prati MC, Faviana P, et al. (2005) Prognostic significance of osteopontin expression in early-stage non-small-cell lung cancer. Br J Cancer 93: 453–457.
49. Boldrini L, Gisfredi S, Ursino S, Faviana P, Lucchi M, et al. (2005) Expression of endothelin-1 is related to poor prognosis in non-small cell lung carcinoma. Eur J Cancer 41: 2828–2835.
50. Davis GL, Dempster J, Meler JD, Orr DW, Walberg MW, et al. (2008) Hepatocellular carcinoma: management of an increasingly common problem. Proc (Bayl Univ Med Cent) 21: 266–280.
51. Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, et al. (2007) Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. Hepatology 45: 938–947.
52. Bae SH, Jung ES, Park YM, Kim BS, Kim BK, et al. (2001) Expression of cyclooxygenase-2 (COX-2) in hepatocellular carcinoma and growth inhibition of hepatoma cell lines by a COX-2 inhibitor, NS-398. Clin Cancer Res 7: 1410–1418.
53. Baek JY, Hur W, Wang JS, Bae SH, Yoon SK (2007) Selective COX-2 inhibitor, NS-398, suppresses cellular proliferation in human hepatocellular carcinoma cell lines via cell cycle arrest. World J Gastroenterol 13: 1175–1181.
54. Wheeler LW, Lents NH, Baldassare JJ (2008) Cyclin A-CDK activity during G1 phase impairs MCM chromatin loading and inhibits DNA synthesis in mammalian cells. Cell Cycle 7: 2179–2188.
55. Krones-Herzig A, Mittal S, Yule K, Liang H, English C, et al. (2005) Early growth response 1 acts as a tumor suppressor in vivo and in vitro via regulation of p53. Cancer Res 65: 5133–5143.
56. Hao MW, Liang YR, Liu YF, Liu L, Wu MY, et al. (2002) Transcription factor EGR-1 inhibits growth of hepatocellular carcinoma and esophageal carcinoma cell lines. World J Gastroenterol 8: 203–207.
57. Elsammak MY, Amin GM, Khalil GM, Ragab WS, Abaza MM (2006) Possible contribution of serum activin A and IGF-1 in the development of hepatocellular carcinoma in Egyptian patients suffering from combined hepatitis C virus infection and hepatic schistosomiasis. Clin Biochem 39: 623–629.
58. Qiao L, Nachbar RB, Kevrekidis IG, Shvartsman SY (2007) Bistability and oscillations in the Huang-Ferrell model of MAPK signaling. PLoS Comput Biol 3: 1819–1826.
59. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123: 917–929.
60. Wang ES, Szabo SJ, Schwartberg PL, Glimcher LH (2005) T helper cell fate specified by kinase-mediated interaction of T-bet with GATA-3. Science 307: 430–433.
61. Schwarz M, Cecconi F, Bernier G, Andrejewski N, Kammandel B, et al. (2000) Spatial specification of mammalian eye territories by reciprocal transcriptional repression of Pax2 and Pax6. Development 127: 4325–4334.
62. Asyali MH, Shoukri MM, Demirkaya O, Khabar KS (2004) Assessment of reliability of microarray data and estimation of signal thresholds using mixture modeling. Nucleic Acids Res 32: 2323–2335.
63. Hinkley DV (1969) On the ratio of two correlated normal random variables. Biometrika 56: 635–639.
64. Brody JP, Williams BA, Wold BJ, Quake SR (2002) Significance and statistical errors in the analysis of DNA microarray data. Proc Natl Acad Sci USA 99: 12975–12978.
65. Cochella L, Green R (2005) Fidelity in protein synthesis. Curr Biol 15: R536–540.
66. Xu Y, Chen SY, Ross KN, Balk SP (2006) Androgens induce prostate cancer cell proliferation through mammalian target of rapamycin activation and post-transcriptional increases in cyclin D proteins. Cancer Res 66: 7783–7792.
67. Novichkova S, Egorov S, Daraselia N (2003) MedScan, a natural language processing engine for MEDLINE abstracts. Bioinformatics 19: 1699–1706.
68. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, et al. (2007) ArrayExpress—a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 35: D747–750.
69. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264.
70. 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.



Connecting the dots: role of standardization and technology sharing in biological simulation

Samik Ghosh[1], Yukiko Matsuoka[1,2] and Hiroaki Kitano[1,3,4]

[1] The Systems Biology Institute, Tokyo 108-0071 Japan
[2] JST ERATO Kawaoka Infection-induced Host-response Network Project, Tokyo 108-8639 Japan
[3] Okinawa Institute of Science and Technology, Okinawa, Japan
[4] Sony Computer Science Laboratories, Tokyo 141-0022 Japan

The role of biological modeling and simulation in enhancing productivity across the drug discovery pipeline has been increasingly appreciated over the past decade by the pharmaceutical industry. However, adoption of *in silico* modeling and simulation techniques has been sparse due to skepticism in the associated pay-offs and knowledge gap in research. While biological simulations have been successfully applied in specific projects, a standardized, community-wide platform is imperative for making the final leap of faith across the domain. This review outlines the issues and challenges involved in fostering a private-public collaborative effort for the development of standard modeling and biosimulation platforms and concludes with insights into possible mechanisms for integrating an *in silico* pipeline into the drug discovery and development process.

