Any problem, please contact with: Zhenzhen Yin evamail.pku@gmail.com
Or Yuheng Lu   lgdeer@gmail.com

INTRODUCTION:

(Reverse engineering and our method)

We adopted the process of Reverse Engineering which in our work means to enumerate all possible network topologies and analyzed whether they fit the objection function or not, thus getting the right topology.

Here we chose the object function as Input-Output Alignment. In order to define IOA precisely for need of calculation, we considered most important characters of IOA and adopted Pearson Correlation Coefficient r to represent Input-Output Linear Relationship in the overall search work (when r>0.99 we consider the network topology having the IOA function), and also, regulated two levels for the initial and ultimate output concentration for the second character – the output range in further search work.(Figure 1)

RESULTS:

Calculation process：

Network enumeration

We use three nodes as a minimal framework: one node that receives input(A in Figure 2), a second node that transmits output(C in Figure 2), and a third node that can play diverse regulatory roles(B in Figure 2). There are 9 direct links among the three nodes and there are altogether 3^9=19683 three-node topologies. With 3,645 topologies that have no direct or indirect links from the input to the output occluded, there remain a total of 16,038 possible three-node topologies that contain at least one direct or indirect causal link from the input node to the output node. For each topology, we sampled 10,000 sets of network parameters with the method of latin hypercube sampling (LHS, Figure 3). In all, we have analyzed a total of 16,038*10,000 different circuits. This search resulted in an **exhaustive** circuit function map used to extract core topological motifs essential for IOA.

Equations set up

Our model is based on the following statements:

(1) The nodes are restricted to TF nodes so that the links stand for TF-TF interactions via DNA. The expression level is quantified by the equilibrium binding probability P of TF binding on its

site and the translation rate constant β, and we adopt a constant λ to modify P to make different

TFs equal status. When it comes to several TF factors, we use the multiplication of their λP or $1-\lambda P$ to indicate their interactions.

(2) We take into consideration only the transcription and translation and TF-DNA interactions because other reactions such as signal-transduction activities typically operate much faster and can be considered to be approximately at steady state on the slow timescales of transcription networks. Also the TF activity levels can be considered to be at steady state within the equations that describe network dynamics on the slow timescale of changes in protein levels. So that the equations contain only the accumulation and degradation of the protein products (here the TFs).

(3) It has been observed that one ordinary gene usually has a nonzero expression level with no TFs on its binding site. We propose that one repressor will lower the initial expression level and one activator will shift it, further on, each TF has its unique contribution to the final expression level, which means for example that the expression level with two repressor and one activator

binding on is different from the level under the regulation of two activators and one repressor, but may not necessarily lower than the latter if the only one activator is very strong.

Consider first the simplest condition under which there is only one link from a node to another. (A→C, <mark>Figure 4</mark>)

It is widely accepted that the possibility of TF binding to the binding site in promoter is

$$P = \frac{(X^* / K_d)^n}{1+(X^* / K_d)^n} \; (X^*: \text{the effective concentration of one TF}; K_d: \text{the dissociation constant})$$

According to hypothesis (1)&(3), the link from A(node1) to C(node3) can be translated as

$$\frac{dX_3}{dt} = \beta_0 + (\beta_m - \beta_0)\lambda P - \alpha X_3 = \beta_0 + (\beta_m - \beta_0)\lambda \frac{(X_1^* / K_d)^n}{1+(X_1^* / K_d)^n} - \alpha X_3 = \beta_0(1-\lambda \frac{(X_1^* / K_d)^n}{1+(X_1^* / K_d)^n}) + \beta_m \lambda \frac{(X_1^* / K_d)^n}{1+(X_1^* / K_d)^n} - \alpha X_3$$

(β_0 :the basal translation rate factor; β_m :the effective translation rate factor.)

As $\; X_1^* = [SSI] = kI[S]^2 = kIX_1^2 \quad$ (<mark>Figure 5</mark>)

Thus $\; \dfrac{dX_3}{dt} = \beta_0(1-\lambda \dfrac{(kIX_1 / K_d)^n}{1+(kIX_1 / K_d)^n}) + \beta_m \lambda \dfrac{(kIX_1 / K_d)^n}{1+(kIX_1 / K_d)^n} - \alpha X_3$ (The subscript of X

is the node number.) Actually, the first component of the equation is the fundamental expression level of the network in which the $1-\lambda P$ is the possibility that the TF is off the DNA target site, and the second component of the equation is the effective expression level in which λP is the possibility that TF is on its site.

As to general conditions, there are

$$P_{1i} = \frac{(kIX_1^2)^n}{K_{1i}^n + (kIX_1^2)^n}, X_i = A(i=1), B(i=2), C(i=3), P_{ji} = \frac{X_j^{*n}}{K_{ji}^n + X_j^{*n}}, j=2,3$$

$$\frac{dX_i}{dt} = \beta_0(1-\lambda_{1i}P_{1i})(1-\lambda_{2i}P_{2i})(1-\lambda_{3i}P_{3i}) + \beta_m[1-(1-\lambda_{1i}P_{1i})(1-\lambda_{2i}P_{2i})(1-\lambda_{3i}P_{3i})] - \alpha_i X_i$$

And in order to make the equations fit hypothesis 3，we deduced the proper range of λ , that is,

no regulation $\lambda_{ij}=0$, activation $0<\lambda_{ij}<1$ and repression $1-\sqrt[3]{\dfrac{\beta_m}{\beta_m - \beta_0}}<\lambda_{ij}<0$. The choice of

other Parameter values and their reference is in <mark>Table 1.</mark>

network topologies' analysis

Aiming at getting the values of r for each circuit, we need to numerically simulate the ODE equations to get the steady-state concentration of output node C under each input concentration, and then making linear fit of input and output concentrations. As our input concentration range is $10^{-9} \sim 10^{-5}$, we select points that have the same logarithmic distance intervals, then simulate output evolution curve to get the steady concentration one point by one. We choose the fourth-order Runge-Kutta method to solve the ODE equations and so as to save calculation time, we adopt Implicit Runge-Kutta algorithm to get the output steady concentration when Input=10^{-9}M and set the very concentration as initial value for the Newton-Raphson method for following different Input concentration. And considering the possibility of bistable network topology, we calculate the

two directions (positive sequence and the reverse) that Input concentration changes to avoid wrongly supposing it as one IOA function circuit.

