文章编号:1000-1735(2013)01-0092-06 doi:10.11679/lsxblk2013010092

基因调控网络研究进展

李庆伟, 全俊龙, 刘 欣

(辽宁师范大学 海洋生物功能基因与蛋白质组学研究所,辽宁 大连 116081)

摘 要:基因调控网络是一种连续而复杂多变的动态网络系统,是细胞内各种信号因子之间相互作用关系的整体表现. 近些年来,很多物种全基因组测序的完成、高通量实验技术的发展以及高性能分子生物学工具的应用,使得构建一个复杂和相对完整的基因调控网络成为可能,从而使绘制整个活细胞内各种基因表达的调控网络成为当前研究的热点. 作为系统生物学的核心领域,构建和分析基因调控网络将有利于我们更系统地剖析细胞的功能,更深刻地洞见生命的本质. 综述关于基因调控网络研究的基本原理和方法,为今后进行更深入的研究和探讨打下良好的基础.

关键词:基因调控网络;拓扑学特征;生物学实验手段;网络模型

中图分类号:Q-3 文献标志码:A

1 基因调控网络

1.1 基因表达调控的类型

细胞是所有活的生命体结构和功能的基本单元.一个生物体的所有细胞均包含有相同的 DNA,但它们的结构和功能却存在明显的不同.由于细胞内存在着一种复杂而有序的调控机制,致使基因在表达上有很多特异性,如组织特异性、细胞周期特异性和外界信号的响应特异性等.真核基因的表达受细胞核内和细胞核外多层次的调节,呈现多级调控,主要包括以下 2 种:一种是遗传调控(genetic regulation),另一种是表观遗传调控(epigenetic regulation).遗传调控又主要包括以下 5 种:(1)基因转录调控:调控基因表达的时间和速率;(2)转录后的加工调控:尤其是调控 mRNA 前体的选择性剪接;(3)翻译调控:选择哪些 RNA 将被翻译成蛋白质;(4)RNA 降解调控:干扰细胞中一些 RNA 的稳定性;(5)蛋白质活性调控:蛋白质形成后的激活、灭活、降解以及区室化.

其中,转录水平的调控是整个真核生物遗传信息传递和接收的过程中第一个具有高度选择性的环

节^[8]. 在遗传调控过程中,构成基因表达调控网络的基础是反式作用因子(trans-acting factor)与顺式作用元件(cis-acting element)间的相互作用. 表观遗传调控是指转录前基因在染色质水平上的结构调整,它是真核基因组一种独特的调控机制^[9]. 近年来,有关表观遗传学(epigenetics)研究已愈来愈成为基因表达调控的研究热点之一.

1.2 基因调控网络及拓扑学特征

基因调控网络(见图 1)^[10]描述了各个元素如基因、反式作用因子以及其他生物分子之间全局性的 互作关系,这些元素都是以一种团队的形式去完成单个蛋白所不能完成的生物学过程,如生物代谢、解 毒作用、对不同环境的响应等等. 对基因调控网络的研究,将会有利于我们更大范围地认识生物在 mRNA 水平上的生理状态;为我们提供治疗复杂疾病的有效方法;并使得通过比较不同生物的基因网络研究生物进化成为可能.

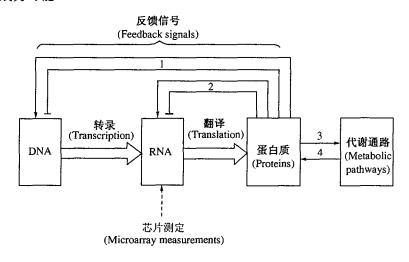


图 1 基因调控网络[10]

Fig. 1 Dynamic system of gene regulatory network

注:基因表达(转录、翻译后)的产物蛋白质形成多蛋白的复合体,彼此之间相互作用,并整合胞内、外各种信号调控代谢通路(箭头3)、接收(信号4)和释放(箭头1、2)各种反馈信号.

以元素之间相互联系的角度来分析,基因调控网络类似于一个由边(调控作用,如反式作用因子与靶基因启动子区特定区域结合)和节点(调控元素,如基因、蛋白质)组成的有向图结构.

