Discriminating direct and indirect connectivities in biological networks
识别生物网络中的直接和间接连接

学院：信息学院
姓名：邢向阳

content

- Background
- Methods
- Results
- Conclusion

background

- Direct and indirect interactions are pervasive in all networks.
- 在生物网络中直接和间接的相互作用是普遍存在的。
- Despite concurrent advances in quality and quantity of data as well as computing resources and algorithms, a fundamental reverse engineering bottleneck is the ability to discriminate between direct and indirect connections.
- 尽管在数据的数量和质量以及计算能力和算法上都在并行进步，但逆向工程的一个根本瓶颈是如何够区分直接和间接连接。
- The inability to disentangle these interactions hampers reverse engineering progress.
- 这些相互作用的难以辨别阻碍了逆向工程的进展。

background

- A number of theoretical approaches have been proposed to overcome this hurdle， but the ability to experimentally verify the conclusions drawn by reverse engineering tools remains paramount.
- 许多的理论方法被提出来克服这个障碍，但能够通过逆向工程工具实验验证这些结论仍然是最重要的。
- we adopt the notions of abstraction, emulation, benchmarking, and validation in the context of discovering features specific to this family of connectivities.
- 研究人员采用概念抽象，仿真，基准测试及环境验证的方法来探索这些网络连接性的特征。

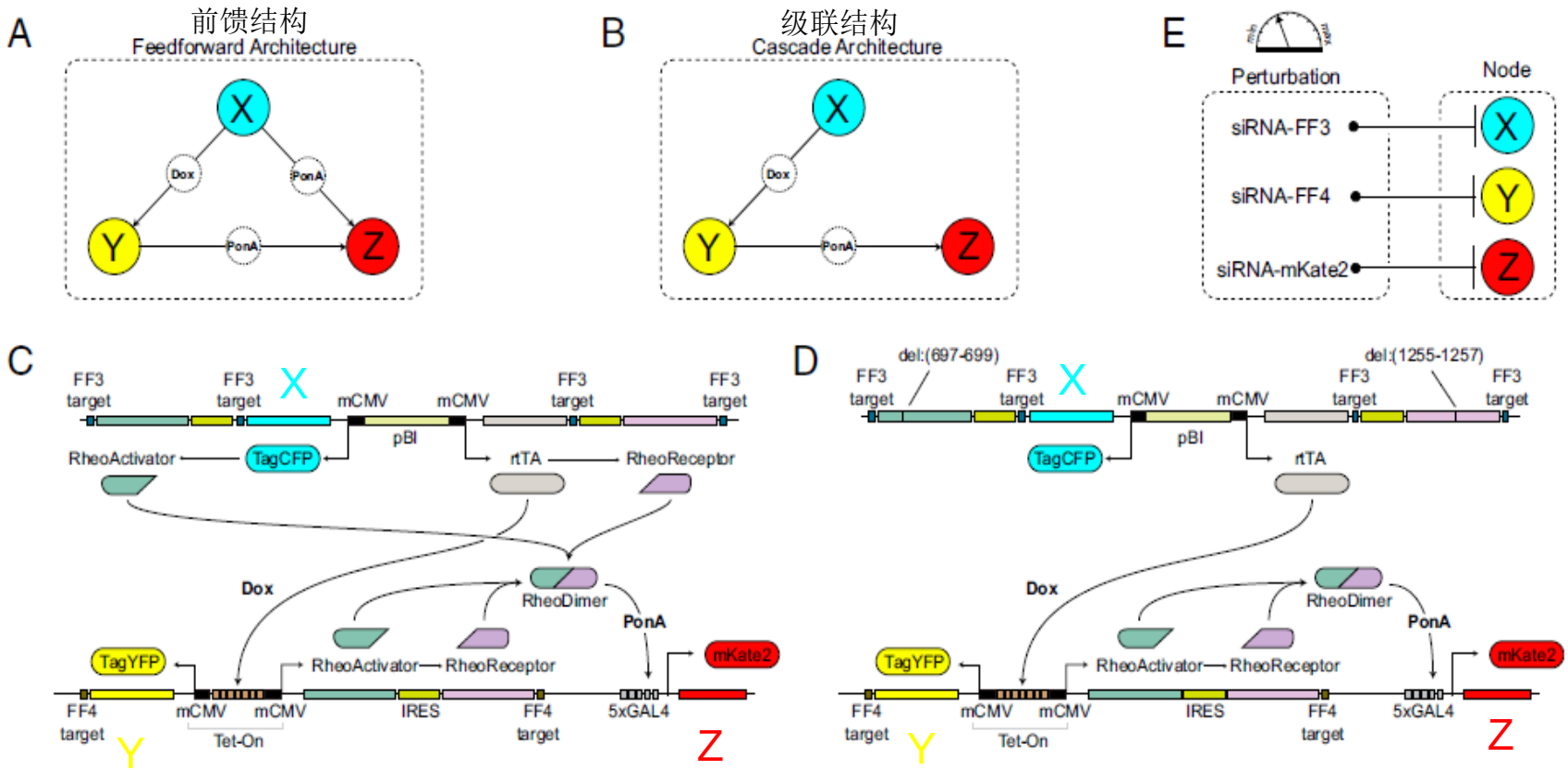
Methods

- Mammalian Cell Culture and Transfections
- 哺乳动物细胞培养和转染
- Fluorescence Microscopy
- 萤光显微镜检查
- Flow Cytometry
- 流式细胞计数法
- Modular Response Analysis
- 模块化的响应分析
- Resampling
- 重采样