Identifying Minimal IOA Networks

Here we define Q value as the number of IOA function circuit among 10000 sets of network parameters, and it indicates the robustness of one topology to finish the IOA function—the larger Q value is, the more robust the topology is. We sort in reverse sequence all the network topologies according to their Q value and the x axis is their rank(Figure 6). We can observe that most network topologies have 0 or low Q value while there's only a small part of the topologies having large Q value.

We firstly analyszed the first 160 network topologies (Q >=705) and list in Figure 7 all the simplest topologies that have only 3 or less direct links between the three nodes.(Figure 7) Among the 14 topologies, there are 12 three-link networks and 2 two-node networks so that we can see that the minimal number of links for the topologies to be functional is two but the most usual number is 13. The common features of the networks capable of IOA are either one negative control loop (NCL) or one negative feedback loop (NFL). Here we define NCL as a topology that has one negative control on the input-receiving node (A) from the intermediate node (B) and one positive regulation on the output node (C) from A node.(as the 1st topology in Figure 7 and the 1st one in Figure 8) Similarly we define NFL as a topology that share the positive-regulation link from A to C, while uniquely have one negative feedback from C to A.(as the 2nd topology in Figure 7 and the 2nd one in Figure 8) And the NCL topology seemingly more robust than the NFL topology, as there are 9 topologies out of the 14 simplest networks contain NCL topology compared to 6 of NFL, so that B node appears to be important, which will be discussed in details in the following part.

Mechanisms of Minimal IOA Networks and Key Parameters Analysis

Aimed at answering the question why the two topologies defined above (Figure 8) is functional in IOA, we unravel their mechanisms using the ODE equations in this part, also getting the parameter restrictions of each topology.

NCL Topology

When the network has built steady state:

$$\frac{dX_1}{dt} = \beta_0(1 - \lambda_{21}\frac{X_2}{K_{21} + X_2}) + \beta_m[1 - (1 - \lambda_{21}\frac{X_2}{K_{21} + X_2})] - \alpha_1 X_1 = 0$$

$$\frac{dX_2}{dt} = \beta_0 - \alpha_2 X_2 = 0$$

$$\frac{dX_3}{dt} = \beta_0(1 - \lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2}) + \beta_m[1 - (1 - \lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2})] - \alpha_3 X_3 = 0$$

Solve the equations:

$$X_2 = \frac{\beta_0}{\alpha_2} \quad \text{is constant;}$$

$$X_1 = \frac{\beta_0}{\alpha_1}(1 - \lambda_{21}\frac{X_2}{K_{21} + X_2}) + \frac{\beta_m}{\alpha_1}[1 - (1 - \lambda_{21}\frac{X_2}{K_{21} + X_2})] \text{ is also constant;}$$

$$X_3 = \frac{\beta_0}{\alpha_3}(1 - \lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2}) + \frac{\beta_m}{\alpha_3}(\lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2})$$

If there is $K_{13} >> kIX_1^2$, then

$$X_3 = \frac{\beta_0}{\alpha_3}(1 - \frac{\lambda_{13}kX_1^2}{K_{13}}I) + \frac{\beta_m}{\alpha_3}(\frac{\lambda_{13}kX_1^2}{K_{13}})I = \frac{\beta_0}{\alpha_3} + \frac{\beta_m - \beta_0}{\alpha_3}(\frac{\lambda_{13}kX_1^2}{K_{13}})I$$

Because X_1 is constant, the coefficient of I is also constant. So there comes the linear correlation between X_3(output concentration) and I.

When $K_{13} >> kIX_1^2$,

$$\frac{dX_3}{dt} = \beta_0(1 - \lambda_{13}\frac{kIX_1^2}{K_{13}}) + \beta_m(\lambda_{13}\frac{kIX_1^2}{K_{13}}) - \alpha_3 X_3 = \beta_0 + (\beta_m - \beta_0)\lambda_{13}\frac{kX_1^2}{K_{13}}I - \alpha_3 X_3$$

And the concentration of B node is actually part of the slope coefficient the linear equation, we can imagine that the function of B node is to lower the concentration of A straightly without the interference of other factors as well as control the concentration of A more precisely and more freely to make the parameter restriction easier to achieve and at the same time the output range is not too small. As we know, the stochastic error may make vague the linear relationship when the values of y axis are too near. The lower concentration X_2 is, the steeper the line is, and so the bigger the range is, which in biology means that the bioreporter is more sensitive to certain environmental signal. Through modifying the parameters of node B, we can get a proper concentration of A node to achieve a good r. In all, the node B is a proportion node.

NFL Topology

$$\frac{dX_1}{dt} = \beta_0(1 - \lambda_{31}\frac{X_3}{K_{31} + X_3}) + \beta_m[1 - (1 - \lambda_{31}\frac{X_3}{K_{31} + X_3})] - \alpha_1 X_1 = 0$$

$$\frac{dX_3}{dt} = \beta_0(1 - \lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2}) + \beta_m[1 - (1 - \lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2})] - \alpha_3 X_3 = 0$$

Also solve the equations:

$$X_1 = \frac{\beta_0}{\alpha_1}(1 - \lambda_{31}\frac{X_3}{K_{31} + X_3}) + \frac{\beta_m}{\alpha_1}(\lambda_{31}\frac{X_3}{K_{31} + X_3}) = \frac{[(\beta_m + \beta_0)\lambda_{31} + \beta_0]X_3 + \beta_0 K_{31}}{\alpha_1(K_{31} + X_3)}$$

$$X_3 = \frac{\beta_0}{\alpha_3}(1 - \lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2}) + \frac{\beta_m}{\alpha_3}(\lambda_{13}\frac{kIX_1^2}{K_{13} + kIX_1^2})$$

If $|(\beta_m + \beta_0)\lambda_{31}| << \beta_0$, then $X_1 = \frac{\beta_0 X_3 + \beta_0 K_{31}}{\alpha_1(K_{31} + X_3)} = \frac{\beta_0}{\alpha_1}$ is constant;

And if $K_{13} >> kIX_1^2$, then $X_3 = \dfrac{\beta_0}{\alpha_3} + \dfrac{\beta_m - \beta_0}{\alpha_3}(\dfrac{\lambda_{13}kX_1^2}{K_{13}})I$, which correlation is the same as the NCL Topology.