网络中节点和边的数目庞大,类型多样,并在特定时期或状态下存在变化.其拓扑学特征主要包括:

- (1)基因调控网络有一个非常重要的特征,就是结构范围自由(scale-free structure)^[9],体现为一种分等级的结构. 网络之中大部分节点间只有一种或两种功能关系;而少部分节点,又称为中心节点(hubs),包含有非常多的功能关系,在支配细胞对特定刺激作出反应的过程中起核心作用. 这种分工使得大多数关联稀少的节点具有强的鲁棒性(robustness)^[10],但是又对中心节点的失活非常敏感.
- (2)基因调控网络组成一种超微小世界(ultra-small)^[9],任意两个节点的关联只需要很少的步骤.这种策略可能有利于信号的高效传递与整合.
- (3)含有很多不同类型的网络特征模块(motifs),描述了单个节点如何与其旁近节点相互作用(见图 2)^[9]. 在网络分析中,往往可以通过富集 motifs 的特征来划分相似的网络结构,减少网络模型的复杂性.

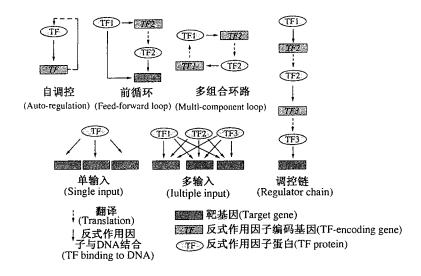


图 2 常见的转录调控网络特征模块[9]

Fig. 2 Motifs in transcriptional regulatory network

2 基因调控网络的研究方法与原理

目前关于基因调控的绝大多数问题还没有得到根本的解决,除了利用生物学方法以外,还得综合利用其他学科,如数学、统计学、神经网络、人工智能等方法在计算机上分析模拟表达调控机制,以便于深入研究基因调控网络^[5].利用现有的生物学背景知识和大量数据构建初步的基因调控网络,从而利用所得到的数据进行仿真和模拟;再根据模拟结果修改模型,并通过生物学实验验证,这样便可以得到一个较为完整的基因调控网络.

2.1 构建基因调控网络的生物学实验手段

构建精确可靠的网络模型首先需要生物学实验的数据^[11],即对网络中各元素及它们之间相互作用的监测,包括蛋白质合成水平、mRNA表达水平以及蛋白质-DNA的互作、常用的实验方法有如下几种:

2.1.1 基因表达分析

- 2.1.1.1 DNA 微阵列技术(DNA microarray) 随着功能基因组学研究的不断深入,迫切需要能同时监测大量靶基因表达的实验手段,迅速准确地在基因组水平上阐述不同生物组织或细胞中各种转录本的变化规律,因此 DNA 微阵列技术应运而生. 它是指在固体表面(玻璃片或尼龙膜)上固定成千上万的 DNA 克隆片段或人工合成的寡核苷酸片段,用荧光或其他标记的 mRNA、cDNA 或基因组 DNA 探针进行杂交,从而同时快速高通量地检测基因表达状况,其基本原理是基于 Southern 杂交或斑点杂交技术. 但是高通量的基因表达实验往往存在实验噪声,这种噪声在一般的微阵列实验结果中要占整个信号的 20%~30%[12].
- 2.1.1.2 时间序列基因表达实验(time series gene microarray)^[13] 这种技术是测定在不同的时间点对细胞周期以及特定状态条件下的基因表达情况,寻找相应的调控基因和靶标基因.时间序列基因表达数据在通常的情况下要比同样大小的静态基因表达数据含有更大的推导基因调控网络的信息量.但因为成本非常昂贵,大多数实验室测量的样本相对于庞大的基因数来说还差得很远,而且实验噪声和转录翻译时延这些影响结果分析的问题始终没有得到很好的解决,以上这些因素都对模型的构建提出了新的要求.

