

# 中国计算系统生物学新近进展及展望

## 摘要

计算系统生物学是近年来新兴的一门交叉学科，受到国内外众多领域研究学者的关注和重视。本报告侧重于运筹学的若干分支与计算系统生物学领域的交叉研究，阐述近年来的一些研究进展，评述其发展趋势以及展望。

## Recent Development and Future Prospect of Computational System Biology in China

Recent advances in genetics, molecular biology, and impetus of the genome projects have motivated the integration of biological, physical, computer and mathematical sciences. Analysis and interpretation of large and complex genomic datasets now require interdisciplinary expertise and the close collaboration of biologists, statisticians, computer scientists and mathematicians. As a result, the computational systems biology is emerging as an interdisciplinary subject in recent years, which attracts attentions from a broad range of areas and a lot of domestic and oversea researchers. In this chapter we report the interfaces between computational systems biology and several branches of operations research including DNA sequencing, the prediction of protein structures, inference of biologic molecular networks and diagnosis of complex diseases. Specifically, we mainly review the recent advances, development trends, and outlook of these areas in near future.

## 一、引言

进入 21 世纪后，随着数学、物理学、计算机科学等定量学科的交叉与渗透，生命科学的研究进入了一个崭新的革命性时代，也带动了计算系统生物学这门新学科的兴起。计算系统生物学是在计算生物学与生物信息学的基础上发展起来的一门新兴的交叉科学，并且三者之间没有明确的界限，它们的研究内容往往交织在一起。它是以核酸、蛋白质等生物大分子及其构成的复杂系统为研究对象，以数学、物理学、计算机科学等为主要手段，构建精确模型，开发和运用更有效的算法、数据结构等，通过整合海量生物数据，使生命科学研究模式化、定量化，从而为探索生命科学与医学中的重大问题提供重要的解决工具。

计算系统生物学研究的兴起不仅具有重大的科学意义，而且具有巨大的社会和经济效益，在国际上受到高度重视，美国和欧洲一直走在该研究领域的前列。1997 年，国际计算生物学学会在美国成立。1999 年，世界上第一个系统生物学家在美国西雅图成立。2004 年，美国卫生研究院启动了“生物信息学和计算生物学”计划，并着手建立了数个“国家生物医学计算中心”。近年来，加州大学、斯坦福大学、德克萨斯大学、芝加哥大学、威斯康星大学等机构纷纷成立了计算生物学中心，德国、法国、澳大利亚、意大利等国也纷纷建立了计算生物学研究机构。2005 年 10 月 13 日，中国科学院上海生命科学研究院计算生物学研究所（中科院-马普学会计算生物学伙伴研究所）正式揭牌，成为我国第一个计算生物学研究所，标志着我国对计算生物学研究的重视达到前所未有的高度。2011 年 9 月 2 日，中国运筹学会计算

系统生物学会正式成立，进一步促进和推动了国内计算系统生物学的发展和国际交流与合作。

本报告仅侧重于运筹学科的若干分支与计算系统生物学领域的交叉研究。首先将从高通量生物实验数据蓬勃发展的角度回顾计算系统生物学交叉研究的缘起、内在驱动、问题和主要研究手段。然后结合计算系统生物学的几个代表性研究分支阐述近年来的一些研究进展，对比、分析该交叉领域的国内外发展现状，评述了其发展趋势以及展望。但限于种种原因，很难全面与精准，不周之处，请读者批评与谅解。

## 二、计算系统生物学的发展

计算系统生物学的最终目的在于理解生命现象背后深刻的自然规律和逻辑，它本质上是由高通量生物数据的发展驱动的。众所周知，我们正处在一个大数据的时代，这种变化尤其体现在生物学领域，芯片、测序等代表性技术生成了海量的数据。一方面这些数据的处理、信息提取超出了传统生物学家的研究范围，驱动了生物学和数学、计算机科学等其他学科的交叉。目前，一些大型的生物学实验室都有做数据分析的计算人员，集中显示了这一趋势。另外，在更高一点的层面，这些不同侧面的观测数据如何整合转化为系统层面的知识，为生物学研究提供更多的定量研究，是计算系统生物学中“系统”一词的内涵，揭示了其更为持久的内在驱动力。运筹学作为数学的一个分支，是一门研究如何“更好”的科学。如何更好地挖掘、解释生物学数据，同时如何更好地用数学模型来探索复杂生命现象，都对运筹学提出了新的问题和挑战。生物学和运筹学的交叉研究，自然而然且生命力旺盛。

计算系统生物学涉及的课题很多，用到的运筹学的方法也因具体问题而异，这里我们以近 10 年来生物数据发展的历程作为主线，回顾一下计算系统生物学的发展及运筹学在其中的作用。

### （一）第一阶段：基因组及生物分子层面数据的积累

人类基因组计划是 20 世纪 80 年代提出，于 2005 年完成，是继曼哈顿计划和阿波罗登月计划之后人类科学史上的又一个伟大工程，已积累了人类和其他许多物种基因组的大量测序数据。这直接推动了基于基因组学的生物信息学研究，也标志着生命科学研究全面进入后基因组时代。一些随之而来的基本问题包括：基因（编码蛋白的 DNA 序列）的识别；基因对应蛋白的结构和功能；对一个群体中的 DNA 序列的差异（称为多态性），如单核苷酸多态性（基于单个核苷酸上的差异）的研究；及非编码区生命信息的挖掘等。

这一时期的生物信息学采用了大量运筹学的传统理论和方法，如动态规划、图论、马氏过程、人工神经网络、非线性能量极小化模型等用于研究序列比对、蛋白质结构比对、蛋白质结构折叠和单体型推断与组装等问题。传统运筹学方法遇到的挑战是海量数据和高度非线性。这部分研究工作还远远没有完成，例如有效获取个人的单核苷酸多态性单体型，以实现基于基因组信息的个性化医疗。

### （二）第二阶段：组学及生物分子网络层面数据积累

全基因组范围内的数据明确了生物系统内部的组成结构，直接促进了系统生物学研究（亦称为后基因组时代的研究）。这些研究的指导思想是：不是单个基因（蛋白），而是一组基因（蛋白）在环境的制约下共同产生某一生物体功能（或导致一种疾病）。组学数据是多层次的，包括 DNA 层次上的基因组、转录层次上的转录组、蛋白质层次上的蛋白质组以及相互作用层

面的相互作用组等高通量生物数据。基因组是某种生物所携带的遗传信息的总和。其主要测量技术是 DNA 测序，包括传统的电泳测序和最新的二代测序。转录组则是细胞在某个特定时刻所有转录的 mRNA 总和。其主要测量技术是 DNA 微阵列技术，利用碱基互补配对的特性可以同时检测几千甚至上万基因的表达式。最近兴起基于二代测序技术的 RNA 深度测序相对于 DNA 微阵列技术可以更精确地检测 RNA 的绝对含量。蛋白质组是细胞在某个特定时刻所有蛋白质的总和。蛋白质质谱技术，利用不同多肽链质量和带电量不同对多肽链的种类和含量进行检测，可一次性检测样本中所有多肽链的种类和含量。相互作用组包括蛋白质相互作用组、蛋白质-基因相互作用组等。蛋白质相互作用组是细胞内所有蛋白质相互作用的总和。酵母双杂交可高通量、快速、直接地测得两个蛋白质是否发生相互作用。染色质免疫沉淀芯片技术和染色质免疫沉淀测序技术则用来检测转录因子蛋白和 DNA 是否发生相互作用。除此之外，组学数据和概念还包括代谢组、表观组、疾病组、外显子组等等。这些“组学”数据都是全基因组范围、高维度、多层面、异源、异质、多尺度、高噪声的，分析、集成这些数据对数学建模和算法提出了巨大的挑战。