Results

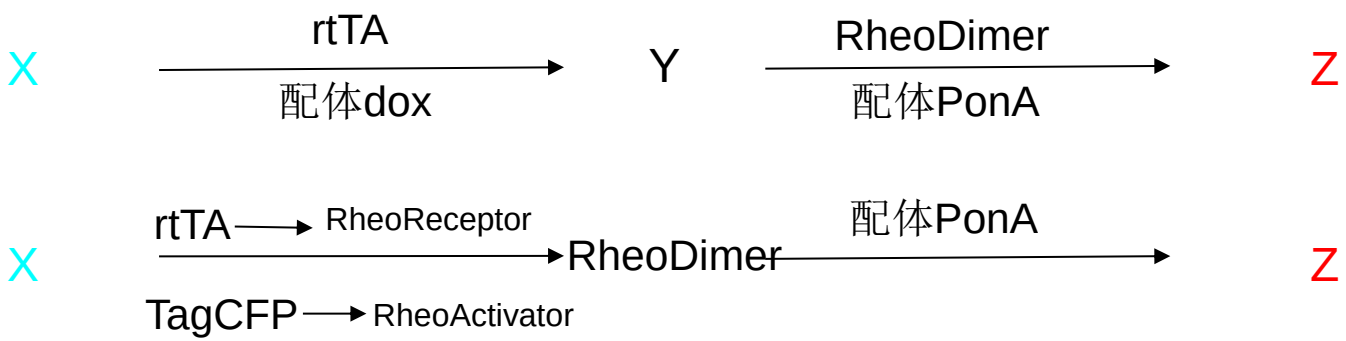
- Design and Assembly of the Benchmark Synthetic Regulatory Networks
- 基准合成调控网络的设计与组装
- Validation of the Synthetic Gene Network Behavior
- 验证合成基因网络行为
- Modular Response Analysis
- 模块化响应分析
- Reverse Engineering of the Benchmark Topologies Using Resampled Single-Cell Data
- 使用单细胞数据重采样进行基准拓扑的逆向工程

Design and Assembly of the Benchmark Synthetic Regulatory Networks
基准合成调控网络的设计与组装



- A图模体是X节点通过直接和间接方式调控Z节点的前馈结构
- B图模体是X节点调控Z节点只能通过激活节点Y的级联结构
- CD图是实现AB结构的合成基因网络的详细图解
- E图系统中每个节点的摄动由siRNA完成，X、Y节点分别受siRNAs-(FF3 、 FF4），Z节点受custom siRNA-mKate2摄动

Design and Assembly of the Benchmark Synthetic Regulatory Networks
基准合成调控网络的设计与组装



- The activity of the three nodes X, Y, and Z can be quantified by the output fluorescent proteins TagCFP, TagYFP and mKate2.
- X,Y,Z三个节点的激活可以分别被产生的3种荧光蛋白TagCFP,TagYFP,mKate2所量化
- To achieve direct activation of node Z by node X, the node X produces the RheoSwitch proteins in addition to TagCFP and rtTA.
- 为实现X节点直接激活Z节点，X节点中TagCFP,rtTA将补充产生RheoSwitch蛋白。
- For the cascade motif, the translation of RheoSwitch dimer protein is prevented by nonsense mutation.
- RheoSwitch二聚体蛋白的转化通过无义突变的手段被阻止，从而实现级联结构的完成。
- the activation of node Y by node X depends on doxycycline，Z requires an EcR agonist such as Genostat or ponasterone A (PonA).
- Y节点通过X节点的激活依靠dox配体，Z节点需要配体PonA

Validation of the Synthetic Gene Network Behavior
验证合成基因网络行为

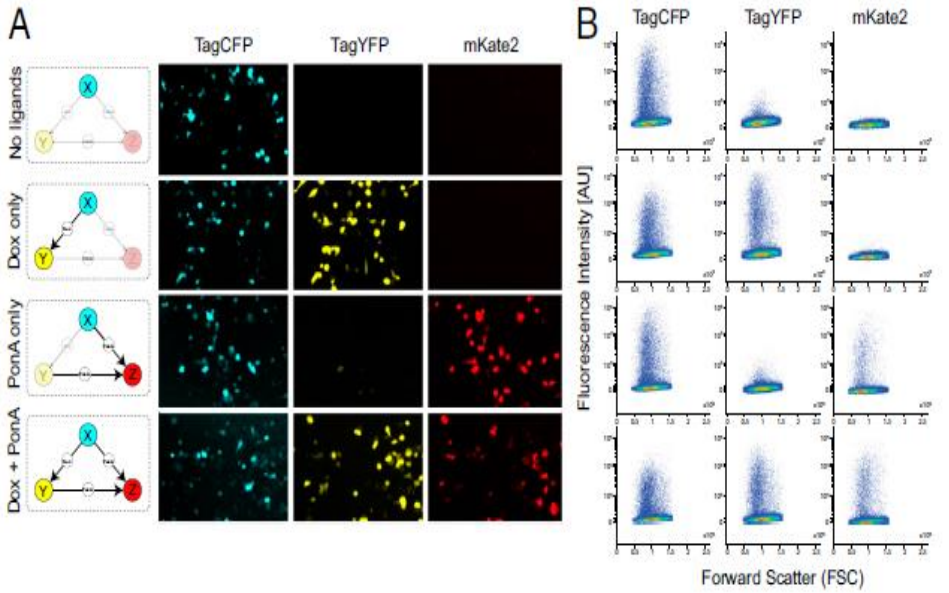


Fig. 2. Validation of the coherent feed-forward architecture. To validate the circuit behavior we tested all combinations of the two small molecules. The result analyzed by fluorescence microscopy (*A*) and flow cytometry (*B*).