2.1.2 蛋白质表达分析

目前,蛋白质表达分析最主要的实验手段为蛋白质微阵列技术(protein-microarray).蛋白质微阵列技术^[9]又称为蛋白质芯片技术,是一种高通量的蛋白功能分析技术,可用于蛋白质表达谱分析,研究

蛋白质与蛋白质的相互作用,甚至 DNA-蛋白质、RNA-蛋白质的相互作用,筛选药物作用的蛋白靶点等.常用的蛋白质芯片是一种固定了保持天然活性的蛋白质微阵列,广义上,将能与蛋白质特异性结合的 DNA 和 RNA、糖类、合成多肽,其他能够从复杂的混合物中特异性地捕获目的蛋白的小分子物质的微阵列也称为蛋白质芯片.蛋白质芯片技术的应用使得对数以千计的蛋白质进行高通量、平行分析成为可能,将极大地推动蛋白组学的发展.

2.1.3 蛋白质-DNA 互作分析

- 2.1.3.1 凝胶阻滞实验(electrophoretic mobility shift assay, EMSA)^[9] 这是一种体外分析 DNA 与蛋白质相互作用的凝胶电泳技术,是一种十分有效的分离纯化特定 DNA 结合蛋白质的实验方法. 其基本原理是,蛋白质与 DNA 结合后将大大增加分子质量,而凝胶电泳中,裸露的 DNA 朝正电极移动的距离与其分子量的对数成反比,没有与蛋白质结合的 DNA 片段跑得快,而与蛋白质形成复合物的 DNA 由于受到阻滞而跑得慢.目的 DNA 片段与细胞提取物混合后,如果其在凝胶电泳中的移动距离缩短了,就证明它与提取物中的某种蛋白质分子发生了结合. 此外, EMSA 还可以研究与蛋白质相结合的 DNA 序列的特异性.
- 2.1.3.2 染色质免疫共沉淀(chromatin immunoprecipitation, ChIP)^[9] 这是一种基于体内分析发展起来的方法,在过去十多年已经成为表观遗传信息研究的重要方法.它为研究者判断在细胞核中基因组的某一特定位置会进行哪种组蛋白修饰提供重要的参考数据. ChIP 除了能使我们了解染色质及其相结合的因子之间的相互作用之外,还可以用来研究组蛋白之间的各种共价修饰和基因表达的关系. 它的原理是在保持组蛋白和 DNA 联合的同时,通过运用对应于一个特定组蛋白标记的生物抗体,染色质被切成很小的片段,并沉淀下来. 无论是考察单个启动子在很短时间内的变化,还是在整个人类基因组中考察某个转录因子,染色质免疫共沉淀技术都能使我们了解在自然状态下基因是如何被调控的.
- 2.1.3.3 全基因组定位分析(genome-wide location analysis) 全基因组定位分析^[3]是一种新的研究蛋白质与 DNA 片段结合的分子生物学技术,综合了染色质免疫共沉淀以及 DNA 微阵列分析技术,可以高通量研究转录因子与全基因组的基因间区序列结合的亲和力,从实验结果可以发现转录因子的调控基因,进而明确基因之间的相互关系.

2.2 基因调控网络的常见模型

伴随着高通量实验数据的产生,利用信息科学和计算机科学构建基因调控网络模型,为我们更进一步的探索生命复杂的过程发挥了不可替代的巨大作用.目前,基于不同的数据类型和理论模式,发展起了若干种建模方法,综合起来主要包括微分方程模型、布尔网络模型、线性网络模型以及贝叶斯网络模型等.

2.2.1 微分方程模型

较早利用微分方程系统作为基因调控网络模型有 Chen^[14]等人,他们用线性常微分方程来描述整个网络系统.该模型建立了很多假设,其中包括每个基因线性转录函数以及翻译产物对转录率的反馈机制,同时考虑了基因的转录、翻译和 RNA、蛋白的降解,并将以上参数共同进行分析.其主要的优点在于强大灵活,有利于描述基因网络中的复杂关系,而且特别适用于有周期性表达的基因^[15].而其缺点在于生物学实验过程中蛋白质的含量比较难以获得,另外也忽略了其他调控物的影响.

2.2.2 布尔网络模型

布尔网络模型^[15]最早由 Kauffman 引入,它是最简单的调控网络模型.这种模型把基因定量为两种不同的状态("开"和"关"). 状态"开"表示一个基因转录表达,状态"关"则代表一个基因未转录. 如果布尔网络模型进一步推广,其可转变为时序布尔网络模型 TBN(temporal Boolean networks)^[13],这有利于处理多于一个单位时间跨度的基因间的依赖性. Shmulevich 等人将马尔可夫链(Markov Chains)和布尔网络结合起来,以便于处理概率框架下的不确定性,同时引进了概率布尔网络模型 PBN(probabilistic boolean networks)^[16],这样就比标准布尔网络模型有了更多的优越性.