In the spring of 2008, a group of scientists and thought leaders from across academia and industry met in Tokyo [1] to brainstorm future challenges in systems biology and its application to the pharmaceutical industry. While unanimously acknowledging the burgeoning role of systems biology in tackling complex diseases, the researchers laid out a bold objective – 'to create over the next 30 years a comprehensive, molecule-based, multi-scale, computational model of the human ('the virtual human'), capable of simulating and predicting, with a reasonable degree of accuracy, the consequences of most of the perturbations that are relevant to healthcare' [2].

The scale and timeline of the project outlined in the Tokyo Declaration [2] underscore the complexity of the problem facing researchers in life sciences and pharmaceutical companies alike. With the current economic scenario, value of impending major blockbuster drug patent expirations and thinning pipelines for anticipated phase III drug approvals [3,4], the pharma industry is undergoing a paradigm shift in its efforts to develop safe and more efficacious drugs for the complex and life-threatening diseases facing humankind. Recent industry reports on future visions for the industry [3,4] have highlighted the need to shift trajectories from target-based drug discovery to more system-oriented, holistic approaches that embrace knowledge from basic academic research and computational and systems engineering techniques. In particular, predictive biosimulation software – based on interconnected sets of mathematical equations with calibrated parameters to represent biological and physiological behaviors – has been successfully applied across the different stages of drug discovery and development, from target identification and validation, lead optimization and candidate selection to clinical trial design and development [5–7].

Simulation is an indispensable tool in all engineering designs and has been successfully applied in the automobile, aerospace and telecommunication industries for decades. Computational fluid dynamics (CFD), for example, is an essential design process in aircraft design, ship design and automobile design. Any high-rise building has to carry out a series of structural integrity simulations even to be approved for construction; chipmakers model, modify and simulate their designs on computers before sending them to the fabrication plants; 'virtual cars' are driven and 'virtual aircrafts' flown under simulated conditions before hitting the manufacturing floor [8]. Although the application of advanced

Corresponding author:. Kitano, H. (kitano@sbi.jp)

1359-6446/06/$ - see front matter © 2010 Published by Elsevier Ltd. doi:10.1016/j.drudis.2010.10.001

modeling and simulation techniques has resulted in immense cost savings and standardized procedures for such R&D-intensive industries, the pharmaceutical industry has historically lacked these approaches, leading to astronomical costs in drug development (~25% of its revenue, almost twice that of other knowledge-driven industries [8]).

Although appreciation and awareness for the potential benefits of computational approaches in biological sciences and drug design have been on an increasing trajectory in both industry and academic circles, it is important to keep in perspective the unique hurdles and cosiderable challenges of applying *in silico* techniques in the life sciences. Identification of the specific features desired from computational tools in the pharmaceutical industry, together with an open, collaborative mindset between all players, would form the key stepping stones in the development of safer, efficacious and cost-efficient drugs for complex diseases such as cancer, metabolic and cardiac disorders.

Issues and challenges

The adoption of simulation techniques in the life sciences requires careful and detailed consideration of the unique challenges of multi-scale modeling– from cells, to tissues and organs, to whole human body and host–pathogen interactions. A series of issues have to be addressed before simulation can be accepted as normal practice in the industry.

First, a set of fundamental technical issues must be solved to further improve accuracy of simulation. Different flavors of simulation technologies exist, from deterministic, differential-equation-based systems to non-deterministic, stochastic techniques, agent-based and discrete-event based simulations; each presents a unique set of assumptions and system conditions that need to be considered before successful application to specific biological problems, as elucidated schematically in Fig. 1 [9]. Cellular modeling or physiological modeling with molecular details will require the integration of heterogeneous computational models that are on different spatial and temporal scales, and the basic equations still need to be defined [10].

The purpose and goal of a simulation system applied in drug design should be clearly defined: for example, in Formula 1 aerodynamics design, the goal is to design an aerodynamically optimal body with maximum down force with minimum drag. This forms a key step in defining and determining the eventual success of a biological simulation system. Merrimack Pharmaceuticals [11], for example, used computer simulation to identify a novel drug target for specific cancer sub-types that resulted in the development of a monoclonal antibody for ErbB3, now in clinical trial. Simulation models need to be designed to sufficiently capture essential features to accomplish the task defined, but features that are unlikely to affect prediction accuracy of the given task can be ignored.

Sophisticated models with molecular details that can predict cellular behaviors in various conditions are crucial for elucidating system-level properties of cellular systems. Such models should be able to provide predictions on how cells and organs respond when certain perturbations, such as drug administration, are given. Although there are some successful cases of computational modeling of limited-scale biological networks, there is no established method for developing high-precision models.