Compare X_1 here with the concentration of node A in NCL topology, one can easily discover that X_1 here is higher and more dull, which makes it more hard to get a balance between satisfying the second parameter restriction and the need for the range to be rational. Then we can understand why NFL topology is less than NCL in high-Q-value topologies.

Aimed at unraveling the relative importance among parameters, we got all functional sets of parameters of NCL&NFL and selected randomly one set, respectively, from the two topologies and processed them with Matlab to compare the differences once the parameters change. The results are in Figure 9. From changes of the two important characters – the output range and r with the change of parameters (Table 2), we got the answer to previous questions. For NCL, the degradation rate of A (α_1) and the value of K_d from node A to node C (K_{13}) are relatively more important to r, and the degradation rate of C (α_3) along with K_d from node A to node C (K_{13}) are essential to the output range, while the constant indicating repression from B to A(λ_{21}), the degradation rate of B (α_2) as well as K_d from node B to node A (K_{21}) influence little both characters. From Table 2 and the analysis above, we can draw the conclusion that in practice, the parameter λ_{13} should be larger and α_3 should be as small as possible. While for NFL, the condition is similar. It is still α_3 and K_{13} that are important to r, but the key parameters on the output range are λ_{13} α_1 α_3 K_{13}, more than the key parameters for output range for NCL. So the suggest to practice is still smaller α_3 and larger λ_{13}, what's more, there should be smaller λ_{31} for larger output range according to the character changes in Table 2.

Parameters Tendency of the IOA Networks
In order to test the correctness of our analysis about the key parameters, we look more closely at the parameters of the simplest IOA networks. For the two simplest networks in Figure 6, we examine the distribution of the parameter values which can realize linear response between input and output out of the total 10000 sets of parameters used in the search.
(Figure10, Figure11)
It can be seen that the constant K of the repression (K21 and K31) are generally lower compared to K13, which means that there should be an intense repression on node A. Such a result also fits well with the condition $K_{13} >> kI[X_1]^2$ used in the derivation above.

We also notice that the degradation rates, α, of the nodes also show apparent tendency. The degradation rate of the node A is generally higher, while the repressing nodes (B or C) have lower degradation rates. Such a tendency also shows that the repression on node A should be intense in order to establish a more linear response curve between input and output.

Analysis of All Possible Three-Node Networks

  The above analyses focused on minimal (less than or equal to 3 links) three-node networks and identified simple topologies that are sufficient for IOA function, also unraveling the mechanism that the topologies work. But whether the topologies are necessary for the IOA function is not understood yet. In other words, are the identified minimal topologies the foundation of all possible networks, or are there more complex higher-order solutions that do not contain these minimal topologies? To answer the questions above, we analyzed the first 160 topologies (Q>705) that are well capable of the IOA function. (Figure 12)

  Analysis of these robust topologies shows that they are overrepresented with NCL and NFL. All 160 topologies contain at least one NCL or NFL motif (or both). These results indicate that at least one of these motifs is necessary for IOA function.

  Supplementarily, the NCL average Q value(AQV)of all 19683 topologies is 17.14 while the NFL AQV is only 9.36, which again indicate that the NCL is more robust than NFL that have drew conclusion in the minimal topology analysis.

Motif Combinations that Improve IOA

   To investigate what additional features can improve the functional performance in some more complex and more robust networks than minimal topologies, we clustered the first 160 networks and then cluster them respectively in three categories: NCL, NFL and the combination of the two. (Figure 11) The results clearly indicate that apart from the link from A to C, there should be no positive regulation, and the NFL topology hates the link from A to A, while the combination topologies show no additional tendency.

In all, by exhaustively searching all network and analyzing the results, we draw the conclusion that NCL meets our need in application well, and we get the proper parameter range for practice.
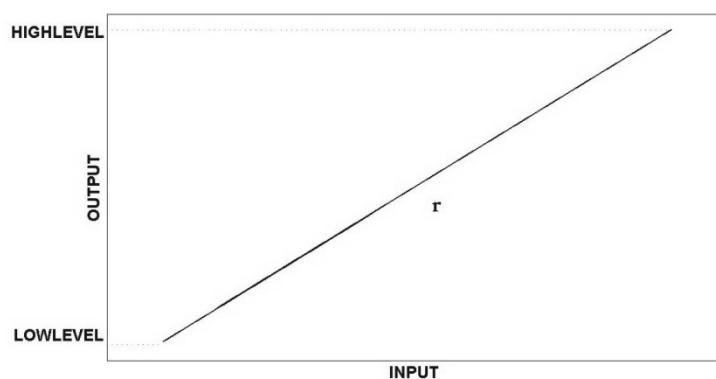
Figure 1 Factors for selection of IOA network topologies. r is the Pearson Correlation Coefficient and the output range is HIGHLEVEL minus LOWLEVEL**.**
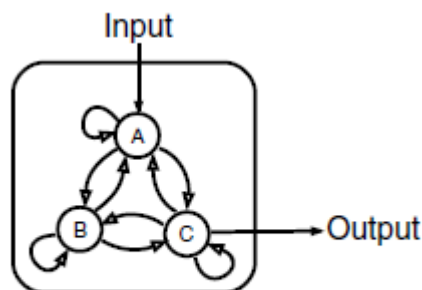
16038 networks

Figure 2 Three-node network with all of its possible directed links(Ref. 7) There are altogether 9 possible links and the input here is Hg(II), and the concentration of C is taken as output.
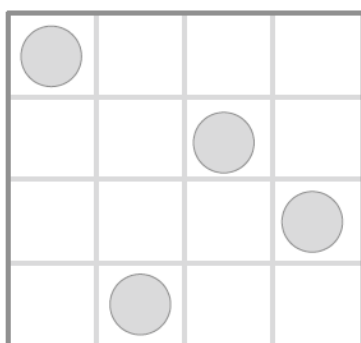
Figure 3   Latin Hypercube Sampling When sampling a function of N variables, the range of each variable is divided into M equally probable intervals, M sample points are then placed to satisfy the Latin Hypercube requirements**.** Then each sample is theonly one in each axis-aligned hyperplane containing it.