布尔网络模型较为简单,然而由于它是一种离散的数学模型,不能很好地反映细胞的实际情况,而在连续网络模型中,各个基因的表达是连续的数值,并且通过具体的数值表示一个基因对其他基因的影响. 2.2.3 线性网络模型

线性组合模型[17]是一种连续网络模型,基本表示形式为:

$$X_i(t+\Delta t) = \sum_{j=1}^{j} w_{ij} X_j(t)$$

其中, $X_i(t+\Delta t)$ 表示基因 i 在 $t+\Delta t$ 时刻的表达水平, $X_j(t)$ 表示基因 j 在 t 时刻的表达水平,而 w_i 代表基因 j 的表达水平对基因 i 的影响,正值表示激活,负值表示抑制,0 表示无影响。如果调控网络参数已给定,那么这个公式中描写的是基因网络的动力学,在工程学中是常见的;如果基因表达时间序列 $X_j(t)$ 已给定,要估计调控网络参数 w_{ij} ,则成为反向工程问题(reverse engineering)。已有将"奇异值分解"方法应用于此反向工程问题的解决 [18-19]。然而,目前的测量均限于基因数要远远超过时间点数,所以它们的解集不是唯一的,因而难有生物学意义。为了解决这个问题需要进行全局聚类,例如在基因共表达的研究中,往往把具有相似表达谱的共表达基因归为一类。

2.2.4 贝叶斯网络模型 贝叶斯网络^[15]是一种重要的概率模型,由条件概率分布和网络结构两部分组成.它的优点是:可以通过借助问题领域自身结构特征和变量间直接影响的局部性,同时使用条件独立的数学概念,将聚合概率分布的计算问题,分解为若干局部条件概率分布的计算问题;具有良好的学习能力;能很好地处理隐变量和数据缺失问题.其不足之处是:不能明显地表现基因调控网络的动力学特征.然而利用动态贝叶斯网络 DBN(dynamical Bayesian networks)^[19]可以在一定程度上克服这种不足. DBNs善于处理非线性关系和由随机现象引起的不确定性,具有很好的概率推理能力和知识表达能力.

3 结论及展望

基因调控是由众多分子组成的复杂网络,在不同生物、不同组织、不同时间和环境条件下有很大的灵活性和鲁棒性[20].利用生物学实验的手段可以直接获取真实的调控数据,但无法对整个网络进行跟踪观察,只能以点代全,并且实验的噪声以及表达时延问题始终没有得到较好的解决;而在实验的基础上建立基因调控网络的模型,可以更方便地研究网络的动态变化.但各种建模方法都有其局限性[5],比如:(1)很多模型由于其本质上具有确定性,所以有时候会把内部的遗传功能和相互作用理解为逻辑规则,而在真实生物系统中,自组织性、自发性则更为重要;(2)基因调控机制是一个"实时"的应激过程,在时间上是连续的,但是计算模拟这种刺激过程是十分昂贵的,只能划分成若干分离的时间单元进行模拟;(3)为了建立真实的动态网络,我们需要整合各种类型的数据信息,同时又要简化系统的复杂度,一种解决办法就是考虑真实网络系统的模块化和鲁棒性,运用一种组合式的方式,但对生物鲁棒性和模块化的校对又是非常难的.尽管现在网络模型存在着不足,但研究前景相当广阔.随着转录组学、蛋白质组学、相互作用组学、代谢组学等高通量实验技术的不断提高以及数学、信息学、计算机科学、分子生物学等学科的不断发展,将算法研究和实验研究相结合,完善建模-模拟-实验这个循环,将会进一步了解生物学中真实基因调控网络的机制,并在生命科学的研究中发挥越来越大的作用.