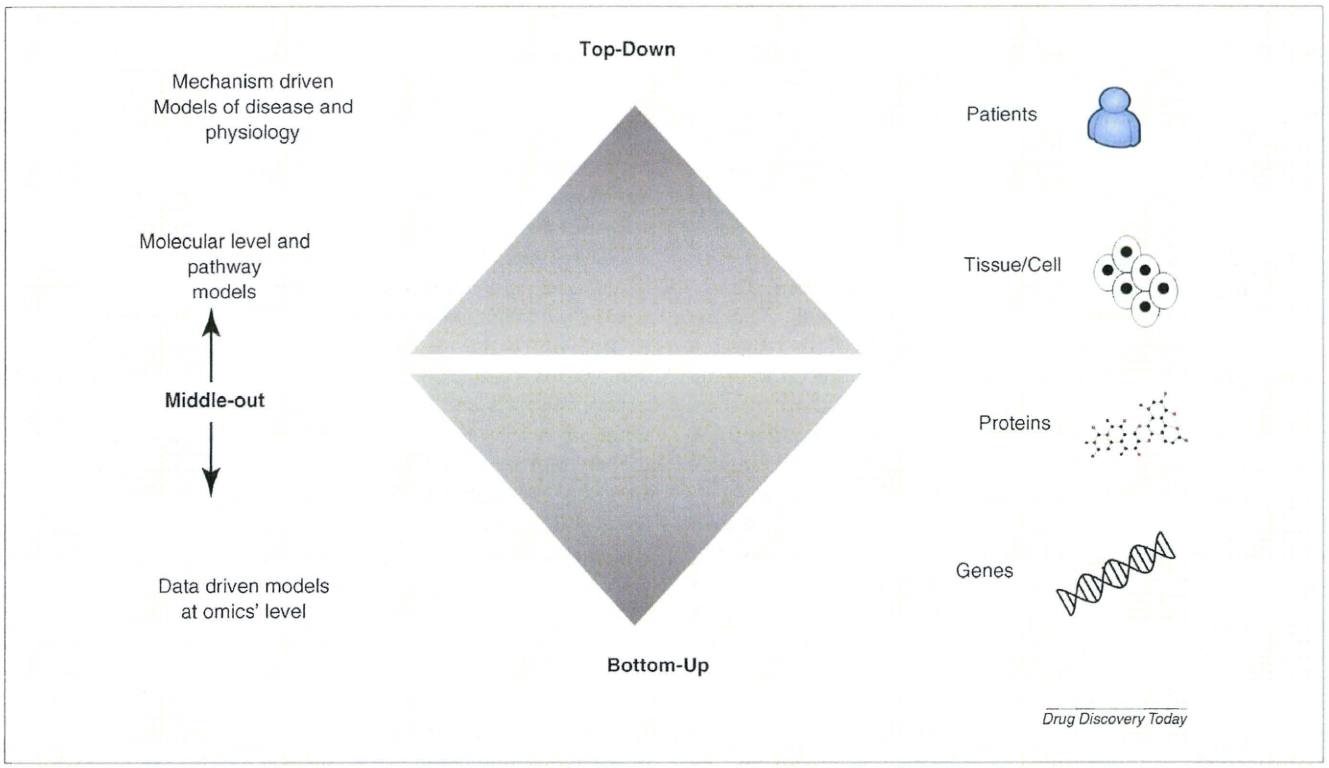


FIGURE 1

Schematic showing different simulation and modeling approaches.

In establishing such practices, it might be prudent to borrow ideas from other fields with more mature and well-established practices of simulation modeling. In particular, the integration of computational modeling and experimental data acquisition has to be promoted. Delving into the design process of Formula 1 racing cars provides a picture of an iterative design cycle – several designs are tested with CFD and some of these designs are then tested using a wind tunnel, leading to the selection of one or two designs that are actually implemented and tested in a test course before one design is selected for final production. In this process, CFD models are calibrated against wind tunnel data for further improvement of accuracy, instead of data from the test course or from actual racing telemetry data.

This comparison delivers two messages. First, we need to develop highly controllable experimental systems comparable to wind tunnels in aerodynamics. This means that we need to be able to precisely control exposure to chemical substances and other environmental conditions. Second, efforts need to be made to create high-precision models against well-controlled experimental systems, instead of uncontrollable systems. The identification and integration of structural, spatial and temporal dynamics of both interaction networks and cellular structures is an essential prerequisite for defining such high-precision models. The dynamics of cellular structure and interaction networks need to be quantified by taking comprehensive, high-resolution, measurements of intracellular status – such as the concentrations, interactions, modifications and localizations of molecules – and of cellular structures in different dimensions and environmental conditions. In addition, the problem of how to identify unknown interactions from such data sets still remains.

These problems are fundamental and require collaborative efforts on a community-wide scale. Although various ongoing efforts exist for tackling the problems of building molecular network maps, simulation tools, data resources and web services for sharing information, an open, integrative platform for sharing and exchange of computational approaches is of fundamental importance.

Standards and technologies

The sharing of knowledge accelerates progress in science. This is perhaps more true for the biological sciences, in which the complexity and size of the problem domain make it imperative for different research groups to focus on different sections of it. Computational tools have already played an important part in the collection, storage and intelligent retrieval of vast amounts of information in the life sciences. With the adoption of biosimulation tools, the ability to store and share information in a seamless, unambiguous fashion became imperative, leading to the definition of The Systems Biology Markup Language (SBML; http://www.sbml.org), a set of standards developed to facilitate effective and efficient sharing of models defined as a set of biochemical reactions. The effort was initiated by the ERATO Kitano Symbiotic Systems Project, funded by the Japanese government but soon grew to be a global community-wide initiative. Importantly, the SBML community has evolved a proper procedure to elect editors, implemented voting procedures and formalized discussion forums. Thus, it is no longer a project belonging to any one institution but is truly a community effort. Currently, more than 180 pieces of bio-modeling and simulation software comply with SBML [12], enabling the sharing of models across them.