Figure 4 simplest network There is only one link from node A to node C and we don't regulate the property of the link. Actually, whether it is activation, repression or no regulation is represented by λ.
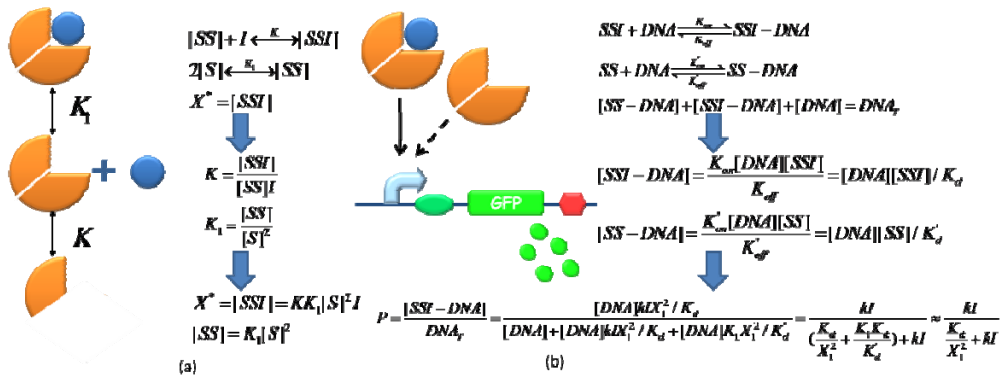
Figure 5 MerR interacts with Hg(Ⅱ) The orange poor arc sector stands for the MerR monomer and the orange major arc sector stands for the MerR dimer. The blue ball is Hg ion that can be bound by the MerR dimer, and only the complex can serve as Transcription Factor that promotes the expression of downstream genes. We list the balance equations, derive the calculation function of X^* as above, and substitute this function to the ODE equations used in our model.
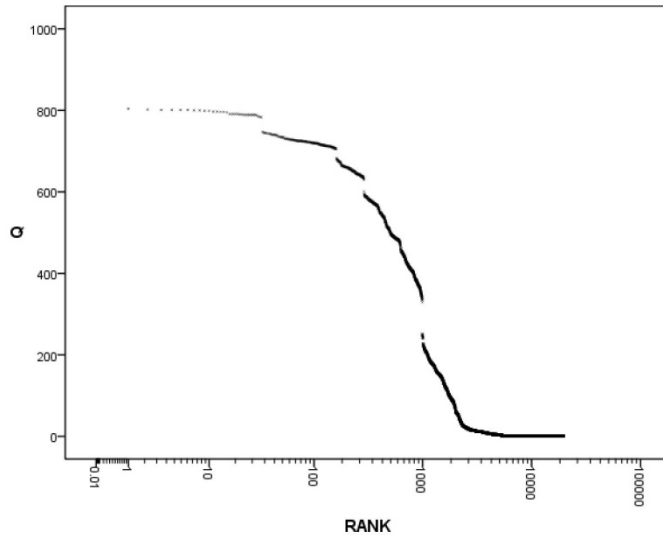
Figure 6 The Q value network topologies Sequence We sort in reverse sequence all the network topologies according to their Q value. X Axis is their ranks, while Y Axis is their corresponding Q values. The figure indicates that most network topologies have 0 or low Q value while there's only a