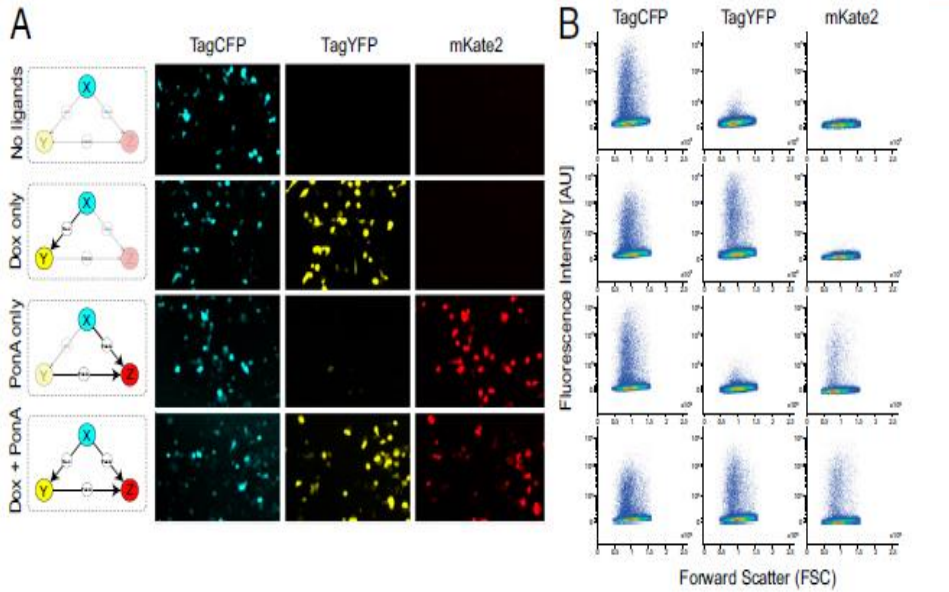


Fig. 3. Validation of the cascade architecture. To validate the circuit behavior we tested all combinations of the two small molecules. The result analyzed by fluorescence microscopy (*A*) and flow cytometry (*B*).

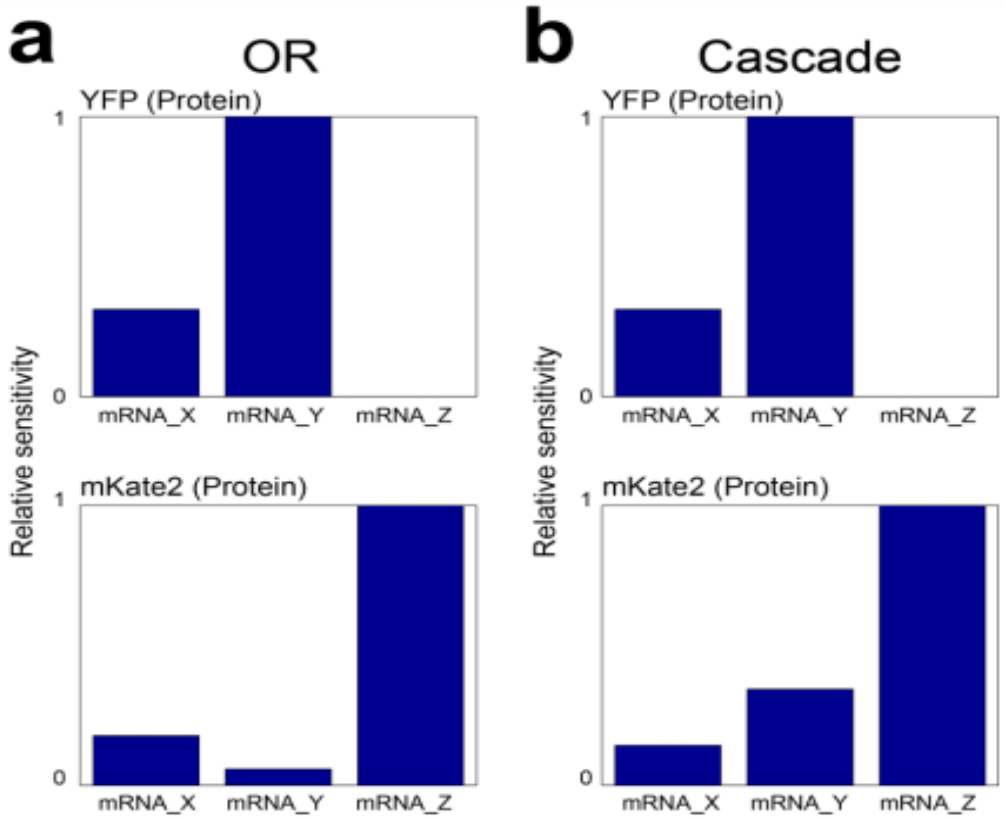
Fig.2 验证前馈结构，测试2种配体小分子的所有组合验证回路行为，使用萤光显微镜检查及流式细胞计数法进行结果分析

Fig.3 验证级联结构，测试2种配体小分子的所有组合验证回路行为，使用萤光显微镜检查及流式细胞计数法进行结果分析

Validation of the Synthetic Gene Network Behavior
验证合成基因网络行为

- the circuits were transfected in human embryonic kidney cell line (HEK293).
- 回路在人类胚胎肾细胞系(HEK293)转染进行实验。
- In the feed-forward loop, node X activity is represented by the constitutively produced fluorescent protein TagCFP and is observed regardless of the ligand conditions (Fig. 2A). The addition of doxycycline, which enables X-to-Y interaction by activating the synthetic transactivator rtTA, results in production of the TagYFP fluorescent protein (Fig. 2A). The activation of node Z is mediated by the active form of RheoSwitch dimer, which is produced by both nodes X and Y. Due to the constitutive activity of node X, PonA is sufficient to activate node Z in the feedforward loop (Fig. 2A).
- 在前馈结构中，节点X的激活与结构性产物荧光蛋白TagCFP有关和配体条件无关。 通过增加dox, 可激活合成反式激活因子rtTA完成x节点到y节点的激活交互,产生的TagYFP荧光蛋白可以验证。节点Z的激活是由活性RheoSwitch二聚体调控(节点X和Y都可产生）。由于节点X激活的结构,在前馈回路中PonA足以激活节点Z。
- In the cascade motif, sequential activation of node X and node Y are necessary for node Z activation. Thus, mKate2 is only observed when Dox and PonA are present (Fig. 3A).
- 在级联结构中，Z节点的激活需要连续激活XY节点。因此mKate2(量化Z的荧光蛋白)只在配体Dox和PonA都存在是才被发现。