Although the definition of a common *lingua franca* is an important step in the standardization of *in silico* technologies, biology has traditionally been a descriptive science, in which the role of pictures and diagrams cannot be overstated. Keeping in mind the need for a common graphical representation standard in the life sciences, a community-wide effort has been undertaken to define The Systems Biology Graphical Notation (SBGN; http://www.sbgn.org/) [13]. The SBGN community is working to formulate a set of rules for human-readable visual representations of biological networks. The goal of SBGN is to define a set of visual glyphs and syntax so that anyone can understand exactly what each diagram means (Fig. 2), in the same vein as the electrical circuit diagrams used by chip designers.

With SBML and SBGN, standards for computational modeling and representation languages are being defined that will have a notable impact on standardization and model sharing. For models to be informative, they must be properly annotated; sufficient information must be attached with the models to enable third parties to use them. With the objective of proper model annotation, the MIRIAM project (http://www.ebi.ac.uk/miriam/) defines the minimum information that has to be attached to a model so that model can be informative by itself. A series of standard formation efforts are now underway to cover the whole process of modeling development and analysis.

Although standardization is an integral part of the process of computational systems biology, the development of a suite of software tools for model building, distribution and running simulations is another important dimension. In this direction, a plethora of modeling building and simulation tools such as Cellarator [14], Copasi [15] and Dizzy [16] in the academic community and SimBiology® from Mathworks Inc. (http://www.mathworks.com/products/simbiology) and PhysioLab® (Entelos Inc; http://www.entelos.com) exist, each catering to different modeling techniques (see Ref. [12] for comprehensive coverage of systems biology tools).

One of the most popular and widely used tools in this space [12] is CellDesigner™ [17] – a modeling and simulation tool to visualize, model, and simulate gene regulatory and biochemical networks. Two major characteristics embedded in CellDesigner boost its usability to create, import and export models: the solidly defined and comprehensive graphical representation (SBGN) of network models and SBML as a model-describing basis, which function as inter-tool media to import and export SBML-based models. Moreover, CellDesigner provides the ability to embed – or smoothly connect via Systems Biology Workbench (http://www.sys-bio.org/research/sbwIntro.htm) – different simulation and analysis packages, which enable the simulation of the pathways using various simulation techniques (Copasi, SBML ODE solver, and so on), as shown in Fig. 3a.

With the explosive growth in proliferation and adoption of the Internet and web services, knowledge in the life sciences has shifted to the World Wide Web with online access to scientific literature, biological databases, and knowledge-sharing and/or discussion forums. As more biological information comes online, it is important to develop an infrastructure that enables the community to leverage the vast web resources. This online, com-