small part of the topologies having large Q value.
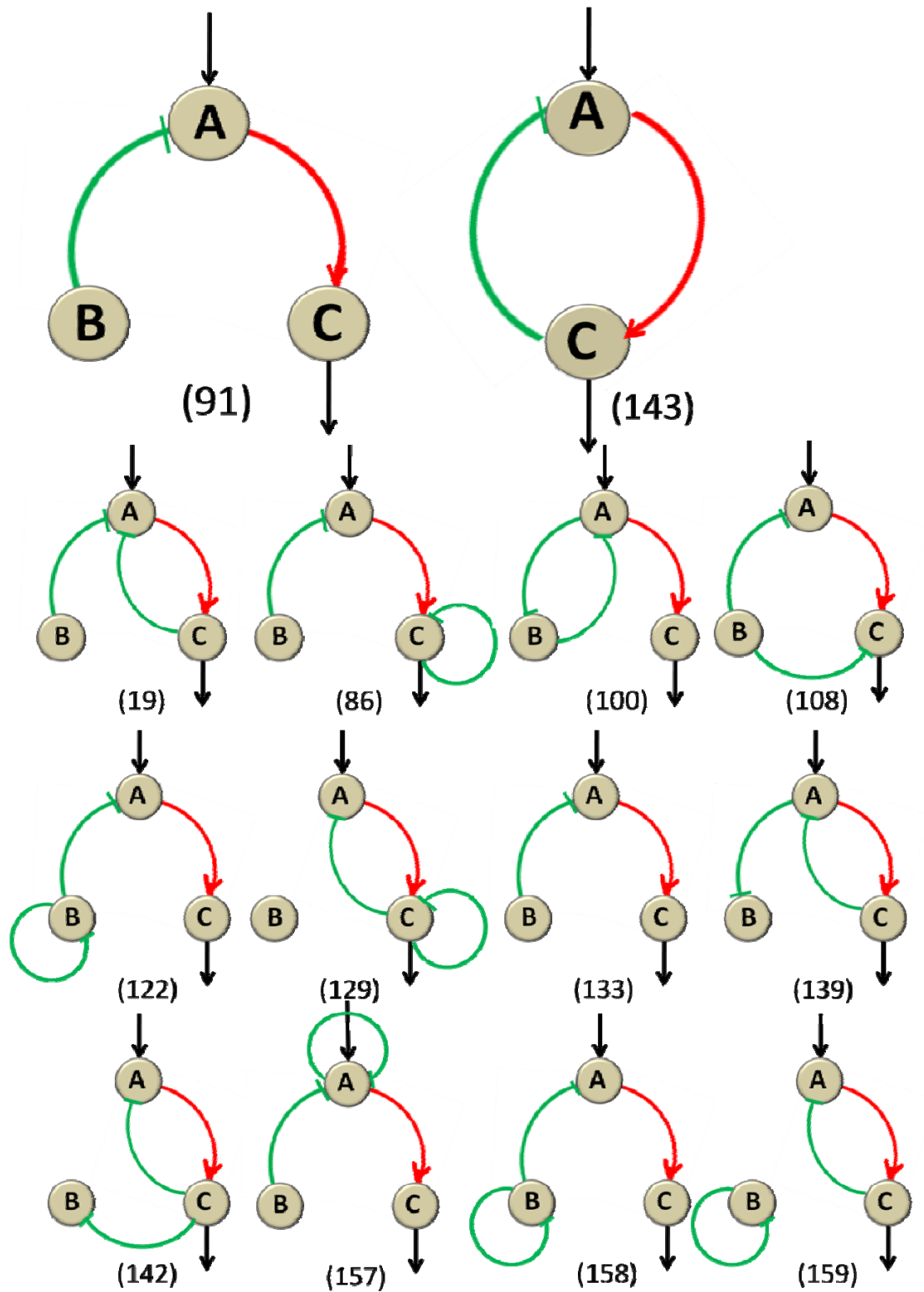
Figure 7 All Functional Networks The numbers below each network are their ranks. The two bigger topologies have 2 links while the other 12 networks have 3 links. In each network, the green arc with one short straight line at one end stands for repression from the start node to the end node and the red arc with one arrow at one end stands for activation.
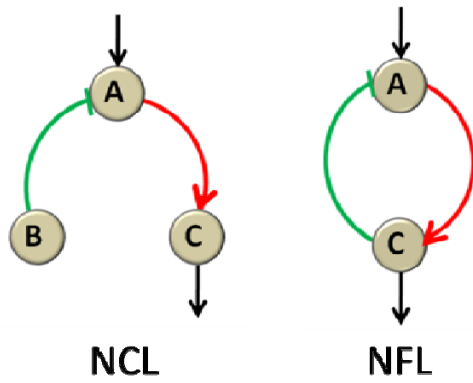
Figure 8 The NCL and NFL topology The green and red line have the same meaning as in Figure 7.
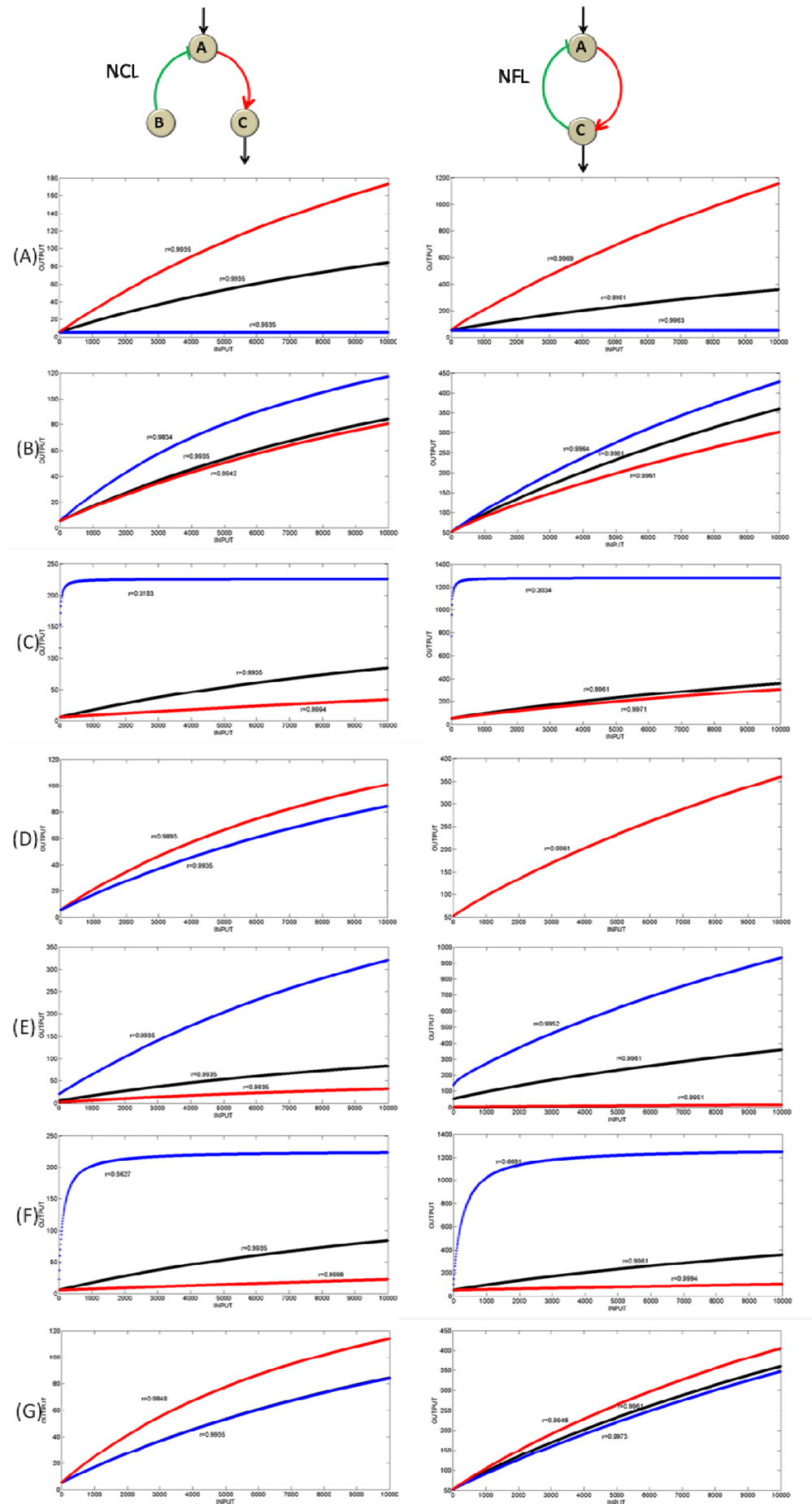
Figure 9 Analysis for key parameters We got all functional sets of parameters of NCL&NFL and selected randomly one set, respectively, from the two topologies and processed them with Matlab to compare the differences once the parameters change. When analyzing one parameter, we only change this very parameter and keep others the same, and when we change the parameter to a lower level, we get the blue line, when to a higher level, we get the red line and the black line is for the unchanged parameter set. Each line has its Pearson Correlation Coefficient r marked in the figure. The X Axis is the concentration of Hg ion as INPUT whose range is 1 to 10000 nM, and the Y Axis is the concentration of node C with the unit of nM. (A) We analyze λ_{13} for both topologies.