对组学数据的研究集中体现在对基因调控机制、代谢机制的生物分子网络的研究。网络成为这一研究的基本工具，所以也称为“网络生物学”，许多生物网络具有幂率性质、小世界性质和其他不同于传统随机网络的性质，对这些网络的静态结构和动态演化的研究以及对生命系统复杂现象的解释才刚刚开始。例如由于生物在进化过程中的自然择优性，最优化原则和模型成为研究中可选的基本框价。一大类数学上的逆问题出现在分析和集成不同来源、不完全、高噪声和高错误率和生物数据来推断或重建生物分子网络及生物模型结构的过程中。主要的挑战是如何最优利用稀少数据和误差数据。例如，由远远少于动力系统维数的时间观察点上得到的数据来精确推断动力系统的结构。

### （三）第三阶段：生物系统层面数据积累

主要包括以研究生物系统层面的特性和以人类健康为直接目的的生物数据积累。例如最近刚部分完成并引起轰动的 DNA 元素百科全书计划。该项目获得了迄今最详细的人类基因组分析数据，这是“人类基因组计划”之后国际科学界在基因研究领域取得的又一重大进展。人类基因组项目发现人的基因组中仅有 1.5% 的序列是编码蛋白质的，以前认为其余 98.5% 的序列是“垃圾”。该项目的一个主要目的就是去分析那些“垃圾”的功能。经过这个合作项目的初步努力，现在科学家们可以说“80% 的基因组是有功能的，集中体现了生物系统的复杂性”。该项目于 2003 年启动，来自五个国家（美国、英国、西班牙、日本和新加坡）的 32 个研究机构参与了此次项目，历时五年，耗资 1.5 亿美元。另一个代表是针对癌症这种复杂疾病的数据。美国政府发起的癌症和肿瘤基因图谱计划，试图通过应用基因组分析技术，特别是采用大规模的基因组测序，将人类全部癌症（近期目标为 50 种包括亚型在内的肿瘤）的基因组变异图谱绘制出来，并进行系统分析，旨在找到所有致癌和抑癌基因的微小变异，了解癌细胞发生、发展的机制，在此基础上取得新的诊断和治疗方法，最后可以勾画出整个新型“预防癌症的策略”。2005 年 12 月 13 日，这一项目由美国国家癌症和肿瘤研究所和国家人类基因组研究所联合进行，预计耗资 1 亿美元。除此之外，这些系统层面的数据也包括流行病学研究、新的药物概念和药物设计流程、人类疾病的重新分类和基于基因组信息的疾病诊断，以及遗传疾病的根本治疗，还包括为微生物学对环境和清洁能源的贡献等。

这一阶段的数学工具需要包括运筹学科研人员在内的数学工作者在传统理论和方法的基础上进行创新。目前，人们还很难简单地描述主要的数学工具，因而解决这些全新的问题给数学与运筹学研究人员提供了难得的机会。

### 三、近年来代表性研究成果及未来发展方向

下面我们将从运筹学与生命科学交叉研究的角度介绍计算系统生物学的几个代表性研究领域和成果，其中前三部分属于生物分子层面的研究，第四部分属于生物分子网络方面，第五部分侧重复杂疾病即系统层面。在每一部分，我们通过阐述近年来的一些研究进展和现状，希望能帮助运筹学工作者了解和体会其中的主要问题、交叉研究的基本流程和思路。此外，每一部分我们都简单介绍了未来的研究方向和一些重点与热点问题，希望广大运筹学工作者积极投身到运筹学与生命科学交叉研究中，促进国内在相关领域迅速赶上国际研究的前沿。

#### (一) 运筹学与 DNA 测序及相关问题

DNA 测序技术是当代生命科学在 20 世纪 90 年代以后得以高速发展的重要基础。在以人类基因组测序完成作为标志的后基因组时代，DNA 测序仍然占有极其重要的地位。

由于生物实验技术的限制，测序实验只能得到较短的序列片段（从几个碱基到几百个碱基）。因此各种测序技术都涉及到如何从测序实验得到的较短的序列片段来重构出原始序列，这个问题往往都被转换为图论问题来求解。其中使用最为广泛的是 de Bruijn 图。De Bruijn 图最早被用于基于微阵列芯片的杂交测序方法中<sup>[1,2]</sup>。杂交测序方法是上个世纪末提出的一个新的高效测序方法，也是基因芯片技术的一个非常有潜力的应用领域。由于生物芯片技术的限制（体积和密度等），杂交测序方法中的重构计算问题（即从生物实验数据重构出目标 DNA 序列）非常复杂。将生物实验得到的序列片段作为图的顶点，杂交测序方法重构问题本质上是一个哈密尔顿路径问题，这是难以求解的 NP 完备问题。而通过使用 de Bruijn 图，杂交测序重构问题可以转化为 de Bruijn 图上的欧拉路径问题，存在多项式时间算法<sup>[2]</sup>。但是如果考虑到实际问题中存在生物实验错误和 DNA 序列的重复子序列，杂交测序重构问题仍然是一个 NP 难问题。欧洲运筹学会联合会的计算生物学、生物信息学和医学工作组的负责人、波兰科学院院士 Jacek Blazewicz 在杂交测序问题上进行了系统的研究<sup>[3-8]</sup>。国内的运筹学研究人员在该问题上也取得了很好的结果<sup>[9,10]</sup>。值得一提的是，虽然由于二代测序技术的出现，使得杂交测序方法最终没能得到广泛应用，但是杂交测序方法的许多研究成果，特别是基于 de Bruijn 图的许多数学模型和算法，都是目前二代测序算法中的核心和基础<sup>[7,11]</sup>。

动态规划在序列分析和比较中有着十分广泛的应用。最著名的莫过于三十多年前提出的用于局部序列比对的史密斯-沃特曼算法<sup>[12]</sup>。而更早的时候，动态规划就已经被用于全局序列比对<sup>[13]</sup>。编著<sup>[14]</sup>中就总结了动态规划方法在序列比较上的应用。对于两两序列比对，基于动态规划的算法可以在多项式时间  $O(n^2)$  内得到精确的最优比对。而对于  $k$  条序列的多序列比对，动态规划算法的时间复杂度为  $O(n^k)$ 。当  $k$  较大时，基于动态规划的算法就只能求解较小规模的问题。目前多序列比对还没有很好的精确算法，大多采用各种启发式算法来得到近似解。例如将多序列比对转化为一系列的两两序列比对问题。在这类方法中，通常需要将前期的多序列比对的结果用一个矩阵表示，称为特征谱，然后通过动态规划算法对两个特征谱进行比对。基于特征谱的方法会造成多序列比对信息丢失，进而影响最终的比对结果。论文<sup>[15]</sup>提出了采用偏序图来表示和保存多序列比对的结果，并设计了基于偏序图的动态规划算法，能避免传统的缺陷。除了 DNA 和蛋白质序列比较，动态规划在蛋白质和 RNA 结构比较与预测等问题上都有着广泛的应用<sup>[16-18]</sup>。

从 DNA 测序得到的片段数据中得到的单体型信息在遗传学和医学研究中占有非常重要的地位，特别是在复杂疾病的关联基因查找方面。但是由于测序技术、规模和费用等原因，直接用实验测定单体型的方法还不成熟，目前主要用数学模型和计算机算法来进行单体型的推断和组装。由于序列的离散特性，这些问题都属于组合优化问题，而且大都已经证明是