Validation of the Synthetic Gene Network Behavior
验证合成基因网络行为



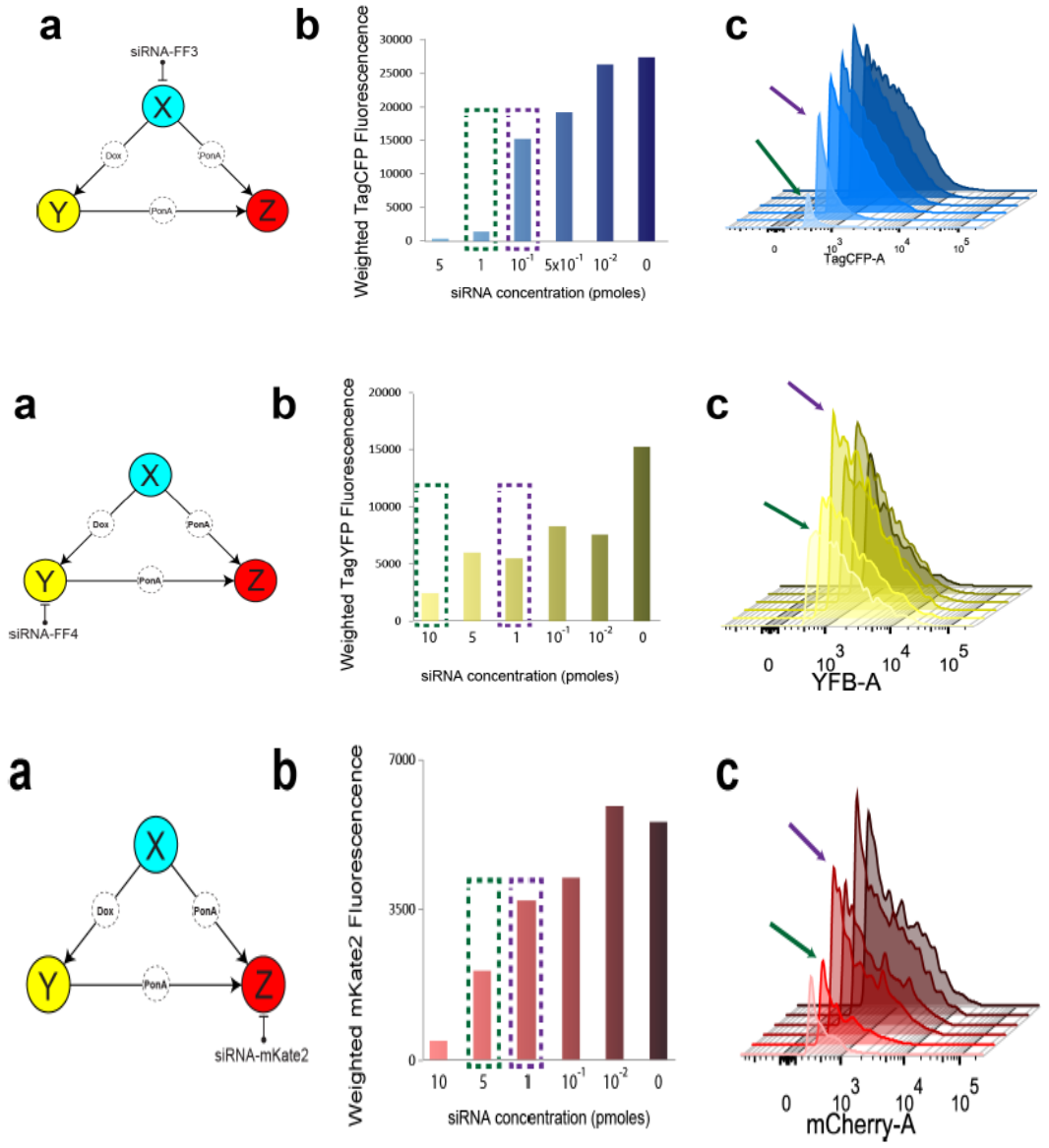
- We performed sensitivity analysis of the output node Z protein concentration against the mRNA species of nodes X andY .
- 进行输出节点Z蛋白质浓度对节点XY mRNA的灵敏度分析实验。

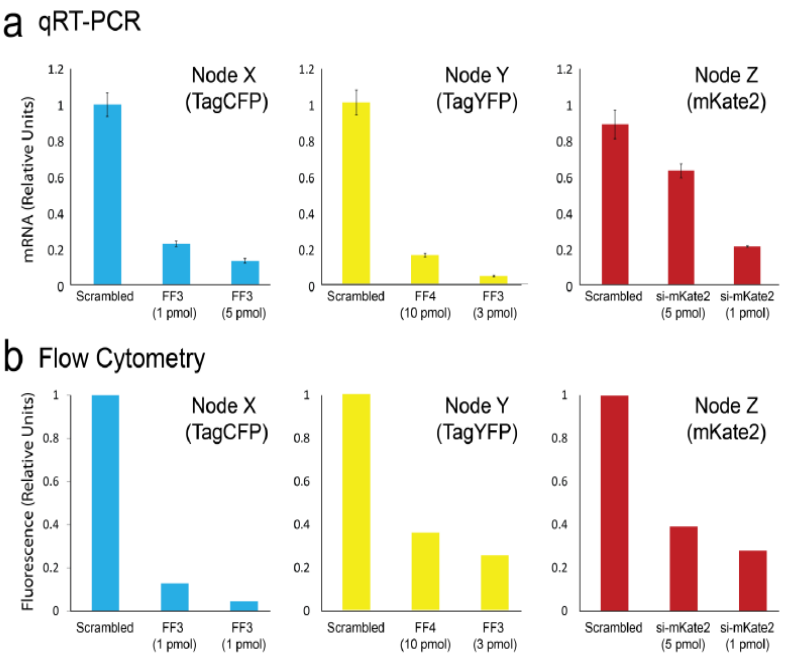
- We observe that, in the feed-forward loop, where node X activates node Z in a direct manner as well as an indirect manner, the cumulative sensitivity of the Z node protein to mRNA species of node X was always higher than that of node Y(SIAppendix, Fig. S3). Conversely, in the cascade, where node X only activates node Z indirectly through node Y, the production of node Z protein was more sensitive to the node Y mRNA.
- 观察到,在前馈回路中,节点X激活节点Z以直接及间接的方式,Z节点蛋白质的产生显示节点X mRNA 敏感性总是高于节点Y mRNA(SI Appendix S3)。相反,级联中,节点X只间接通过节点Y激活节点Z中,Z节点蛋白质的产生对节点Y mRNA更敏感。

Figure S3. In silico sensitivity analysis of the model benchmark networks. Using the mathematical model of the synthetic networks as presented in Supplementary Figure 1, sensitivity analysis of YFP and mKate2 protein against mRNA species of each node was performed.