(B) λ_{21} for NCL and λ_{31} for NFL (C) The degradation rate of A (α_1)for both topologies (D) The degradation rate of B(α_2)for both (E)The degradation rate of C(α_3) for both (F) The dissociation rate of A to the binding site on the gene of C(K_{13}) for both (G) K_{21} for NCL and K_{31} for NFL.
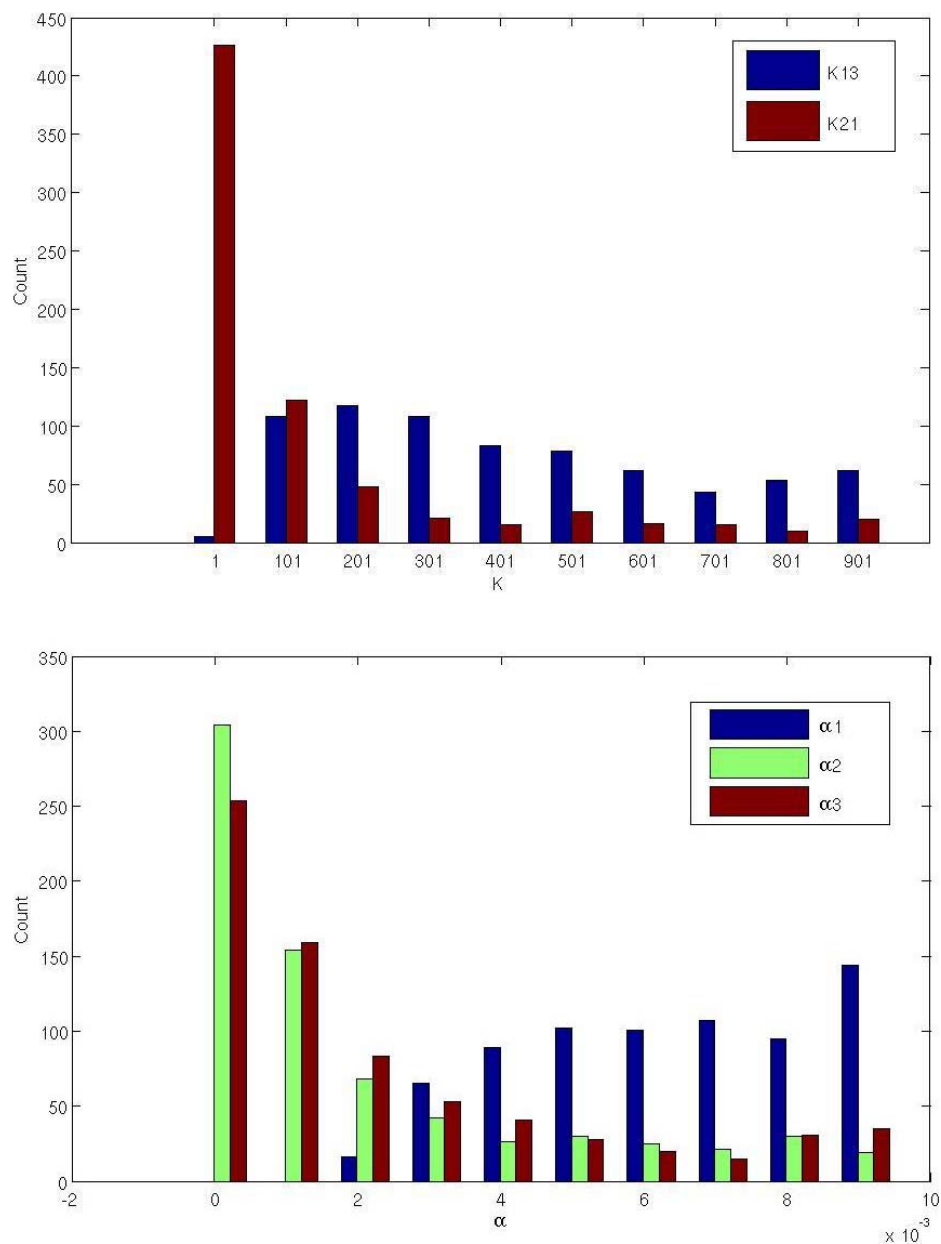
Figure 10 Distribution of NCL topology parameters which can establish linear response curves.
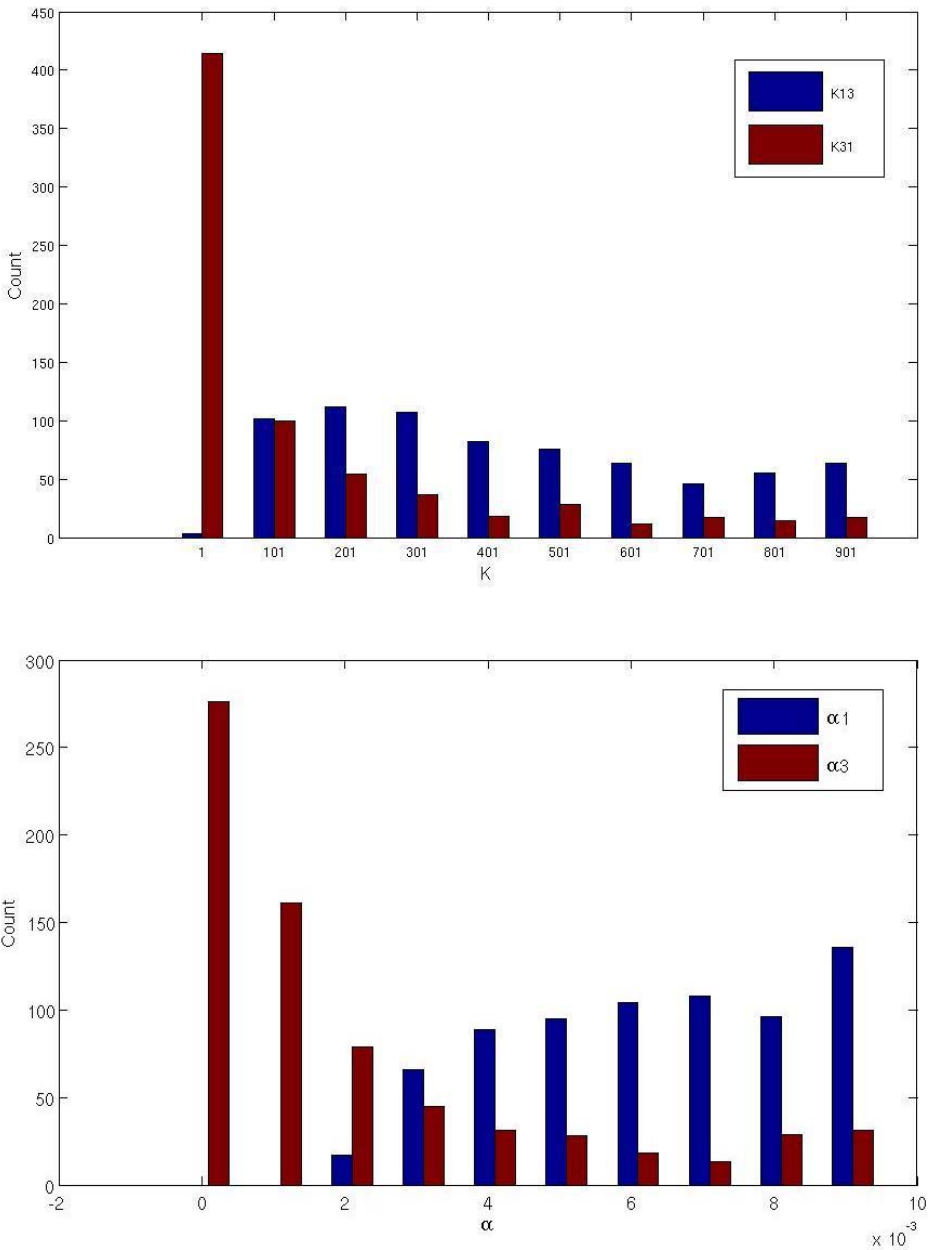
Figure 11 Distribution of NFL topology parameters which can establish linear response curves.
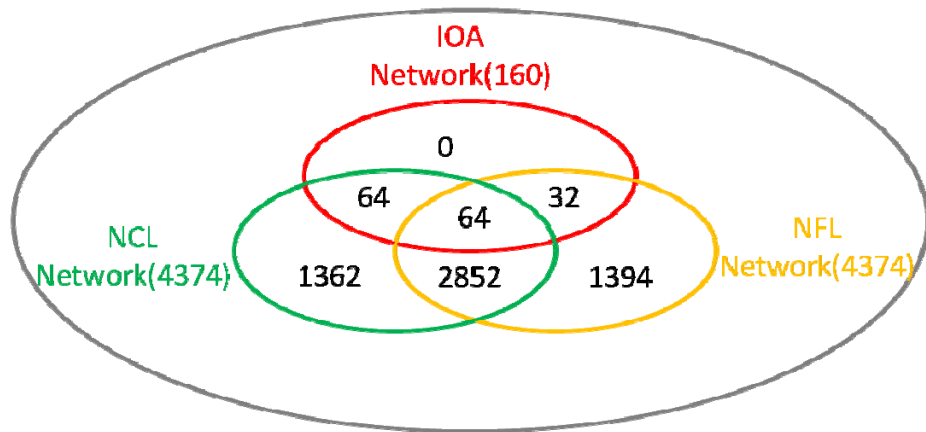
Figure 12 Analysis of the first 160 networks We count all the NCLs and NFLs and also the IOA networks, and discover that all of the