NP-难甚至 APX-难问题<sup>[19,20]</sup>。参数计算理论被认为是较为成功的解决典范之一。图的顶点覆盖问题的参数化被证明对于解决单体型组装问题有效<sup>[21]</sup>。利用各种组合优化模型，国内外的运筹学者<sup>[22-30]</sup>提出了许多针对小规模问题的精确算法和针对大规模问题的近似算法和启发式算法，用于求解单体型推断和装配问题。

该方向未来的一些研究重点与热点问题包括：

(1) 在 DNA 测序中，主要是提高算法的效率和处理数据的规模。目前的二代测序仪的数据通量越来越大，对后续数据处理的计算机软件和硬件提出了更高的要求。在计算机硬件没有太大变化的情况下，就需要在测序数据的分析处理算法上寻求突破，这需要更多数学家和运筹学家的介入。

(2) 在 RNA 测序中，目前的算法还不是十分理想。特别是复杂生物中的大量可变剪切现象，增加了测序数据的处理难度。对于已知基因组序列的物种，人们往往使用映射的方式来分析测序数据，而对于未知基因组序列或者部分知道序列的物种，则需要采用从头拼接算法。这也是目前 RNA 测序的一个研究热点。

(3) 宏基因组测序算法是未来测序算法研究的一个重要方向，需要考虑不同物种的基因组之间的干扰，以及在完成序列拼接的同时将不同的物种进行聚类或者区分。

(4) 序列比对的研究已经十分成熟，但是多序列比对仍然是一个瓶颈。此外，RNA 测序算法中需要用到大量的短序列比对，这些短序列比对对于时间和空间都有着非常高的要求，这方面的研究还有较大的需求和提升空间。

(5) 目前的单体型组装和推断研究，都是独立于测序技术进行的，没有和测序算法有机地结合起来。在测序数据分析处理算法中加入单体型信息和约束，既可以提高测序数据的处理精度，也可以极大地提高单体型组装的效率。这将是单体型研究领域的一个可能的方向。

## (二) 运筹学和蛋白质结构预测

蛋白质结构预测<sup>[31]</sup>是分子生物学领域最具挑战性的问题之一，它直接关联着特定的生物功能，具有重大的科学意义和实际应用。其本质是如何从蛋白质的氨基酸序列出发预测它的构象问题。从数学上看，这是一个三维空间的距离几何问题。该全局优化问题尽管困难，但还是涌现了一些优秀的算法。1953 年，矩阵分解算法，用来求解已知完全距离的蛋白质结构问题的<sup>[31]</sup>，也是后续多个算法的基础<sup>[32-40]</sup>：嵌入算法处理的是一般给定距离上下界的问题；全局光滑算法用来求解一些中小规模的问题；几何增长算法可以进行快速的求解，并且能够求解出带小误差的大规模问题；半定规划松弛算法能够容忍比较大的数据误差。

该领域未来的总体发展趋势应该是朝大规模算法和与生物科学的交叉发展。一方面，该问题的研究用到了优化领域一些最近的进展，如半定规划求解，反过来也推动了这些算法的研究。另一方面，这是一个有着非常强的应用背景的实际问题，现有的算法目前只能处理比较理想的情况，还不能完全解决真正复杂的实际问题。从而，研究这个问题具有重要的意义。该方向未来的一些研究重点与热点问题包括：

欧几里得距离矩阵的性质。我们知道，如果能得到全部的距离，就能用矩阵分解算法准确地求得解。嵌入算法正是基于这一基本思想，但它只是利用到了三角不等式等最基本的性质，并没有从整体上去考虑。如果能了解更多的欧几里得距离矩阵的性质，从而有效地从整体上进行矩阵补全，将会诱导出一个有效的算法。

有效地控制计算舍入误差。几何增长算法的求解速度非常之快，本质上是一个  $O(n)$  的算法，这是非常诱人的。但计算机舍入误差对该算法影响很大，如何根据这个问题本身的结构，找到一些消减舍入误差的方法，提高该算法的计算精度，是一个非常值得研究的方向。

大规模半定规划问题的求解。半定规划相对于几何增长算法的优势在于能够从整体上去控制误差，也能够容忍更大的数据误差，是一个非常有前途的方向。如何快速有效地求解大规模问题变得很关键，当然，这也是半定规划领域大家都关心的基本问题。

分布式算法。这是另一个处理大规模问题的方向，现在已经有了一些这方面的尝试，但还不是很有效。针对该问题设计有效的分布式算法，跟并行计算进行结合，值得开拓研究。

给定上下界的问题。已有的一些算法只能处理给定距离数值的问题，但实际的问题常常是只能给出一个上下界的估计，如何将已有的算法有效地扩展到能处理这些实际问题，也是一个很有意义的方向。

运筹与生物学真正的交叉。现有的研究都是基于生物测量的数据，然后用纯数学的方式进行求解。如何把这些数学方法跟生物技术确定结构的方法有效地结合非常值得研究。只有这样，才算真正意义上的交叉。论文<sup>[42]</sup>提到可以把这些方法求出的解作为能量极小方法的初值，这些想法都是值得参考的。

### （三）运筹学与分子模拟

分子模拟是在分子原子水平上对生物体系进行模拟和计算，研究生物分子的结构、动力学、及功能之间的关系，并应用于生物、医学、健康、工程等领域<sup>[41-43]</sup>。运筹学的优化技术在分子模拟中主要用于能量优化、分子对接、蛋白质结构预测、参数优化等，涉及连续优化、非线性优化、约束优化、随机优化等，许多算法如能量局部极小化方法、遗传算法、神经网络、模拟退火等已实现于众多应用软件中，如 AutoDock, Amber, Charmm 等软件包。国内生物分子模拟领域的研究与应用整体上主要集中在计算生物学工具的应用上，在算法及应用软件的自主开发方面的研究力量还相当薄弱。其中应用优化方法的研究工作零星地出现在蛋白质结构预测<sup>[34]</sup>、分子力场参数化<sup>[44]</sup>、分子结构域的划分等方面<sup>[45]</sup>。该方向未来的一些研究重点与热点问题包括：

（1）时空尺度和力场一直是分子模拟要解决的两大任务。主要的优化方法和应用也与这两个任务相关。由于分子模拟是在复杂的高维相空间中采样，处理采样的比较有效的优化算法常常与随机性及统计方法相结合<sup>[46]</sup>。

（2）前面提到的分子模拟中的传统优化问题仍然是未来关心的课题。其中，能量优化需将局部优化与全局优化技术结合，改进优化的结果，同时提高对大体系的计算效率。分子对接从“刚性”对接向“柔性”对接技术发展，即允许分子内的更多的原子坐标可以变化，因而有更多的自由度。蛋白质折叠预测也将处理越来越大的大分子，实际上也是增大处理自由度的空间。力场参数优化方面将主要发展极化力场、粗粒化模型力场及多尺度结合的力场。极化力场考虑了电场作用的极化效应<sup>[44]</sup>，它随动力学过程变化的而变化，因而要求做大量高效的实时优化。而不同层次和多尺度的粗粒化模型力场则是为适应越来越多的大尺度长时间的模拟需要。

（3）分子模拟在理论方法以及各个应用层面的技术细节中都可以挖掘出不同的优化问题。这里列举一些在其它方面新出现的问题。（3.1）优化算法将面临处理更高维空间、大系统、大数据量的任务，因而算法与软件必须考虑并行与分布式计算的特点和效率。（3.2）除基于能量函数的方法外，发展一些基于知识和数据的（如实验测量信息）预测分子性质的优化方法。