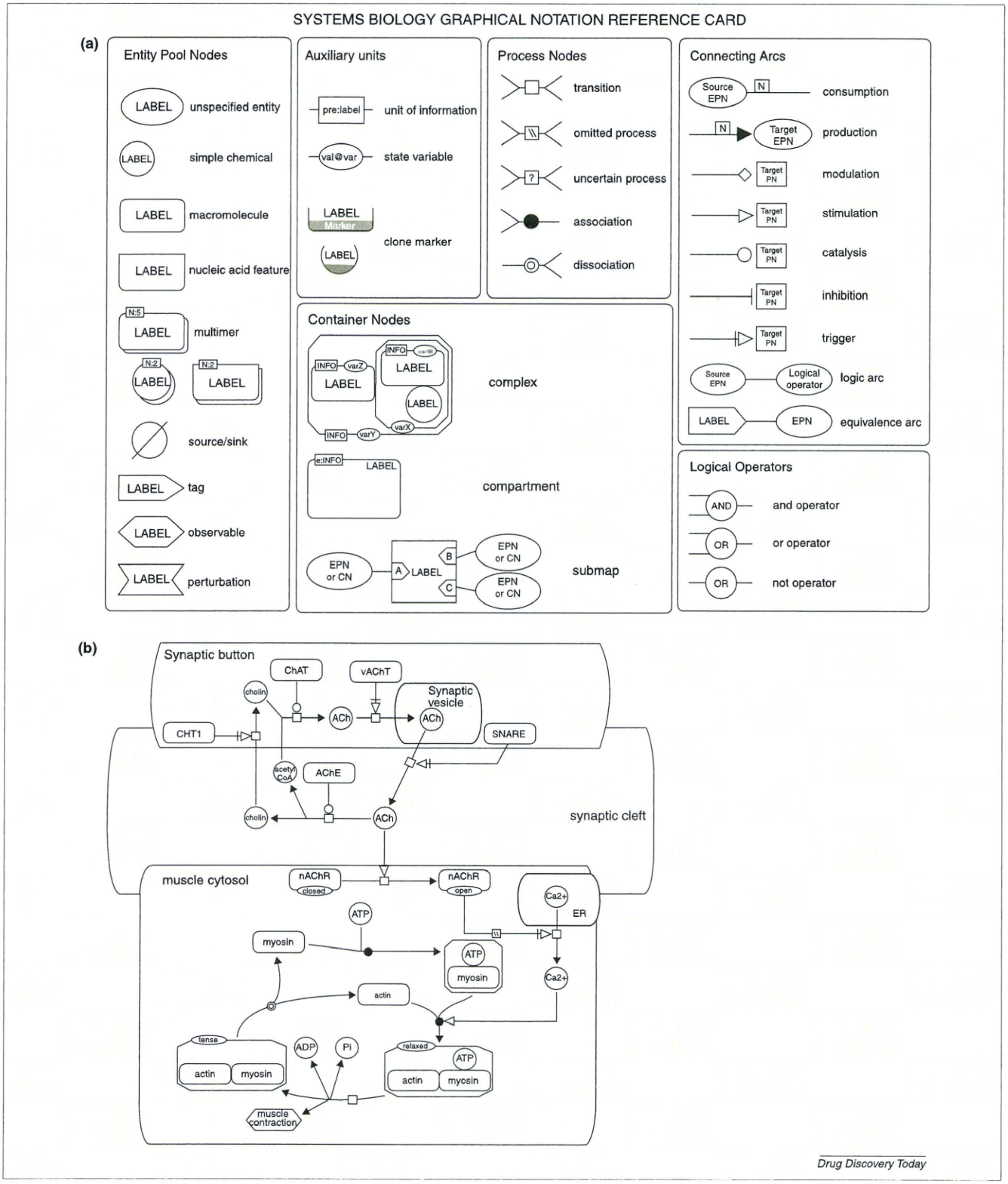


FIGURE 2

(a) The Systems Biology Graphical Notation (SBGN) glyphs and **(b)** sample model representation.

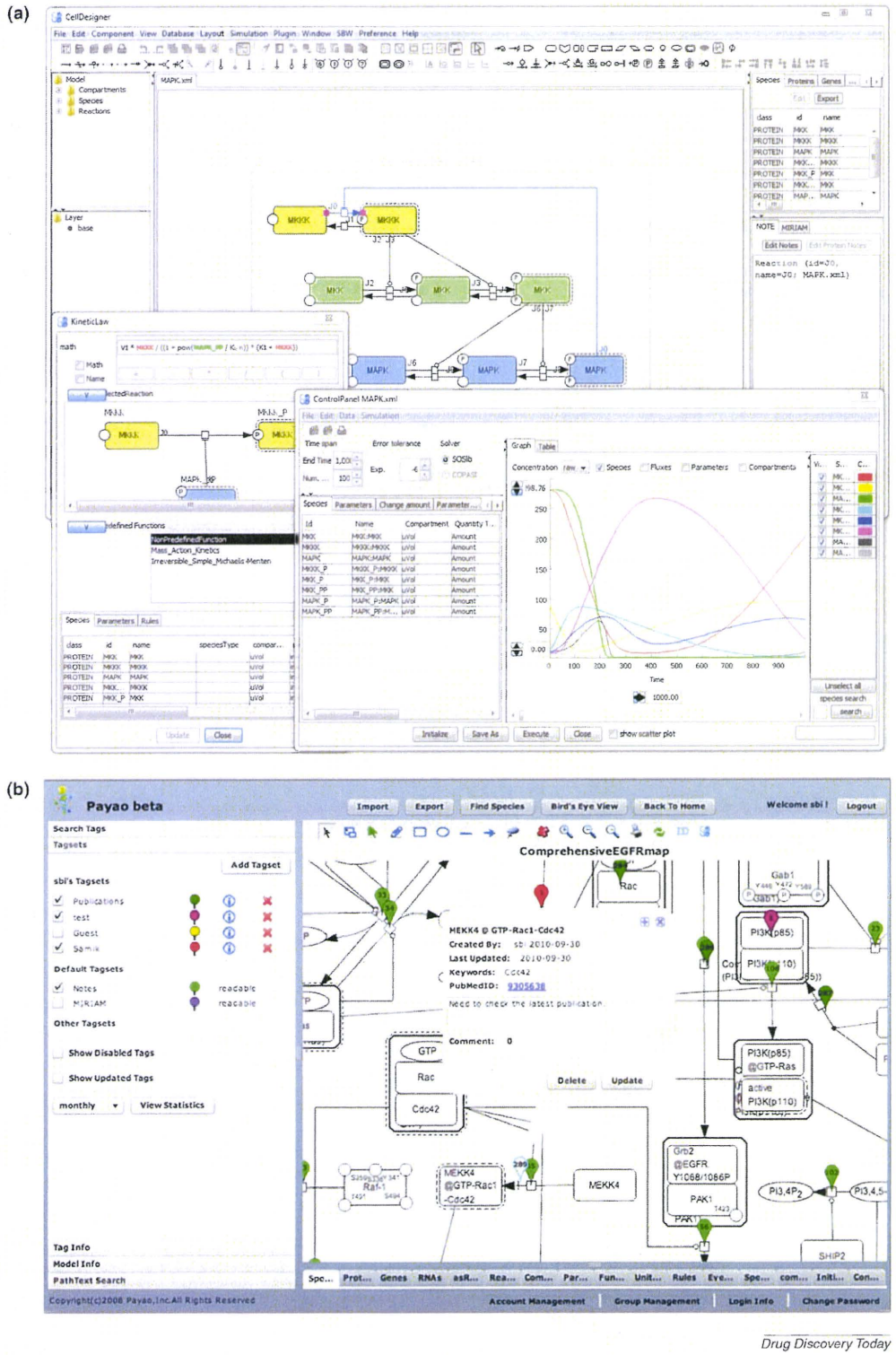


FIGURE 3

(a) CellDesigner snapshot and **(b)** Payao web interface snapshot.