- This hypothesis intriguing scenario where the properties and outcome of signal propagation after custom perturbation experiments can be exploited toward distinguishing direct from indirect connectivities.
- 这个假设的场景,信号传播的性质和结果，之后可以利用定制的扰动实验区分直接和间接的连接性。

Modular Response Analysis
模块化响应分析

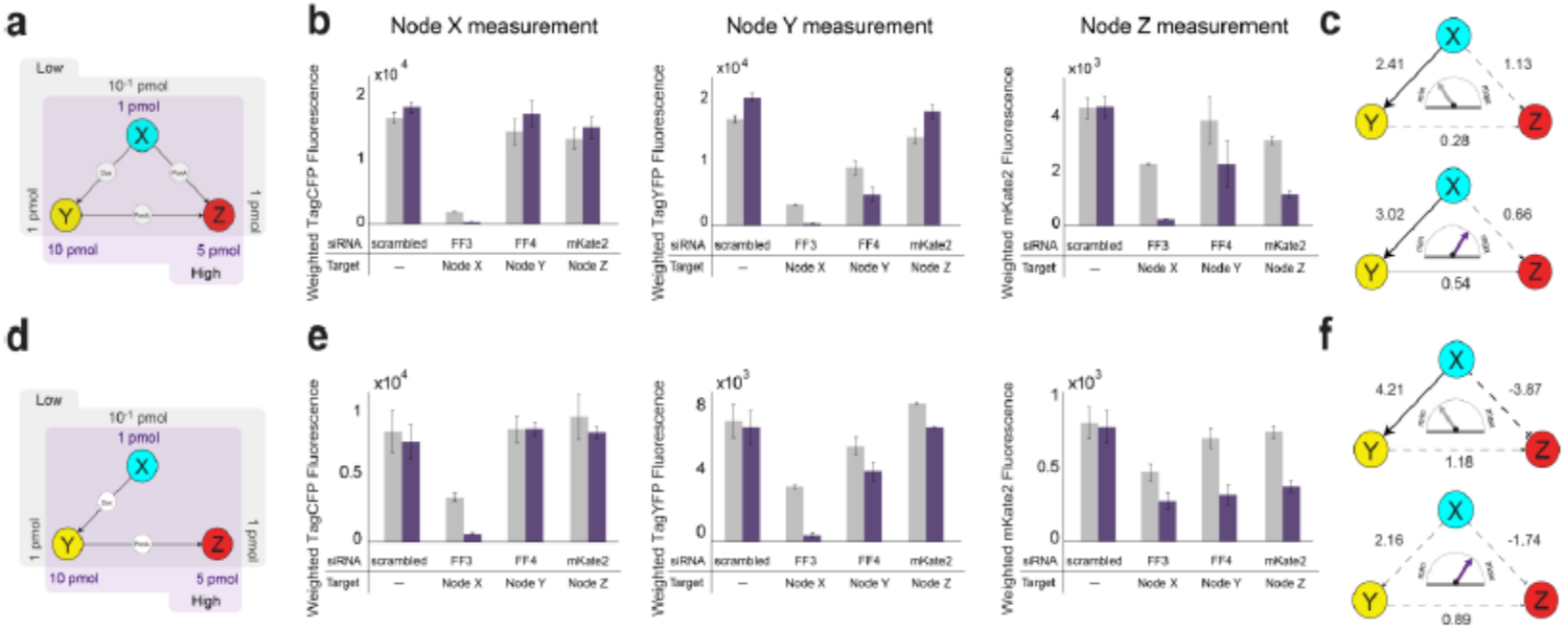




- We first tested the efficacy of siRNA and calibrated the perturbation dosage against the feed-forward loop architecture plasmid (SI Appendix, Figs. S4–S6; quantitative RTPCR results in SI Appendix, Fig. S7). Selecting the perturbation magnitude.

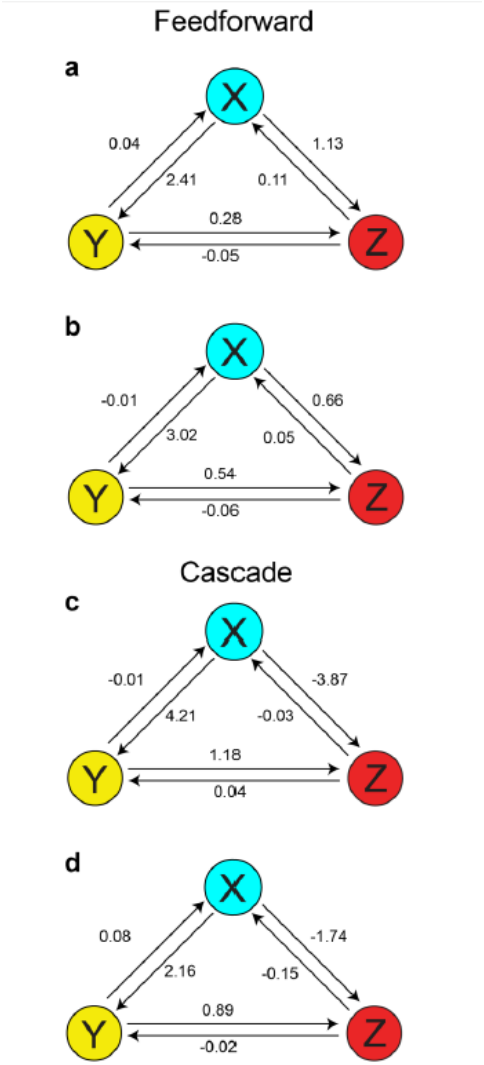
- 对前馈回路结构质粒测试siRNA的功效和校准摄动用量 (左123,S4-S6;定量RTPCR结果,图右1，S7)。选择摄动量的大小。