比如基于 X-ray 和 NMR 等实验数据得到的分子内一些原子对之间的距离,通过距离几何优化方法预测蛋白质的空间结构<sup>[35]</sup>,以及根据同源蛋白片段在结构上的类似性进行的结构预测等。

(3.3) 优化与分子动力学结合,寻求改进或克服分子模拟中遇到的一些困难,尤其是自由能的计算和采样效率。自由能计算是生物分子动力学模拟的一个核心任务,当前的主要困难和瓶颈是采样或模拟时间不足导致的计算准确度较差。采样不足是因为体系的运动较长时间地处于某些局部能量极小的状态,而对另一些对系统性质也有重要影响的状态没有或很少达到。优化方法的优点是可以避开实际动力学过程而直接针对目标函数搜索状态空间。可利用这一特点与当前分子模拟中的一些自由能计算方法,如反应路径采样、伞形采样和其它加导向力采样的方法等相结合来加快采样效率,并在计算模拟中可以设计处理更复杂更合理的约束条件。对于简单相互作用粒子体系的自由能计算,可尝试转化为应用优化算法(结合部分分子模拟)求解一个最优状态几率分布的问题,而避免进行长时间的分子动力学模拟。

(4) 结合优化方法求解一些新的模型,如隐式溶剂模拟中的变分法模型<sup>[47]</sup>及偏微分方程模型<sup>[48]</sup>。这些模型通常是多物理多尺度耦合模型,不容易直接求解。例如,结合优化理论,应用约束条件下的非线性优化理论与技巧,设计新的算法,克服求解困难<sup>[49]</sup>。快速求解的隐式模型可望结合到传统的分子模拟中去,以处理时间空间上较大尺度的对象和过程。

(5) 一些与分子有关的几何优化的问题。如分子表面的定义可通过极小化一个能量曲面来实现<sup>[47]</sup>。另外,近年发展的对分子的数学计算和几何模拟需要用到的分子表面与立体网格。网格的预处理也经常用优化技术来光滑、删减、或重排网格,而同时保持其拓扑性质与其他主要特征不变。如把立体及表面网格的优化转化成求解一个无约束二次优化的问题<sup>[50]</sup>。

## (四) 运筹学与生物分子网络推断

生物体在系统层面的复杂性更多地体现在基因的调控,即基因之间的相互影响关系及其动态的变化<sup>[51-56]</sup>,进一步导致宏观表型层面的演化主要观点是研究细胞通过 DNA, RNA, 蛋白质和化学小分子之间相互作用形成的网络,即广义意义下的基因调控网络,来实现高度复杂精确的调控<sup>[57-65]</sup>。基因调控网络的研究和运筹学有着密切的关系。运筹学的大量分支学科都可以在基因调控网络的研究中得到应用。基因调控网络的表示需要用到运筹学中的图论,即基因与基因之间的相互作用构成基因调控的有向图。然而基因之间的相互关系很难通过生物实验得到。通过基因芯片技术产生的高通量数据,衍生出的运筹学问题是如何构建一个基因调控网络的数学模型,使得模型和观测到的数据吻合地最好,广义地讲,这是一个最优化的问题,这种研究思路在计算机、工程和统计学中取得了长足的进展,它们分别被称为机器学习、系统辨识和统计学习<sup>[66]</sup>。目前主流的数学模型,包括共表达的统计和信息论模型、布尔网络模型、贝叶斯网络模型和微分方程模型<sup>[67-87]</sup>,使得利用数学建模方法来预测复杂的基因调控系统成为可能。

推断基因调控网络的主要的困难在于模型复杂性和时间序列的数据严重不足之间的鸿沟,也就是通常讲的“维度灾难”。“维度灾难”问题的克服,为运筹学带来了新的机遇和挑战。其中稀疏优化,作为最近发展起来的处理高维数据的模型,得到了广泛的应用。因为对于基因调控网络而言,某个基因和大多数基因之间都会有相互影响关系,这明显有悖于生物学常识,细胞中真实的基因调控网络是稀疏的<sup>[88]</sup>。该网络拓扑结构上的特点应该反映到网络推断的建模中<sup>[89]</sup>。这是基因调控网络研究中的热点,对优化模型和算法都提出了很多新的问题。

从实际应用角度出发,“维度”灾难问题的解决策略之一是数据集成。即基因芯片数据,各种低、高通量的实验手段积累了异源的海量数据<sup>[90]</sup>。这些实验数据包括序列<sup>[91]</sup>,三维结构<sup>[92]</sup>,功能<sup>[93]</sup>,定位<sup>[94]</sup>,丰度<sup>[95,96]</sup>,调控<sup>[97,98]</sup>,表现型<sup>[99]</sup>,蛋白质物理和遗传相互作用<sup>[100]</sup>,进化中保

守的相互作用、发育谱、基因相邻关系、共进化、基因簇以及结构域融合数据<sup>[101,102]</sup>等。因此在系统生物学的框架下最大程度、最优地集成这些不同来源、提供不同角度信息的功能基因组数据，进而重构出更精确的基因调控网络也是一个大的挑战<sup>[103]</sup>。这里的最优集成，同时包含了对数据来源、数据精度等多因素的考量，提出一个具体的最优化目标，在满足一定的约束条件下得到最优化模型，设计优化算法，得到生物上满意的最优解<sup>[104,105]</sup>。

该方向未来的一些研究重点与热点问题包括：

(1) 目前值得关注的基因调控网络中与运筹学有关的重要研究内容包括：网络的重建的数学建模，即对生物系统扰动，测量分子层面的数据，进而反向来建立生物系统的定量数学模型。其中涵盖了对高噪声的生物数据进行处理和分析，同时整合不同层面的、异源的生物数据，利用运筹学方法从相互作用、功能模块、信号通路、全局网络等不同层面推断出分子生物网络。不难看到，基于模型和算法的研究方法是生命科学常规实验研究手段的必要补充。前期研究<sup>[106]</sup>已经看到，最优化模型的凸性和正则化项的引入对于保证高效的计算和问题的简化有着非常重要的作用。所以通过深入研究最优化理论和算法设计，还有很大的空间可以改进目前存在的方法并且提出新的方案。

(2) 基因调控网络的重要性体现在它是研究的一个新的平台，利用它作为生物系统的抽象表示对其鲁棒性、时空特异性、多样性、适应性等复杂的系统行为进行分析研究，考虑网络结构特性与生物系统进化、衰老、疾病、外界条件交互的关系，并对网络结构特性以及动力学行为进行研究<sup>[107,108]</sup>。此外对基因调控网络的控制也是未来研究重点。例如针对生物分子系统的结构特点和噪声的基本特性，从控制的角度出发，构造相应的参数辨识、状态估计和反馈控制等控制策略。进一步，在合成生物学的框架内，设计与构建具有特定功能的生物分子元件和网络模块，例如基因开关、基因振荡子、基因计数器和生物传感器等，为实验设计提供理论基础与指导作用，并可以进一步应用于工程与医疗中。这些方面都有大量的运筹学问题。过去几十年，运筹学与控制论的基本思想与方法逐步渗透到基因调控网络的研究中。同时，来源于生命科学的运筹与控制问题也为我们提出了新的机遇与挑战。基因调控网络的控制对生命科学中困扰人类的基本问题，如延长寿命，治愈癌症，糖尿病等顽疾有着非常重要的现实意义。此外，基因调控网络控制研究对合成生物学，网络医学，个性化医学等相关学科的发展具有潜在的应用价值。