munity collaboration paradigm motivated the development of Payao [18], a web-based biological pathway sharing and tagging service (http://www.payaologue.org). A snapshot of the interface is shown in Fig. 3b. The goal of Payao is to provide a Google Maps equivalent for biological pathways, wherein researchers can share large-scale, curated and annotated (MIRIAM-compliant) network maps (SBGN and SBML compliant) using software such as CellDesigner and publish it to the community online. With the built-in tagging and collaborative system, the community can participate in enhancing the biological entities in the map or navigate their specific areas of interest.

Other important assets for online community collaboration are the databases that provide pathway models fully curated and compliant with standards. Two major online data resources providing access to biological pathway models are the BioModels.net Initiative (http://www.biomodels.net), which provides a one-stop shop for SBML-compliant simulation models of biological pathways published in various literature, and The Panther Pathways database [19], which provides a database of annotated pathways created in CellDesigner (SBML and SBGN compliant).

We have provided a slice of the spectrum of activities initiated by different research groups in developing biosimulation tools and technologies. Whereas each of these tools provides a niche solution to a particular biological problem in isolation, an integrative framework leveraging the advantages of the diverse techniques holds the promise of providing a comprehensive computational pipeline.

Community collaboration and open flow model

The standards and technologies in systems biology, as elucidated in the previous section, represent various pieces of a comprehensive computational pipeline for the pharmaceutical industry. Crucial to the success of such a pipeline, however, is the clarification of the promises and pitfalls of the different components and the identification of a cohesive set of strategies to add value to the drug discovery process.

In June 2008, a representative group of systems biology scientists from academia, biotechnology and the pharmaceutical sector gathered in Portofino, Italy [20], with the goal of brainstorming a set of recommendations for such a strategic path to systems biology. While identifying various action items like setting data standards, modeling drug actions and toxicity, the leaders proclaimed that a 'network solution' [20] (i.e. community-wide collaboration, communication and outreach) was the key to solving complex biological network problems.

As outlined in the 'Issues and challenges' section, the size and complexity of living systems present several challenges to the systems biology community. Overcoming the obstacles of simulation speed, accuracy, high-precision experiments and knowledge sharing would require the expertise of scientists and engineers from diverse backgrounds and disciplines. Thus, developing an ecosystem of communication and information sharing empowered by powerful computational simulation tools and web services can provide the right plan for advancing biosimulation approaches in the pharmaceutical domain.

Concomitant with the development of a community ecosystem is the need for an open flow model of research, in which researchers from industry and academia share knowledge in a collective commons of data, analytical tools, information technology, biospecimens and disease models [21]. The past decade has seen several efforts in open-source biomedical science – the Cancer Genome Atlas (http://www.cancergenome.nih.gov/), Pathway Commons (http://www.pathwaycommons.org) and World Community Grid (http://www.worldcommunitygrid.org). Efforts for establishing collaboration and outreach through joint projects have also been launched in North America and Europe, including the Alliance for Cellular Signaling by the NIH (http://www.afcs.org) and the HepatoSys project (http://www.hepatosys.de) in Germany. The Open Source Drug Discovery project (http://www.osdd.net), initiated by the Council of Scientific and Industrial Research, India, is a uniquely novel effort dedicated to developing drugs for neglected tropical diseases (*Mycobacterium tuberculosis*) that plague developing countries and have very thin pipelines (in terms of the number of candidate drug compounds in development) in the pharmaceutical industries.

An open innovation strategy has also gained traction recently in the pharmaceutical industry, fueled by a need to reconfigure and streamline the drug discovery pipeline, particularly in the early phases of target biology and biomarker development [22]. An industry–academic consortium involving researchers from The University of California at San Diego, the California Institute of Technology, the Massachusetts Institute of Technology, the University of Massachusetts, Entelos, and several research groups from Pfizer has initiated a collaborative project on insulin-resistive pathways [23]. More recently, in May 2010, London-based GlaxoSmithKline and Novartis deposited over 300,000 structures of chemical compounds active against the malaria parasite *Plasmodium falciparum* in an open database archive, ChEMBL Neglected Tropical Disease from the European Bioinformatics Institute (http://www.ebi.ac.uk/chemblntd) [22].

Several socioeconomic obstacles to leveraging the power of an open-source approach exist, however – participant willingness to share data, incompatibility of formats, quality control of data, intellectual property conflicts and sustainable funding avenues. To develop a generic model for collaboration, it is necessary to develop a framework that is open, providing standardized, license-free access to biological data; integrative, providing common sets of pathway curation, annotation, modeling, analysis and simulation tools; and 'share-and-care', an incentive system for providing pre-competitive, early access to the results and data to participants.