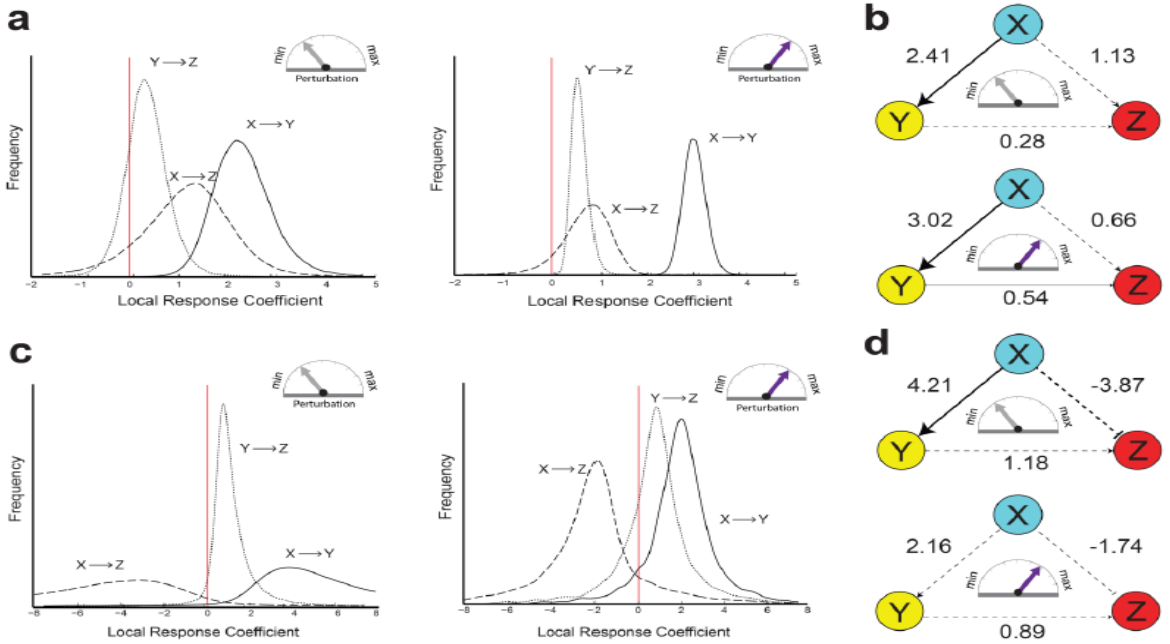
Modular Response Analysis
模块化响应分析



- Specifically, as illustrated in (上图SI Appendix, Fig. S8A), we selected 1pmol as “high” perturbation and 0.1 pmol as “low”. we performed a node-wise perturbation of the feed-forward circuit using the siRNAs that target each node supplemented with scrambled siRNA to control for the total mass.
- 选定1pmol作为高摄动量，0.1pmol为低摄动量。进行一个能有效摄动每个节点的不同siRNA对前馈结构node-wise摄动的影响实验。
- For node X, a decrease in TagCFP is observed only after direct perturbation; for node Y, a decrease in TagYFP is observed after perturbation of nodes X andY; and for node Z, a decrease in mKate2 is observed after perturbationof nodes X, Y, and Z (SI Appendix, Fig. S8B).
- 对X节点，TagYFP的减少只在对X的直接摄动下发生；对Y节点，TagCFP的减少在对X节点和Y节点的摄动下发生；对Z节点，mKate2的减少在对X节点，Y节点和Z节点的摄动下发生。

Modular Response Analysis
模块化响应分析





- Graphical representation of the complete circuit topology derived from population-level statistics
- （左1）源于群体统计数据的完整拓扑结构的图形化表示形式
- Monte Carlo error propagation analysis of modular response analysis
- （上1）蒙特卡洛的误差传播分析模块化的响应分析

Reverse Engineering of the Benchmark Topologies Using Resampled Single-Cell Data
使用单细胞数据重采样进行基准拓扑的逆向工程