(3) 基因调控网络来源于生物实际问题，它的研究还要回到具体实际问题中。我们上面讨论的针对一般方法论的研究，需要和具体的应用实际结合，才能更加有的放矢，产生有意义的结果，这是基因调控网络研究一个非常重要的动向，就是将基因调控网络研究与重大的生命科学问题结合，例如发育与分化、复杂疾病、衰老、干细胞等问题<sup>[109-111]</sup>，这些都需要研究新的模型和算法。基因调控网络的研究需要跟上生命科学和相关交叉学科的发展步伐，关注新的技术和数据，要重视实际应用。通过和实验生物学家实质性的交流合作，选择重要的问题和适当的切入点，实质性地加入生命科学的研究，促进运筹学与基因调控网络的交叉研究。

## (五) 运筹学与复杂疾病

复杂疾病的网络生物学研究<sup>[112-145]</sup>正蓬勃发展。复杂疾病网络生物学的研究主要是指利用网络的模型和方法，针对影响人们身体健康的重大疾病如癌症、糖尿病、老年痴呆症等进行致病机理分析，将疾病的发生发展看作相关分子网络的动态演化过程。网络的建立、特征分析、小的子网络分解、网络标记物的筛选与组合、关键蛋白质的挑选、药物靶点的设计等均可运用运筹学的多种模型与方法来解决。



大规模的基因组学研究已经产生了大量的生物数据, 这些数据必将加深我们对许多常见疾病的分子机制的理解. 对于从这些数据中挖掘信息, 可视化数据并结合数据建立复杂网络, 利用图论和复杂网络的方法寻找疾病相关的基因的需求已经激发了计算方法在转化生物学中的发展, 论文<sup>[137]</sup>通过结合基因以及基因之间的功能关系信息构造了一个人类基因功能网络, 并利用这个网络来排序候选基因. 论文<sup>[138]</sup>改进了传统的直接邻居和最短路方法, 在蛋白作用网络上利用重启的随机游走和扩散核这两种全局的网络距离测度来寻找潜在的致病基因. 论文<sup>[139]</sup>提出了名为 CIPHER 的方法, 该方法结合了人类的蛋白质相互作用网络信息、疾病之间的性状类似性网络信息和已知的疾病基因关联网络去揭示疾病和基因之间的潜在关联, 另外两篇论文<sup>[143,144]</sup>提出基于保守共表达网络和模块化本质预测人类致病基因. 论文<sup>[142]</sup>提出了模块化的候选致病基因排序算法, 可以识别潜在的和疾病相关的基因. 论文<sup>[143-145]</sup>提出了基于网络的方法预测疾病相关的 miRNA、miRNA-环境因子作用和 lncRNA, 尤其后两个工作为领域内第一个解决该问题的计算方法。

基于运筹学的理论与方法在复杂疾病研究中的具体研究方向主要包括, 重大疾病的网络分子标记物的筛选, 基于网络方法的药物靶点的设计以及癌症等重大疾病的早期诊断技术的开发等方面. 未来的研究方向和研究重点主要包括: 标记疾病进程的网络分子标记物的筛选; 疾病发生发展的特异局部网络特征的探求; 研究重要疾病蛋白的相互作用和可能药物靶点; 建立疾病病程变化过程中局部网络特征库; 分子网络标记物的最优组合设计及其对疾病自然病程的分类; 研究正常与疾病之间的差异与过渡特征; 设计开发基于网络生物学的癌症早期新型诊断工具。

## 参考文献

- 【1】. Pevzner, P. A., H. Tang, M. S. Waterman. An Eulerian path approach to DNA fragment assembly. PNAS, 98(17): 9748-9753, 2001.
- 【2】. Pevzner, P. A., H. Tang. Fragment Assembly with Double-Barreled Data. Bioinformatics, 17(suppl 1): S225-S233, 2001.
- 【3】. Blazewicz, J., P. Formanowicz, M. Kasprzak, W.T. Markiewicz, A. Swiercz. Tabu search algorithm for DNA sequencing by hybridization with isothermic libraries. Computational Biology and Chemistry, 28(1): 11-19, 2004.
- 【4】. Blazewicz, J., C. Oguz, A. Swiercz, J. Weglarz. DNA sequencing by hybridization via genetic search. Operations Research, 54(6): 1185-1192, 2006.
- 【5】. Blazewicz, J., F. Glover, M. Kasprzak, W.T. Markiewicz, C. Oguz, D. Rebholz-Schuhmann, A. Swiercz. Dealing with repetitions in sequencing by hybridization. Computational Biology and Chemistry, 30(5): 313-20, 2006.
- 【6】. Blazewicz, J., M. Kasprzak. Computational complexity of isothermic DNA sequencing by hybridization. Discrete Applied Mathematics, 154(5): 718-729, 2006.
- 【7】. Blazewicz, J., M. Bryja, M. Figlerowicz, P. Gawron, M. Kasprzak, E. Kirton, D. Platt, J. Przybytek, A. Swiercz, L. Szajkowski. Whole genome assembly from 454 sequencing output via modified DNA graph concept. Computational Biology and Chemistry, 33(3): 224-230, 2009.
- 【8】. Blazewicz, J., E.K. Burke, M. Kasprzak, A. Kovalev, M.Y. Kovalyov. The simplified partial digest problem: Approximation and a graph-theoretic model. European Journal of Operational Research, 208(2): 142-152, 2011.
- 【9】. Zhang, J.-H., L.-Y. Wu, and X.-S. Zhang. Reconstruction of DNA sequencing by hybridization. Bioinformatics, 19(1): 14-21, 2003.
- 【10】. Zhang, J.-H., L.-Y. Wu, Y.-Y. Zhao, X.-S. Zhang. An optimization approach to the reconstruction of positional DNA sequencing by hybridization with errors. European Journal of Operational Research, 182,

413-427, 2007.