IOA networks can be classified into NCL/NFL/the combination of the two. And there are more IOA functional network featured in NCL than those characterized with NFL.
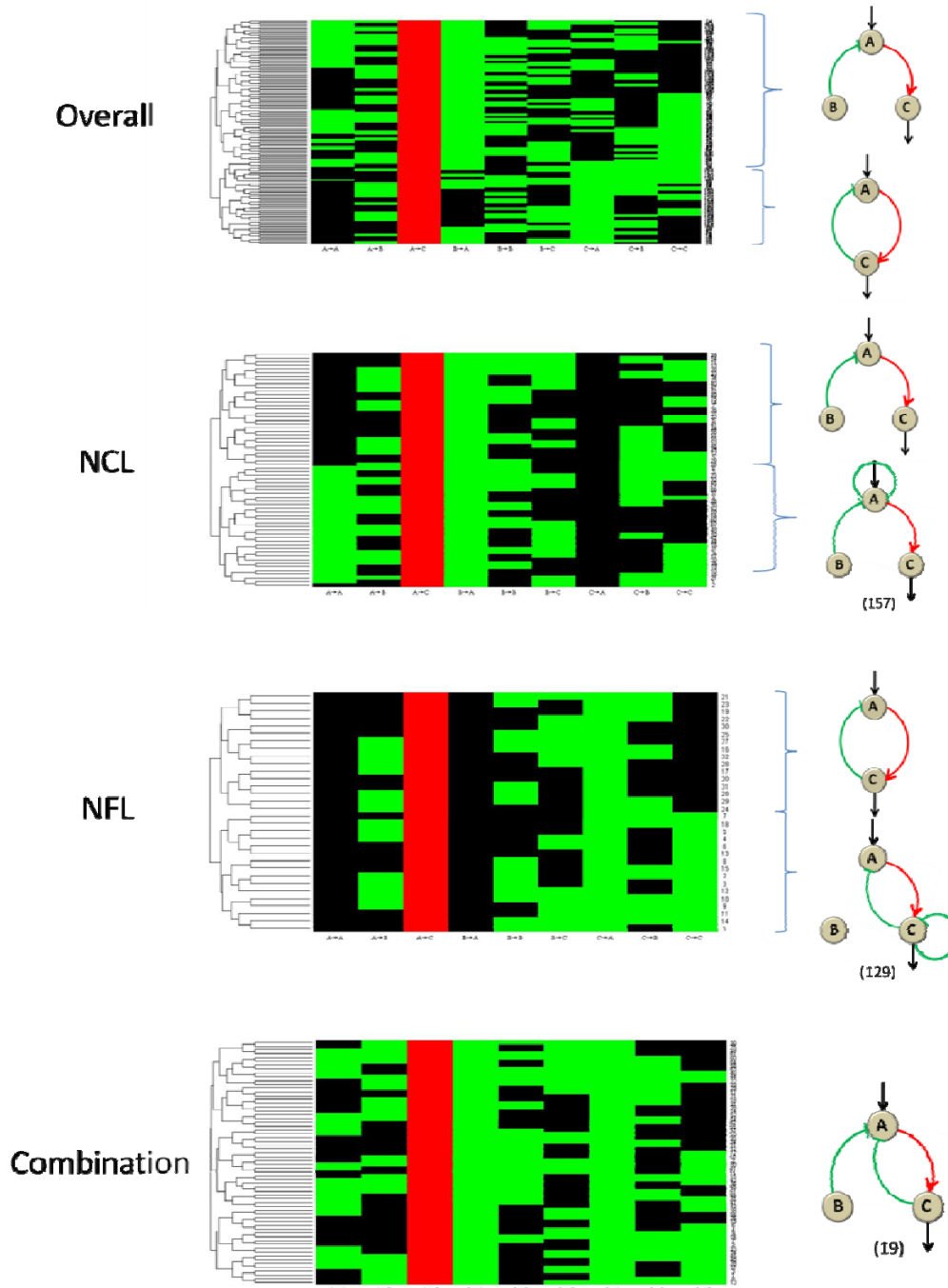
Figure 13 The clustergrams of the networks We use the clustergram command in matlab to get the additional features of the functional networks. The nine vertical rectangle bar stand for nine links in Figure 2 which are, respectively, from A to A, from A to B, from A to C, from B to A, from B to B, from B to C, from C to A, from C to B, from C to C. And red stands for activation, green for repression and black for no regulation. The topologies on the right are corresponding minimal topologies that is shown in the clustergrams on the left.