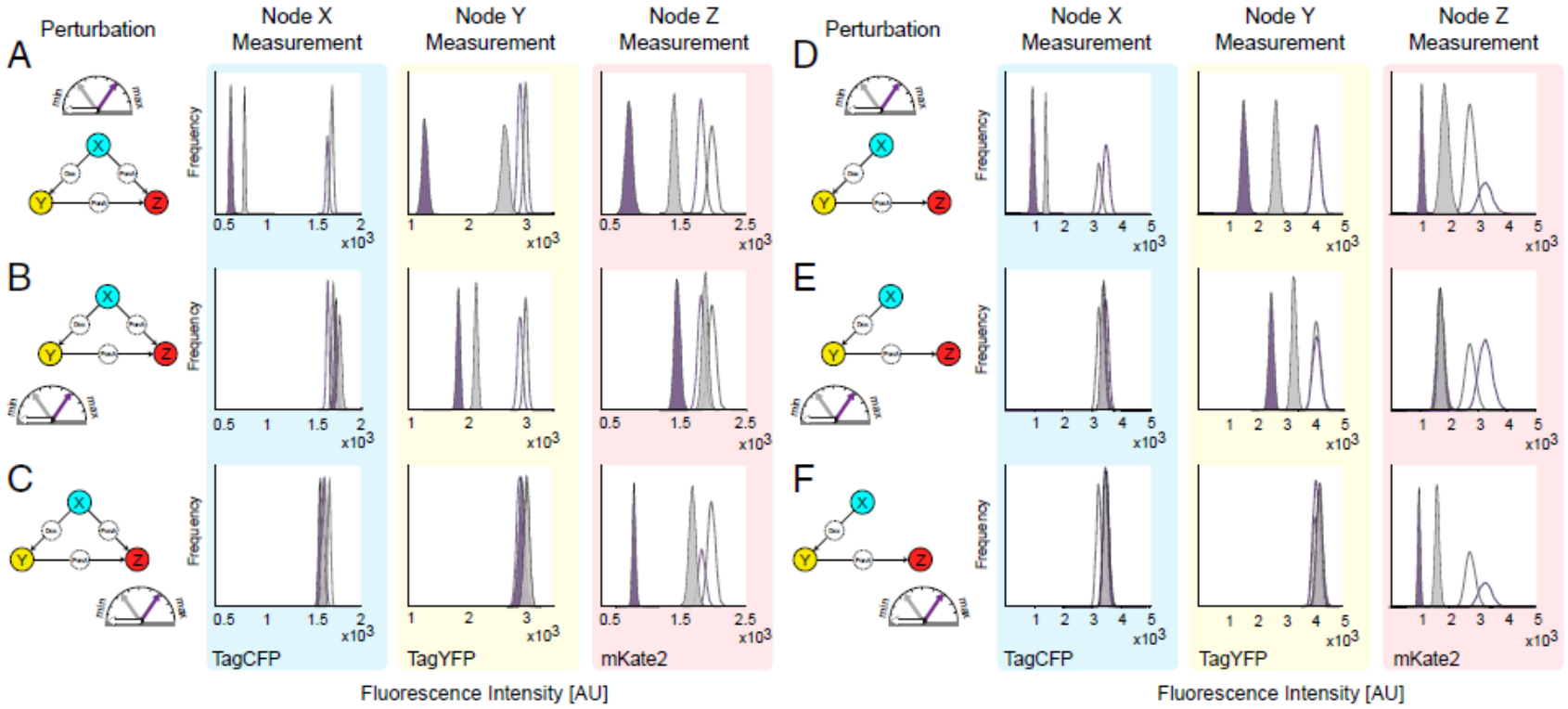


Fig. 4. Resampling the single-cell flow cytometry data after node-wise perturbations. Forty-eight h after siRNA perturbation, the expression level of the fluorescence reporters that represent nodes X, Y, and Z (TagCFP, TagYFP and mKate2, respectively) are measured using flow cytometry. To calculate the mean fluorescence of each population and the associated uncertainty, bootstrap resampling was performed. The resulting probability distributions of the resampled mean before perturbation (empty) and after perturbation (color-filled) are shown for the feed-forward loop (*A–C*), and the cascade (*D–F*). The colors of the peaks indicate the relative strength of the suppression applied (gray is used to indicate the low and purple the high perturbation). (*A*) The graphical representation of the X-node perturbation and the corresponding nodal responses using the feed-forward architecture. Probability distributions are composed of bootstrapped mean of the fluorescence reporters TagCFP, TagYFP and mKate2 (left to right) following perturbations to node X at two different siRNA concentrations. Color of the peak indicates the relative degree of suppression. (*B* and *C*) The graphical representation of the Y- and Z-node perturbations and the corresponding nodal responses using the feed-forward architecture. (*D–F*) Results from the same process using the cascade architecture.

Reverse Engineering of the Benchmark Topologies Using Resampled Single-Cell Data
使用单细胞数据重采样进行基准拓扑的逆向工程

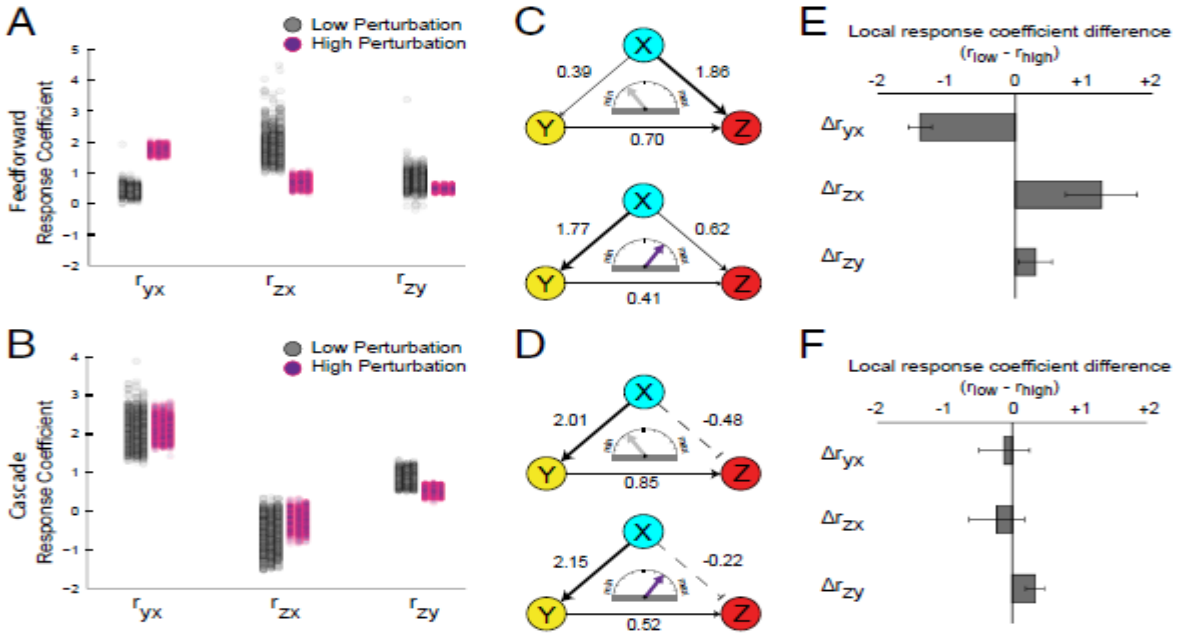


Fig. 5. Reverse engineering of the benchmark topologies using resampled single-cell data. (*A* and *B*) The complete reconstruction of the network with modular response analysis performed after two perturbations. For every set of subsampled means that make up the probability distributions, the MRA results along with the 95% confidence interval of the distribution are plotted as a 1D scatter plot. (*C* and *D*) The graphical representation of the reconstructed synthetic networks. (*E* and *F*) To probe the effect of response coefficient change due to perturbation magnitude shift we calculated the difference between coefficients of equivalent edges (*C* and *D*). The error bars were obtained using a propagation of error among the pair of local response coefficient distributions used to calculate this difference.