- 【11】 . Zerbino, D. R., E. Birney. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18 (5): 821-829, 2008.
- 【12】 . Smith, T. F., and M. S. Waterman. Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, 147: 195-197, 1981.
- 【13】 . Needleman, S. B., and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48 (3): 443-53, 1970.
- 【14】 . Sankoff, D., J. B. Kruskal. (eds.), *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison*, Addison-Wesley, Reading MA, 1983
- 【15】 . Levitt, M., and M. Gerstein. A unified statistical framework for sequence comparison and structure comparison. *PNAS*, 95(11): 5913-5920, 1998.
- 【16】 . Mathews, D. H., M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *PNAS*, 101(19): 7287-7292, 2004.
- 【17】 . Waterman, M. S., and T. F. Smith. Rapid dynamicprogramming algorithms for RNA secondary structure. *Advances in Applied Mathematics*, 7(4): 455-464, 1986.
- 【18】 . Lee, C., C. Grasso, and M. F. Sharlow. Multiple sequence alignment using partial order graphs. *Bioinformatics*, 18(3): 452-464, 2002.
- 【19】 . Zhang, X.-S., R.-S. Wang, L.-Y. Wu, L. Chen. Models and algorithms for haplotyping problem. *Current Bioinformatics*, 1(1), 105-114, 2006.
- 【20】 . Wu, L.-Y. Haplotype Inference Models and Algorithms. Chapter 36 in *Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications*, Mourad Elloumi, Albert Y. Zomaya, John Wiley & Sons, Inc., 843-864, USA, 2011.
- 【21】 . Zhao, Y. Y., L.-Y. Wu, J.-H. Zhang, R.-S. Wang and X.-S. Zhang. Haplotype assembly from aligned weighted SNP fragments. *Computational Biology and Chemistry*, 29(4), 281-287, 2005.
- 【22】 . Wang, R.-S., L.-Y. Wu, Z.-P. Li, X.-S. Zhang. Haplotype reconstruction from SNP fragments by minimum error correction. *Bioinformatics*, 21(10), 2456-2462, 2005.
- 【23】 . Zhang, J.-H., L.-Y. Wu, J. Chen, X.-S. Zhang. A fast haplotype inference method for large population genotype data. *Computational Statistics and Data Analysis*, 52(11): 4891-4902, 2008.
- 【24】 . Geraci, F. A comparison of several algorithms for the single individual SNP haplotyping reconstruction problem. *Bioinformatics*, 26(18): 2217-2225, 2010.
- 【25】 . He, D., A. Choi, K. Pipatsrisawat, A. Darwiche, and E. Eskin. Optimal algorithms for haplotype assembly from whole-genome sequence data. *Bioinformatics*, 26(12): i183-i190, 2010.
- 【26】 . Wang, J., M. Xie, and J. Chen. A Practical Exact Algorithm for the Individual Haplotyping Problem *MEC/GI. Algorithmica*, 56(3), 283-296, 2010.
- 【27】 . Xie, M., J. Wang, and J. Chen. A model of higher accuracy for the individual haplotyping problem based on weighted SNP fragments and genotype with errors. *Bioinformatics*, 24(13): i105-i113, 2008.
- 【28】 . Genovese, L. M., F. Geraci, M. Pellegrini. SpeedHap: An Accurate Heuristic for the Single Individual SNP Haplotyping Problem with Many Gaps, High Reading Error Rate and Low Coverage. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(4): 492-502, 2008.
- 【29】 . Li, Z.-P., W.-F. Zhou, X.-S. Zhang, L. Chen. A parsimonious tree-grow method for haplotype inference. *Bioinformatics*, Vol. 21, No. 17, pp. 3475-3481, 2005.
- 【30】 . Wu, L.-Y., Z. Li, R.-S. Wang, X.-S. Zhang, L. Chen. Self-organizing map approaches for the haplotype assembly problem. *Mathematics and Computers in Simulation*, 79(10), 3026-3037, 2009
- 【31】 . Blumenthal, L.M. (1953) *Theory and applications of distance geometry*. Clarendon, Oxford.
- 【32】 . [http://en.wikipedia.org/wiki/Protein\\_structure](http://en.wikipedia.org/wiki/Protein_structure)

- 【33】. Crippen, G. M. and Havel, T. F. (1988) Distance geometry and molecular conformation, John Wiley & Sons.
- 【34】. Saxe, J.B. Embeddability of weighted graphs in k-space is strongly NP-hard. In: Proc. 17 th Allerton Conference in Communications, Control and Computing, pp.480-489,1979.
- 【35】. Sit, A. Wu, Z. and Yuan, Y. A geometric buildup algorithm for the solution of the distance geometry problem using least-squares approximation, *Bulletin of Mathematical Biology* 71:1914-1933, 2009.
- 【36】. Blumenthal, L.M. Theory and applications of distance geometry. Clarendon, Oxford. 1953.
- 【37】. More, J. and Wu, Z. Global continuation for distance geometry problem, *SIAM journal on Optimization*, Vol. 7, No. 3: 814-836, 1997.
- 【38】. Dong, Q. and Wu, Z. A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances, *J. Global Optim.* 22: 365-375,2002
- 【39】. Biswas, P., Liang, T., Wang, T. and Ye, Y., Semidefinite programming based algorithm for sensor network localization. *ACM J. Trans. Sensor Netw.* 2,188-220,2006.
- 【40】. Yoon, J., Gad Y. and Wu. Z.,Mathematical modeling of protein structure using distance geometry. Technical report TR00-24, Rice University,2000.
- 【41】. Schlick T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Springer-Verlag, Berlin-New York, 2002.
- 【42】. Leach A.R., 分子模拟的原理和应用 (第2版) (*Molecular Modelling: Principles and Applications*), 世界图书出版公司北京公司, 2003
- 【43】. Brooks C., Karplus M. and Pettitt B., *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Wiley, 1988.
- 【44】. Wang Y.T., Izvekov S., Yan T.Y. and Voth G.A.,Multiscale coarse-graining of ionic liquids. *J Phys Chem B.* 110 (8): 3564-75., 2006.
- 【45】. Zhang Z.Y., Lu L.Y., Noid W.G., Krishna V., Pfandtner J., Voth G.A.. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophysical Journal*, (95): 5073-5083, 2008
- 【46】. Swanson J.M.J., Adcock S.A. and McCammon J.A. Optimized radii for Poisson-Boltzmann calculations with the AMBER force field, *J. Chem. Theory Comput.* 3: 484-493, 2005.
- 【47】. Dzubiella J., Swanson J.M.J. and McCammon J.A.,Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models, *Phys. Rev. Lett.* 96: 087802,2006.
- 【48】. Lu B.Z. and Zhou Y.C., Poisson-Nernst-Planck Equations for Simulating Biomolecular Diffusion-Reaction Processes II: Size Effects on Ionic Distributions and Diffusion-reaction Rates, *Biophysical J.* 100: 2475-2485, 2011.
- 【49】. Zhou S.G., Wang Z.M. and Li B., Mean-field description of ionic size effects with non-uniform ionic sizes: A numerical approach, *Phys. Rev. E*, 84: 021901,2011
- 【50】. Gao Z., Yu Z. and Wang J., An Optimization-Based Iterative Approach to Tetrahedral Mesh Smoothing, invited book chapter in *Image-based Geometric Modeling and Mesh Generation* (Edited by Y. Zhang), pages 143-157, Springer,2012.
- 【51】. Venter, J. C., M. D. Adams, et al., The Sequence of the Human Genome, *Science* 291(5507): 1304-1351,2001.
- 【52】. Huang, R., A. Wallqvist, et al., Targeting changes in cancer: assessing pathway stability by comparing pathway gene expression coherence levels in tumor and normal tissues, *Molecular Cancer Therapeutics* 5(9): 2417,2006
- 【53】. Huelsenbeck, J. P. and F. Ronquist, *Bioinformatics*, *Bioinformatics*: 754-755,2001.
- 【54】. Ideker, T., O. Ozier, et al., Discovering regulatory and signalling circuits in molecular interaction networks, *Bioinformatics* 18(suppl\_1): S233-240,2002.
- 【55】. Kitano, H. (2002). Computational systems biology, *Nature* 420(6912): 206-210.