An open, integrative, share-and-care model

The goal of an open, integrative, share-and-care framework is to provide a common set of tools, principles and practices for the application of biosimulation techniques in a systematic and cohesive manner. It would involve a standardization platform that enables the incorporation of standards and interfaces for the systems biology community; a computational bio-networking platform that provides a suite of network building tools, databases and web services for sharing information in a standards-compliant, interoperable manner; and an advanced simulation platform that incorporates the different simulation tools and technologies and is capable of encompassing information from the standard compliant biological network resources and databases. A schematic representation of such a scheme is shown in Fig. 4.

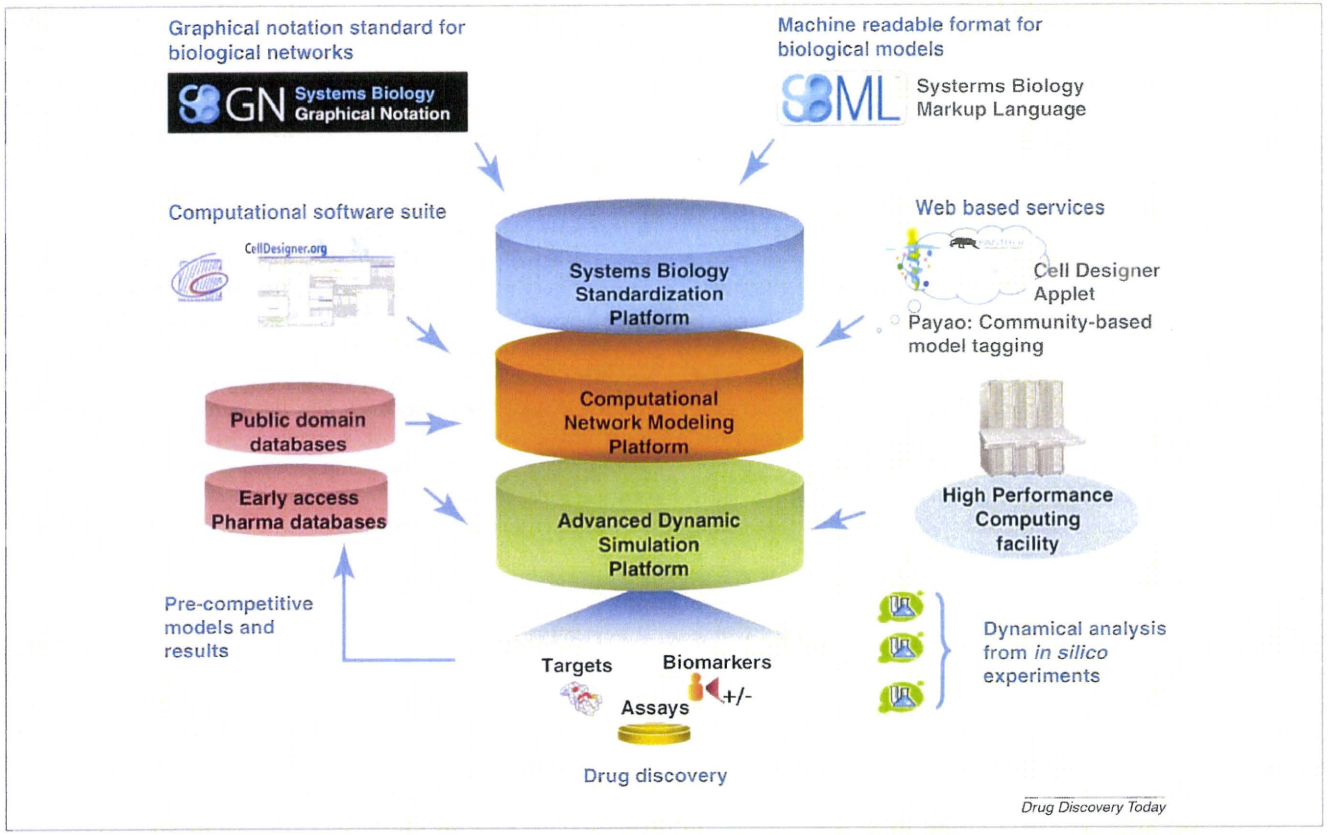


FIGURE 4

An open, integrative, share-and-care model for systems biology.

Although the technical components of such a framework are all or mostly in place, the success of the model hinges on being able to foster a bi-directional flow of knowledge between academia and industry. An open access paradigm for information and data sharing would enable the fast and effective dissemination of knowledge and the identification and filling in of information gaps in crucial biological processes and pathways.

A standard-driven, integrative framework would enable a plug-and-play model for diverse simulation and modeling tools, providing researchers with the freedom to test various tools and techniques.

The most important part, perhaps, is ensuring the flow of information between diverse research groups. In this respect, a share-and-care model would provide incentives to all participants: academic researchers would gain access to specific data on drugs and clinical trials (e.g. access to failed drugs from phase II trial databases [24]), enabling them to develop models with higher predictive powers, whereas the business imperatives of pharmaceutical companies would be respected through early, pre-competitive access to model predictions and insights.