参考文献:

- [1] MCADAMS HH, SHAPIRO L. Circuit simulation of genetic networks[J]. Science, 1995, 269(5224):650-656.
- [2] ARKINN A, ROSS J, MCADAMS HH. Stochastic kinetic analysis of developmental pathway bifurcation inphageλ-infected Escherichia coli cells[J]. Genetics, 1998, 149(4):1633-1648.
- [3] WYRICK JJ, YOUNG RA. Deciphering gene expression regulatory networks[J]. Current Opinion in Genetics & Development, 2002,12(2):130-136.
- [4] SHINOZAKI K, YAMAGUCHI-SHINOZAKI K, SEKI M. Regulatory network of gene expression in the drought and cold stress

- responses[J]. Current Opinion in Plant Biology, 2003, 6(5): 410-417.
- [5] DAVIDSON E, LEVIN M. Gene regulatory networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(14):4935.
- [6] DAVUDSON EH, LEVINE MS. Properties of developmental gene regulatory networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(51): 20063-20066.
- [7] EDWARD R. Genetic Regulatory Networks[J]. EURASIP Journal on Bioinformatics and Systems Biology, 2007, 2007 (1): 17321-17322.
- [8] DEWEY T,GALAS D. Power Laws, Scale-Free Networks and Genome Biology[M]. Berlin; Springer, 2006; 106-122.
- [9] GELENBE E. Bio-Inspired Computing and Communication[M]. Berlin: Springer, 2008:19-32.
- [10] TREPODE NW, ARMELIN HA, BITTNER M, et al. A robust structural PGN model for control of cell-cycle progression stabilized by negative feedbacks[J]. EURASIP Journal on Bioinformatics and Systems Biology, 2007, 2007(1):73109.
- [11] CHEN T, FILKOV V, SKIENA SS. Identifying gene regulatory networks from experimental data[J]. Parallel Computing, 2001, 27(1-2):141-162.
- [12] ROCKE DM, DURBIN B. A model for measurement error for gene expression arrays[J]. Journal of Computational Biology, 2001, 8(6):557-569.
- [13] 金志超,吴骋,高青斌,等.基于时间序列表达数据基因调控网络模型的研究进展[J].第二军医大学学报,2008,29(9).
- [14] CJEN T, HE HL, CHURCH GM. Modeling gene expression with differential equations[J]. Pacific Symposium on Biocomputing, 1999.4.29-40.
- [15] 刘化锋,王文燕,基因转录调控网络模型[J].山东大学学报:理学版,2006,41(6):103-108.
- [16] MARSHALL S, YU L, XIAO Y, et al. Inference of a probabilistic Boolean network from a single observed temporal sequence[J]. EURASIP Journal on Bioinformatics and Systems Biology, 2007, 2007(1):32454-32468.
- [17] D'HAESELEER P, WEN X, FUHRMAN S, et al. Linear modeling of mRNA expression levels during CNS development and injury[J]. Pacific Symposium on Biocomputing, 1999, 4, 41-52.
- [18] HARTEMINK AJ. Reverse engineering gene regulatory networks[J]. Nature Biotechnology, 2005, 23(5):554-555.
- [19] LUNA IT, HUANG Y, YIB Y, et al. Uncovering Gene Regulatory Networks from Time-Series Microarray Data with Variational Bayesian Structural Expectation Maximization[J]. EURASIP Journal on Bioinformatics and Systems Biology, 2007, 2007(1): 71312-71325.
- [20] 卢欣,李衍达. 基因调控过程中的典型控制环节[J]. 自动化学报,2000,26(5):637-644.

Progress in the study of gene regulatory network

LI Qingwei, QUAN Junlong, LIU Xin

(Institute of Marine Genomics & Proteomics, Liaoning Normal University, Dalian 116081, China)

Abstract: Gene regulatory network is a continuous and complex dynamic system which is also the overall performance of the intracellular gene interaction. In recent years, the accomplishment of genome sequencing for many species, the development of high-throughput experimental technology, and the application of molecular biology tools have promoted the construction of complex genetic regulatory network. In this case, the researches on the gene regulatory network in living cells have become a hot-spot. As the core of systematic biology, construction and analysis of gene regulatory network will help us to understand the function of cells and the nature of life more systematically. This paper reviews the basic principle and main researching methods of gene regulatory network which will provides theoretical foundation for the further research in this field.

Key words: gene regulatory network; topological features; biological experiment method; network model