Table 1 Physiological range of the model parameters

Parameter	Description	Value(in vivo)	Principal molecular determinants	Reference
β_0	The basal maximum translation rate	2e-11M/sec	The gene character	1,2
β_m	The effective maximum translation rate	2e-9M/sec	The promoter activity – strength of TF-RNAp interaction	2
k	$K*K_1$	1e15	The stability of MerR dimer and the binding affinity of MerR dimer and Hg ion	2,3
λ	The interaction factor	0-1(activator) 0(no regulation) $1-\sqrt[3]{\frac{\beta_m}{\beta_m-\beta_0}}$ -0(repressor)	The difference of strength of activators and repressors	Derived ourselves
K	Dimer dissociation constant	1e-9-1e-6M	Monomer-monomer affinity	2,4,5
α	Protein degradation rate	1e-4-1e-2 M/sec	Growth rate(dilution), protein stability, degradation tag(proteolysis)	2,6
n	Cooperativity in promoter activity	1	Number of operator-bound TFs interacting with RNAp	2,5,7

Table 2 The change of two important characters as the parameters rise

Parameter that changes bigger	NCL		NFL	
	Range	r	Range	r
λ_{13}	↑ (red)	—	↑ (red)	—
λ_{21} for NCL λ_{31} for NFL	↓ (green)	↑ (red)	↓ (green)	—
α_1	↓ (green)	↑ (red)	↓ (green)	↑ (red)
α_2	↑ (red)	↓ (green)	—	—
α_3	↓ (green)	—	↓ (green)	—
K_{13}	↓ (green)	↑ (red)	↓ (green)	↑ (red)
K_{21} for NCL K_{31} for NFL	↑ (red)	↓ (green)	↑ (red)	↓ (green)

* has the meaning that the character value rises as the parameter rises. has reverse meaning.

means that the character value keep relative stability as the parameter changes.

Reference

1 Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2002)*Molecular Biology of the Cell* (Garland, New York).

2 Nicolas E. Buchler, Ulrich Gerland, Terence Hwa, Nonlinear protein degradation and the function of genetic circuits[J],PNAS,2005,102(27),9559-9564.

3 John D. Helmann, Barry T. Ballard, Christopher T. Walsh, The MerR Metalloregulatory Protein Binds Mercuric Ion as a Tricoordinate, Metal-Bridged Dimer[J], Science, 1998, 247, 946-948.

4 Diana. M, Ralston and Thomas V. O Halloran, Ultrasensitivity and heavy-metal selectivity of the allosterically modulated MerR transcription complex[J], PNAS, 1990, 87, 3846-3850.

5 URI ALON, An Introduction to Systems Biology—Design Principles of Biological Circuits[J], Chapman&Hall/CRC,2007.

6 Wickner, S., Maurizi, M. R. & Gottesman, S. (1999) Science, 286, 1888–1893

7 W. Ma, A. Trusina, H. El-Samad, W.A. Lim, C. Tang. Defining network topologies that can achieve biochemical adaptation[J]. Cell. 2009, **138**:760–773.