Reverse Engineering of the Benchmark Topologies Using Resampled Single-Cell Data
使用单细胞数据重采样进行基准拓扑的逆向工程

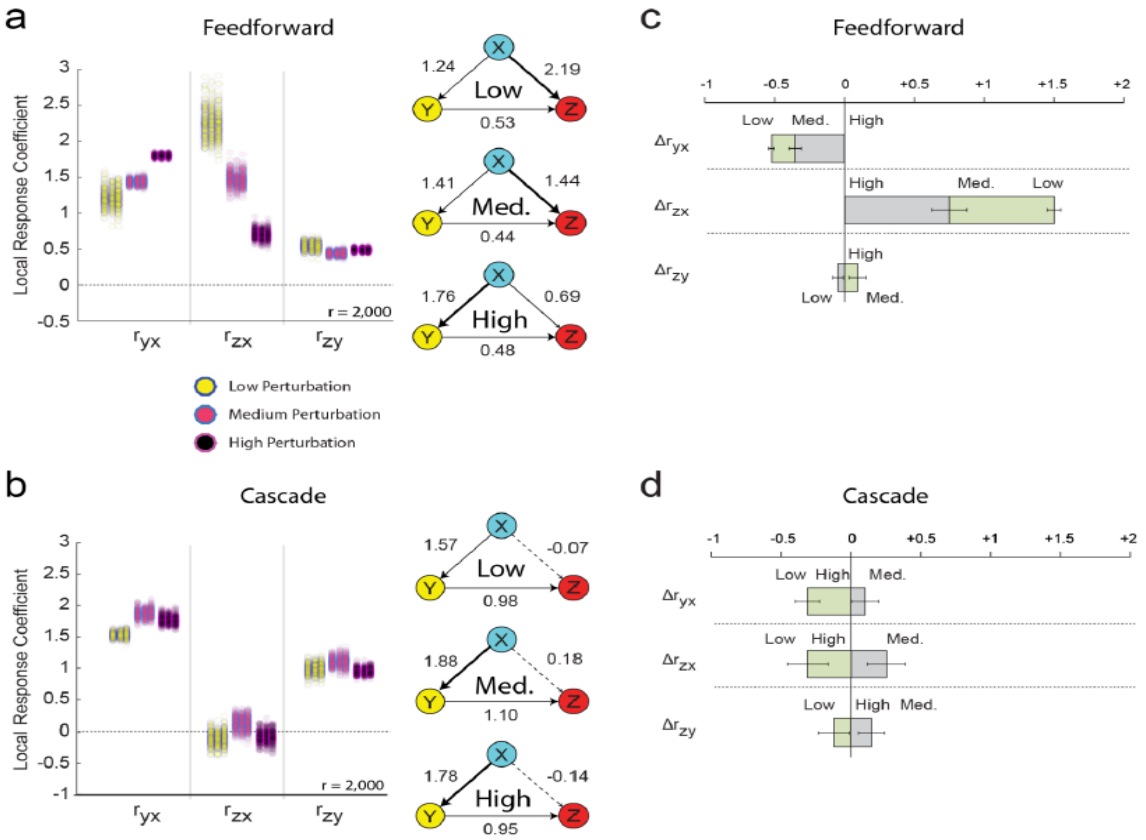


Figure S13. Reverse engineering of the benchmark topologies using resampled single-cell data. The complete reconstruction of the feedforward (a) and cascade (b) with modular response analysis performed after three different magnitudes of perturbations. For every set of subsampled means that make up the probability distributions, local response coefficients are calculated. This process cycle is performed 2,000 times, and the resulting local response coefficient distribution is plotted as a 1 dimensional scatter plot, and the corresponding graphical representation of the reconstructed synthetic networks with the mean values of these distributions are shown on the right (From top to bottom: low, medium and high perturbations, respectively). After reconstruction of the synthetic networks using three distinct sets of systemic perturbation, the change in response coefficients of equivalent edges are calculated after subsequent decrease in perturbation magnitude for feedforward (d) and cascade (d). The response coefficients recovered after the strongest perturbation sets ("High") are used as a reference point. The uncertainty associated with these values was obtained by error propagation based on the standard deviations of the original distributions.

Results

- Rooted in metabolic control analysis, Modular Response Analysis (MRA) uses steady state data obtained from node-wise perturbation to express the network in terms of pair-wise interaction sensitivities.
- 源于代谢控制分析，模块化的响应分析（MRA）使用从node-wise摄动得到的稳态数据分析表达出pair-wise交互作用敏感性。
- the experimental procedure consists of the following steps: (i) measure the steady-state xi corresponding to the unperturbed set of inputs pi, (ii) perform a perturbation to each pi individually and measure the new steadystate xi', (iii) calculate the global response coefficients using the steady-state data, and (iv) convert global response coefficients to local response coefficients by inversion of the global response matrix

- To increase our confidence in predictions we developed a technique based upon bootstrapping, an alternative to the sample statistics obtained from an aggregate population.
- 为增加可靠性,开发了一种基于bootstrapping（拔靴法)的另一个集合数据的样本分析法。
- Fig S13 To qualitatively probe our observations we developed a phenomenological model of the architectures. we analytically calculated the local response coefficients under low and high perturbations and we indeed confirmed the divergent shifts in interaction strengths.
- 定性分析所发现的结构，作者开发了一现象学模型架构。分析计算在低和高摄动的局部响应系数并确认交互作用的变化大小。

Conclusion

- 主要结论：在不同摄动条件下拓扑结构的变化可能会解决识别生物网络中直接和间接连接这一长期存在的问题。
- 创新点：使用多种验证分析手段，利用不同算法分析数据，增加了可信度，并从实验发现中自行开发现象学模型具有应用性。
- 启发：从模型设计到实验验证需要有缜密的思维，并学会用不同数学分析手段系统全面分析数据，通过实验结果联系可行的应用。
- 改进：设计的模体除了coherent feed-forward loop和cascade motif还可以有Incoherent feed-forward loop等其他结构，可以从其他模体结构分析可能的直接间接连接。