Such a system would work in a positive feedback loop, because high-precision data from the drug industry would drive the precision of biosimulation models, which, in turn, would be able to provide better decision-making conditions for the drug companies. In a recent development, a Boston, MA based startup called EnlightBio (http://www.enlightbio.com), working on similar paradigms, started operations focused on developing breakthrough innovations through partnership with multiple drug companies. Another effort initiated in early 2010 is Sage Bionetworks (http://www.sagebase.org), a non-profit, open-source research organization that aims to develop and share large-scale network models of diseases. Such initiatives would foster collaborative research while paving the way for systems biology to empower the life sciences industry.

Future perspectives

The value of applying systems biology techniques to the different stages of the drug design and development process has been demonstrated through various academic–industry collaborations spanning projects of different scales and sizes. Although such isolated success stories further motivate the large-scale adoption of *in silico* biosimulation practices, the development of a structured and coherent route to facilitate and accelerate the process is imperative. As outlined in this review, collaboration is the key to achieving the goal – not only in terms of joint projects but also – in developing a suite of standards-compliant platforms for sharing domain-specific knowledge and expertise. It might not be a distant dream that the community will develop a set of recommendations for good simulation practices (GSP), as already exists for good manufacturing practices (GMP) and good engineering practices (GEP) – standards and protocols accepted and complied with in the

application of computational simulation tools for drug design and development.

Acknowledgements

This research is, in part, supported by the ERATO-SORST Program of the Japan Science and Technology Agency (JST); the Genome Network Project of the Ministry of Education, Culture, Sports, Science, and Technology; the NEDO Fund for International Standard Formation from the New Energy Development Organization; and the Okinawa Institute of Science and Technology.

References

1 Editorial, (2008) All systems go. *Nat. Rev. Drug Discov.* 7, 278–279
2 FCSB, (2008) *The Tokyo Declaration, International Workshop on Future Challenges for Systems Biology*. http://www.systems-biology.org/~myukiko/FCSB2008/doku.php?id=workshop:statement
3 PriceWaterhouseCoopers, (2008) *Pharma 2020: Virtual R7D – which path will you take?* http://www.pwc.com/extweb/pwcpublications.nsf/docid/91BF330647FFA402852572F2005ECC22
4 Futurology, P. (2007) *Joined-up healthcare, 2016 and beyond.* British Telecommunications http://www2.bt.com/static/i/media/pdf/BT_Pharma_Lowres.pdf
5 Kansal, A.R. and Trimmer, J. (2005) Application of predictive biosimulation within pharmaceutical clinical development: examples of significance for translational medicine and clinical trial design. *Syst. Biol. (Stevenage)* 152, 214–220
6 Rullmann, J.A.C. *et al.* (2005) Systems biology for battling rheumatoid arthritis: application of the Entelos PhysioLab platform. *Syst. Biol. (Stevenage)* 152, 256–262
7 Sams-Dodd, F. (2006) Drug discovery: selecting the optimal approach. *Drug. Discov. Today* 11, 465–472
8 No authors listed. (2005) Models that take drugs. *The Economist* June, S23–S24.
9 Michelson, S. and Scherrer, D. (2003) Predictive biosimulation for lead optimization. *Curr. Drug Discov.* 1, 18–22
10 Kitano, H. (2010) Grand challenges in systems physiology. *Frontiers in Physiology.* doi:10.3389/fphys.2010.00003.
11 No author listed. (2008) Merrimack Pharmaceuticals initiates enrollment in a phase 1 study of MM-121, an ErbB3 antagonist. *Medical News Today* August.
12 Klipp, E. *et al.* (2007) Systems biology standards – the community speaks. *Nat. Biotechnol.* 25, 390–391
13 Le Novère, N. *et al.* (2009) The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741
14 Shapiro, B.E. *et al.* (2003) Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics* 19, 677–678
15 Hoops, S. *et al.* (2006) COPASI – a complex pathway simulator. *Bioinformatics* 22, 3067–3074
16 Ramsey, S. (2005) Dizzy: stochastic simulation of large-scale genetic regulatory networks. *J. Bioinform. Comput. Biol.* 3, 415–436
17 Funahashi, A. *et al.* (2008) CellDesigner 3.5: a versatile modeling tool for biochemical networks. In *Proceedings of the IEEE 96* pp. 1254–1265
18 Matsuoka, Y. *et al.* (2010) Payao: a community platform for SBML pathway model curation. *Bioinformatics* 26, 1381–1383
19 Thomas, P.D. *et al.* (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141
20 Henney, A. and Superti-Furga, G. (2008) A network solution. *Nature* 455, 730–731
21 Patlak, M. (2010) Open-source science makes headway. *J. Natl. Cancer Inst.* 102, 1221–1223
22 Strauss, S. (2010) Pharma embraces open source models. *Nat. Biotechnol.* 28, 631–634
23 No authors listed. (2008) The Insulin Resistance Pathways Project. http://www.pfizercambridge.com/home.php?id=groups/tgu/focus_insulin
24 Petsko, G.A. (2010) When failure should be the option. *BMC Biol.* 8, 61