- 【56】 . Liu, Z-P, Wang, Y., Zhang X-S, Chen L. Network-based analysis of complex diseases, *IET Systems Biology*, 6(1): 122-133, 2012.
- 【57】 . Richards, F. M., "The protein folding problem." *Scientific American* 264(1): 54-63, 1991.
- 【58】 . Wang, Y., Zhang, X-S, and Chen, L. Modelling biological systems from molecules to dynamical networks, *BMC Systems Biology*, 6(Suppl 1):S1,2012.
- 【59】 . Wang, Y., Zhang X-S, and Chen, L. Computational systems biology: integration of sequence, structure, network, and dynamics, *BMC Systems Biology*, 5(Suppl 1):S1, 2011.
- 【60】 . Wang, Y., Zhang, X-S, and Chen, L. Optimization meets systems biology. *BMC Systems Biology*, Vol. 4, Suppl 2, S1, 2010.
- 【61】 . 陈洛南, 王勇, 费敏锐, 刘曾荣, 从理工科视角探索系统生物学, *科技导报*, Vol. 25, No. 10: 5--9, 2007.
- 【62】 . Barabasi, A. L. and Z. N. Oltvai, *Network biology: understanding the cell's functional organization*, *Nature Reviews Genetics* 5(2): 101-113,2004.
- 【63】 . Bredel, M. and E. Jacoby, *Chemogenomics: an emerging strategy for rapid target and drug discovery*, *Nature Reviews Genetics* 5(4): 262-275,2004.
- 【64】 . Carthew, R. W. , *Gene regulation by microRNAs*, *Current Opinion in Genetics & Development* 16(2): 203-208,2006.
- 【65】 . Chuang, H. Y., E. Lee, et al., *Network-based classification of breast cancer metastasis*, *Mol Syst Biol* 3: 140,2007.
- 【66】 . Csermely, P., V. Agoston, et al., *The efficiency of multi-target drugs: the network approach might help drug design*, *Trends in Pharmacological Sciences* 26(4): 178-182,2005.
- 【67】 . Gardner, T. S. and J. J. Faith, *Reverse-engineering transcription control networks*, *Physics of Life Reviews* 2(1): 65-88,2005.
- 【68】 . Yamada, T., M. Kanehisa, et al., *Extraction of phylogenetic network modules from the metabolic network*, *BMC Bioinformatics* 7(1): 130.,2006.
- 【69】 . Akutsu, T., S. Miyano, et al., *Identification of genetic networks from a small number of gene expression patterns under the Boolean network model*, *Pac Symp Biocomput* 17: 28, 1999.2007.
- 【70】 . Auffray, C., *Protein subnetwork markers improve prediction of cancer outcome*, *Molecular Systems Biology* 3: 141,2007.
- 【71】 . Bansal, M., V. Belcastro, et al., *How to infer gene networks from expression profiles.* *Mol Syst Biol* 3: 78,
- 【72】 . Bansal, M., G. D. Gatta, et al., *Inference of gene regulatory networks and compound mode of action from time course gene expression profiles*, *Bioinformatics* 22(7): 815-822,2006.
- 【73】 . Chen, T., H. L. He, et al., *Modeling gene expression with differential equations*, *Pacific Symposium on Biocomputing* 4: 29-40,1999.
- 【74】 . D'Haeseleer, P., S. Liang, et al., *Genetic network inference: from co-expression clustering to reverse engineering*, *Bioinformatics* 16(8): 707-726, 2000.
- 【75】 . de Jong, H. *Modeling and Simulation of Genetic Regulatory Systems: A Literature Review*, *Journal of Computational Biology* 9(1): 67-103,2002.
- 【76】 . Demeter, J., C. Beauheim, et al., *The Stanford Microarray Database: implementation of new analysis tools and open source release of software*, *Nucleic Acids Research* 35(Database issue): D766,2007.
- 【77】 . Friedman, N., *Inferring cellular networks using probabilistic graphical models.* *Science* 303(5659): 799-805,2004.
- 【78】 . Gardner, T. S., D. di Bernardo, et al., *Inferring genetic networks and identifying compound mode of action via expression profiling*, *Science* 301(5629): 102-5, 2003.
- 【79】 . Guimera, R., M. Sales-Pardo, et al., *A network-based method for target selection in metabolic networks*,

Bioinformatics 23(13): 1616,2007.

- 【80】 . Guo, Z., Y. Li, et al. , Edge-based scoring and searching method for identifying condition-responsive protein protein interaction sub-network,Bioinformatics 23(16): 2121-2128,2007.
- 【81】 . Hartemink, A. J.,Reverse engineering gene regulatory networks." Nature Biotechnology 23: 554-555,2005.
- 【82】 . Husmeier, D., Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks." Bioinformatics 19(17): 2271-2282, 2003.
- 【83】 . Margolin, A. A., I. Nemenman, et al., RACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, BMC Bioinformatics 7(Suppl 1):S7,2006.
- 【84】 . Markowitz, F. and R. Spang ,Inferring cellular networks: review." BMC Bioinformatics 8(Suppl 6): S5,2007.
- 【85】 . Schlitt, T. and A. Brazma,Current approaches to gene regulatory network modelling." BMC Bioinformatics 8(Suppl 6): S9,2007.
- 【86】 . Scott, M. S., T. Perkins, et al.,Identifying Regulatory Subnetworks for a Set of Genes." Mol Cell Proteomics 4(5): 683-692,2005.
- 【87】 . Steven Zheng, X. F. and T. F. Chan,Chemical genomics in the global study of protein functions, Drug Discovery Today 7(3): 197-205,2002.
- 【88】 . Wang, R. S., Y. Wang, et al. Inferring transcriptional regulatory networks from high-throughput data, Bioinformatics 23: 3056-3064,2007.
- 【89】 . Yeung, M. K., J. Tegner, et al.,Reverse engineering gene networks using singular value decomposition and robust regression." Proceedings of the National Academy of Sciences USA 99(9): 6163,2002.
- 【90】 . Wang, Y., T. Joshi, et al. , Inferring gene regulatory networks from multiple microarray datasets." Bioinformatics 22(19): 2413,2006.
- 【91】 . Eils, J., C. Lawerenz, et al., Databases for systems biology." Computational Systems Biology,2006.
- 【92】 . Bairoch, A., R. Apweiler, et al. The SWISS-PROT protein sequence data bank and its supplement TrEMBL, Nucleic Acids Research 25(1): 31-36,1997.
- 【93】 . Sussman, J. L., D. Lin, et al.,Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules, Acta Crystallogr D Biol Crystallogr 54(1 Part 6): 1078-1084,1994.
- 【94】 . Camon, E., M. Magrane, et al. ,The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology." Nucleic Acids Research 32(90001): W313-W317,2004.
- 【95】 . Huh, W. K., J. V. Falvo, et al.,Global analysis of protein localization in budding yeast." Nature 425(6959): 686-691,2003.
- 【96】 . Lu, L. J., Y. Xia, et al. , Assessing the limits of genomic data integration for predicting protein networks." Genome Research 15(7): 945, 2005.
- 【97】 . Xia, Y., L. J. Lu, et al., Integrated Prediction of the Helical Membrane Protein Interactome in Yeast, Journal of Molecular Biology 357(1): 339-349, 2006.
- 【98】 . Wingender, E., X. Chen, et al.,TRANSFAC: an integrated system for gene expression regulation, Nucleic Acids Research 28(1): 316-319,2000.
- 【99】 . Ge, H., Z. Liu, et al. , Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae." Nat Genet 29(4): 482-486, 2001.
- 【100】 . Kahraman, A., A. Avramov, et al., PhenomicDB: a multi-species genotype/phenotype database for comparative phenomics." Bioinformatics 21(3): 418-420, 2005.
- 【101】 . Stark, C., B. J. Breitkreutz, et al.,BioGRID: a general repository for interaction datasets, Nucleic Acids Research,2004.
- 【102】 . Yu, H., N. M. Luscombe, et al., Genomic analysis of gene expression relationships in transcriptional regulatory networks, Trends in Genetics 19(8): 422-427,2003.

- 【103】 . Bowers, P. M., M. Pellegrini, et al., Prolinks: a database of protein functional linkages derived from coevolution, *Genome Biology* 5:R35,2004.
- 【104】 . Wang, Y., Wang, R-S Wang, Joshi, T., Xu, D., Zhang, X-S, Chen, L., and Xia, Y. A Linear Programming Framework for Inferring Gene Regulatory Networks by Integrating Heterogeneous Data, *Computational Methodologies in Gene Regulatory Networks*, Sanjoy Das, Doina Caragea, W. H. Hsu, Stephen M. Welch (eds.), IGI Global, 2009.
- 【105】 . Rui-Sheng Wang, Yong Wang, Xiang-Sun Zhang, Luonan Chen. Inferring transcriptional regulatory networks from high-throughput data. *Bioinformatics*, Vol. 23, pp. 3056-3064, 2007.
- 【106】 . Wang, Y., X.-S. Zhang, L. Chen. A network biology study on circadian rhythm by integrating various omics data. *OMICS: A Journal of Integrative Biology*, Vol. 13, No. 4, 2009.
- 【107】 . P. Wang, J. Lu, M. J. Ogorzalek. Global relative parameter sensitivities of the feed-forward loops in genetic networks. *Neurocomputing*, 78(1): 155-165, 2012.
- 【108】 . 王沛, 吕金虎. 基因调控网络的控制: 机遇与挑战. *自动化学报*, 2013, in press.
- 【109】 . Zhang, T., Statistical behavior and consistency of classification methods based on convex risk minimization, *Ann. Statist* 32(1): 56-85,2004.
- 【110】 . Xia, K., H. Xue, et al., Identification of the Proliferation/Differentiation Switch in the Cellular Network of Multicellular Organisms, *PLoS Comput Biol* 2(11): e145,2006.
- 【111】 . Xue, H., B. Xian, et al., A modular network model of aging. *Molecular Systems Biology* 3:147: 1-11, 2007.
- 【112】 . Wang, Y.. *Bioinformatics studies on induced pluripotent stem cell*, *Current Bioinformatics*, Vol. 8, 2013.
- 【113】 . Alaoui-Ismaili MH, Lomedico PT, and Jindal S, Chemical genomics: discovery of disease genes and drugs. *Drug Discovery Today*, 7(5):292-294 , 2002.
- 【114】 . Alon U, *An Introduction to Systems Biology: Design Principles of Biological Circuits* Chapman & Hall/CRC, 2007.
- 【115】 . Barabasi AL, Albert R, Emergence of scaling in random networks. *Science*, 286(5439):509-512, 1999.
- 【116】 . Barabasi AL, Network Medicine -- From Obesity to the "Diseasome". *N Engl J Med*, 357(4): 404-407, 2007.
- 【117】 . Barabasi AL and Oltvai ZN, Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101-113, 2004.
- 【118】 . Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol*. 2007;3:140.
- 【119】 . Chung CC, Magalhaes W, Gonzalez-Bosquet J, Chanock SJ, Genome-wide Association Studies in Cancer - Current and Future Directions. *Carcinogenesis*. in press ,2009.
- 【120】 . Cui Q, et al., Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol*, 2:46, 2006.
- 【121】 . Duarte NC, et al., Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA*. 104(6): 1777, 2007.
- 【122】 . Elowitz MB and Leibler S, A synthetic oscillatory network of transcriptional regulators. *Nature*, 2000, 403(6767):335-338.
- 【123】 . Ergun A, et al., A network biology approach to prostate cancer. *Mol Syst Biol*, 3: 82, 2007.
- 【124】 . Fung E, Wong WW, Suen JK, Bulter T, Lee S, and Liao JC, A synthetic gene-metabolic oscillator. *Nature*, 435:118-122, 2005.
- 【125】 . Gardner TS, Cantor CR, and Collins JJ, Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767):339-42, 2000.
- 【126】 . Goh KI, et al., The human disease network. *Proc Natl Acad Sci USA*, 104(21): 8685-8690, 2007.
- 【127】 . Kann MG, Protein interactions and disease: computational approaches to uncover the etiology of diseases.

Brief Bioinform, 8(5):333-346, 2007.

- 【128】. Kitano H, Computational systems biology. Nature, 420(6912):206-210, 2002.
- 【129】. Kreeger PK, Lauffenburger DA, Cancer Systems Biology: A Network Modeling Perspective. Carcinogenesis, 2009.
- 【130】. Li F, Long T, Lu Y, Ouyang Q and Tang C, The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci USA, 101(14): 4781-4786, 2004.
- 【131】. Ma W, Trusina A, El-Samad H, Lim W and Tang C, Defining Network Topologies that Can Achieve Biochemical Adaptation, Cell, 138(4):760,2009.
- 【132】. Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN, Means AR, Kasturi SP, Khan N, Li GM, McCausland M, Kanchan V, Kokko KE, Li S, Elbein R, Mehta AK, Aderem A, Subbarao K, Ahmed R, Pulendran B. Systems biology of vaccination for seasonal influenza in humans. Nat Immunol. Jul 10;12(8):786-795,2011.
- 【133】. Tu K, Yu H, Hua YJ, Li YY, Liu L, Xie L and Li YX, Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms. Nucleic Acids Research, 37(18):5969-5980, 2009.
- 【134】. van Rooij E and Olson EN, MicroRNAs: powerful new regulators of heart disease and provocative therapeutic targets. Journal of Clinical Investigation, 117(9): 2369, 2007.
- 【135】. Workman, CT, Mak, HC, McCuine, S, Tagne, JB, Agarwal, M, Ozier, O, Begley, TJ, Samson, LD, Ideker, T, A Systems Approach to Mapping DNA Damage Response Pathways. Science, 312 (5776):1054-1059, 2006.
- 【136】. Yang K, Bai H, Ouyang Q, Lai L, Tang C. Finding multiple target optimal intervention in disease-related molecular network. Mol Syst Biol, 4:228, 2008.
- 【137】. Franke L, van Bakel H, Fokkens L, De Jong ED, Egmont-Petersen M, Wijnenga, Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. Am J Hum Genet, 78:1011–1025, 2006.
- 【138】. Kohler S, Bauer S, Horn D, Robinson PN, Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet, 82: 949–958,2008.
- 【139】. Wu XB, Jiang R, Zhang MQ, Li S, Network-based global inference of human disease genes. Mol Syst Biol, 4: 189–199, 2008.
- 【140】. Ala U, Piro RM, Grassi E, Damasco C, Silengo L, Oti M et al, Prediction of human disease genes by human– mouse conserved co-expression analysis. PLoS Comput Biol, 3: e1000043, 2008.
- 【141】. Oti M, Reeuwijk VJ, Huynen MA, Brunner HG, Conserved co-expression for candidate disease gene prioritization. BMC Bioinformatics, 9:208–210, 2008.
- 【142】. Chen X, Yan GY, Liao XP. A novel candidate disease genes prioritization method based on module partition and rank fusion. OMICS: A Journal of Integrative Biology 14:337-356, 2010.
- 【143】. Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA-disease associations. Molecular BioSystems 8: 2792-2798, 2012.
- 【144】. Chen X, Liu MX, Cui QH, Yan GY. Prediction of disease-related interactions between microRNAs and environmental factors based on a semi-supervised classifier. PLoS ONE, 7:e43425, 2012.
- 【145】. Chen X, Yan GY. Novel human lncRNA-disease association inference based on lncRNA expression profiles. Bioinformatics, 29: 2617-2624, 2013.

总顾问：章祥荪

执笔人：闫桂英、吴凌云、王勇

编写组（依照姓氏拼音排序）：陈兴、戴彧虹、刘治平、卢本卓、吕金虎、邵伟文、盛镇醴  
索引

计算系统生物学、生物分子、基因组、测序、复杂疾病、  
网络推断、蛋白质结构、分子模拟、调控网络