(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107391961 B
(45)授权公告日 2020.11.17

(21)申请号 201710237916.2
(22)申请日 2012.09.07
(65)同一申请的已公布的文献号
申请公布号 CN 107391961 A
(43)申请公布日 2017.11.24
(30)优先权数据
61/532,972 2011.09.09 US
(62)分案原申请数据
201280043499.3 2012.09.07
(73)专利权人 菲利普莫里斯生产公司
地址 瑞士纳沙泰尔
(72)发明人 F·马丁
(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
代理人 鲍进
(51)Int.Cl.
G16B 5/00(2019.01)
审查员 刘秀

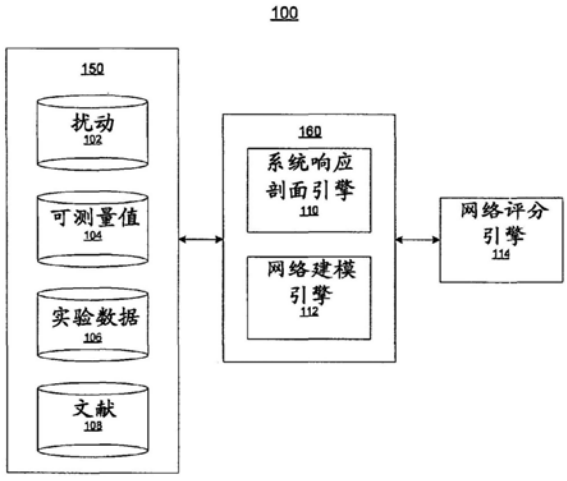
权利要求书3页 说明书29页 附图18页

(54)发明名称

用于基于网络的生物活性评估的系统与方法

(57)摘要

本发明公开涉及用于基于网络的生物活性评估的系统与方法。本文公开了基于从生物系统中的实体子集测出的活性数据来量化生物系统对一个或多个扰动的响应的系统与方法。基于该活性数据和描述测量和未测量的实体之间关系的生物系统的网络模型，推断未测量的实体的活性。推断出的活性用于导出量化生物系统对扰动的响应，诸如对治疗条件的响应的得分。该得分可以代表网络对扰动的响应的量级和拓扑分布。



CN 107391961 B

1.一种用于估计生物实体的活性值的计算机化方法，该方法包括：

识别包括代表生物系统中的生物实体的节点和代表节点之间的关系的边的网络模型，其中节点包括代表第一组生物实体的第一组节点和代表第二组生物实体的第二组节点，第一组生物实体中的至少一个生物实体与第二组生物实体中的至少一个生物实体交互，其中第二组生物实体不能够被直接测量，并且其中边将第一组节点中的至少一些节点连接到第二组节点中的至少一些节点并且表示第一组生物实体和第二组生物实体之间的关系；

在处理器接收第一组治疗数据和第二组治疗数据，其中第一组治疗数据和第二组治疗数据与第一组生物实体对应，第一组治疗数据对应于暴露给制剂的第一组生物实体，第二组治疗数据对应于不暴露给制剂的第一组生物实体；

基于第一组治疗数据与第二组治疗数据之间的差异为第一组节点获得活性测量；

基于网络模型和第一组节点的活性测量，为第二组节点生成活性值，其中第二组节点不能够被直接测量，该生成包括：

基于第二组节点中的每个特定节点的活性值与该特定节点利用网络模型中的边所连接到的节点的活性值或活性测量之间的差异来为该特定节点生成活性值，其中对于第二组节点中的每个特定节点，活性值是最小化所述差异的活性值；及

基于网络模型和第二组节点的活性值，生成代表由制剂造成的生物系统的扰动的网络模型的得分。

2.如权利要求1所述的方法，其中生成的活性值与推断的数据对应。

3.如权利要求1所述的方法，其中所述得分是至少部分地基于第二组节点中的每个节点的外出边的数量、第二组节点中的每个节点的进入边的数量、以及由连接第二组节点中的节点的边定义的邻接矩阵来生成的。

4.如权利要求3所述的方法，还包括确定指示生物系统的扰动的得分的统计显著性。

5.如权利要求4所述的方法，其中得分的统计显著性是通过比较得分与多个测试得分来确定的，这多个测试得分是从多个随机生成的测试网络模型计算出来的。

6.如权利要求5所述的方法，其中所述多个随机生成的测试网络模型包括节点标签、将第一组节点连接到第二组节点的边和连接第二组节点中的节点的边中的一个或多个的随机置换。

7.如权利要求1-6中任一项所述的方法，其中每个活性值是活性测量的线性组合，其中该线性组合包括计算代表在第一组节点和第二组节点之间连接的边的第一矩阵和代表连接第二组节点中的节点的边的第二矩阵。

8.如权利要求1-6中任一项所述的方法，还包括通过形成第一组节点的活性测量的每个活性测量的变差估计的线性组合来为每个活性值提供变差估计。

9.如权利要求1-6中任一项所述的方法，还包括：

将活性值表示为第一活性值向量；

将第一活性值向量分解成第一起作用向量和第一不起作用向量，使得第一起作用向量和第一不起作用向量之和是第一活性值向量。

10.如权利要求9所述的方法，还包括：

基于第一组生物实体的第三组治疗数据和第四组治疗数据识别与第二组节点的活性值对应的第二活性值向量；

将第二活性值向量分解成第二起作用向量和第二不起作用向量;及

比较第一起作用活性值向量和第二起作用活性值向量。

11.如权利要求10所述的方法,其中比较第一起作用向量和第二起作用向量包括计算第一起作用向量和第二起作用向量之间的相关性以指示第一组治疗数据和第三组治疗数据的可比性。

12.如权利要求1-6中任一项所述的方法,还包括:

基于第三组治疗数据与第四组治疗数据之间的差异为第二网络模型中的与第三组生物实体对应的第三组节点获得活性测量;

为第二网络模型中的第四组节点中每个特定节点生成活性值;及

比较第二组节点的活性值与第四组节点的活性值。

13.如权利要求12所述的方法,其中,包括第一组节点和第二组节点的网络模型是第一网络模型,并且所述比较包括应用基于与第一网络模型关联的带符号拉普拉斯算子和与第二网络模型关联的带符号拉普拉斯算子的内核规范相关性分析。

14.如权利要求12所述的方法,其中包括第一组节点和第二组节点的网络模型是第一网络模型,并且第一网络模型和第二网络模型对应于包括试管系统、活体系统、小鼠系统、大鼠系统、非人灵长类动物系统及人体系统的组中的两个不同元素。

15.如权利要求1-6中任一项所述的方法,其中活性测量是倍数变化值,且每个节点的倍数变化值包括由对应节点表示的生物实体的对应各组治疗数据之间的差异的对数。

16.如权利要求1-6中任一项所述的方法,其中对于第二组节点中的每个特定节点,差异至少部分地基于该特定节点的活性值与该特定节点利用网络模型中的边所连接到的节点的活性值或活性测量之间的差异的平方的和。

17.如权利要求16所述的方法,其中差异的平方被由连接第二组节点中的节点的边定义的邻接矩阵加权。

18.如权利要求1-6中任一项所述的方法,其中网络模型是用带符号的拉普拉斯矩阵来表示的,该带符号的拉普拉斯矩阵通过代表第二组节点中的每个节点的外出边的数量的第一对角矩阵、代表第二组节点中的每个节点的进入边的数量的第二对角矩阵、以及由连接第二组节点中的节点的边定义的邻接矩阵而被定义。

19.如权利要求18所述的方法,其中带符号的拉普拉斯矩阵对应于第一总和与第二总和之间的差,第一总和是第一对角矩阵和第二对角矩阵的总和,第二总和是邻接矩阵和该邻接矩阵的转置的总和。

20.如权利要求19所述的方法,其中第二组节点中的特定节点的差异是由活性值向量的转置、带符号的拉普拉斯矩阵以及活性值向量的积来定义的。

21.如权利要求18所述的方法,其中带符号的拉普拉斯矩阵被分解成四个子矩阵,这四个子矩阵包括代表第一组节点中的边的第一子矩阵、代表第一组节点和第二组节点之间的边的第二子矩阵、作为第二子矩阵的转置的第三子矩阵、以及代表第二组节点中的边的第四子矩阵。

22.如权利要求20所述的方法,其中第二组节点中的特定节点的差异是由活性值向量的转置、带符号的拉普拉斯矩阵以及活性值向量的积的偏导数来定义的。

23.如权利要求21所述的方法,其中活性值对应于第四子矩阵的逆矩阵、第三子矩阵和

活性测量的积。

24.一种计算机可读存储介质，包括计算机可读指令，当在包括至少一个处理器的计算机化系统中执行时，该计算机可读指令使处理器执行如权利要求1-23中任一项所述的方法。

用于基于网络的生物活性评估的系统与方法

[0001] 本申请是申请号为201280043499.3,申请日为2012年9月7日,题为“用于基于网络的生物活性评估的系统与方法”的中国发明专利申请的分案申请。

背景技术

[0002] 人体由于暴露给可能有害的制剂(agent)而不断地被扰动,就长期而言,这些制剂会造成严重的健康风险。暴露给这些制剂会危及人体内部的生物机制的正常机能。为了理解并量化这些扰动对人体的影响,研究人员研究了生物系统对暴露给制剂作出响应的机制。有些组大量利用活体动物测试方法。但是,因为关于其可靠性和相关性还存在疑问,所以动物测试方法不总是充分的。不同动物的生理系统中存在许多差异。因此,不同的物种会对暴露给一种制剂有不同的响应。因此,关于从动物测试获得的响应是否可以外推到人类生物学还存在疑问。其它方法包括通过人类志愿者的临床研究评估风险。但是这些风险评估是后验执行的而且,因为疾病可能要花几十年才表现出来,所以这些评估可能不足以说明把有害物质关联到疾病的机制。还有其它方法包括试管(in vitro)实验。虽然,作为其基于动物的对等体的完全或部分代替方法,基于试管细胞与组织的方法已经获得普遍认同,但是这些方法具有有限的价值。因为试管方法集中在细胞和组织机制特定方面;所以它们不总是考虑到发生在整个生物系统中的复杂相互作用。

[0003] 在过去的十年中,核酸、蛋白质与代谢物水平的高吞吐量测量结合传统的依赖剂量的疗效与毒性检测已经作为一种用于说明许多生物过程作用机制的手段出现。研究人员已经尝试结合来自这些全异测量的信息与来自科学文献的关于生化过程的知识来组成有意义的生物模型。为此,研究人员已经开始使用可以挖掘大量数据的数学与计算技术,诸如群集与统计方法,来识别可能的生物作用机制。

[0004] 之前的工作还探究了揭开基因表达变化的特征签名的重要性,这种基因表达变化是由于对生物过程的一个或多个扰动导致的,以及还探究了对那种签名在附加数据集中的存在进行后续打分,作为那个过程的具体活性量值(activity amplitude)的测量。这方面的大部分工作涉及识别与疾病表型(phenotype)关联的签名并给其打分。这些得自表型的签名提供了显著的分类能力,但是缺乏单个具体扰动与该签名之间的机制或因果关系。因此,这些签名可能代表多个截然不同的未知扰动,这些扰动通过常常未知的机制导致相同的疾病表型或者得自于相同的疾病表型。

[0005] 一个挑战在于理解生物系统中各个个别生物实体的活性如何使得能够激活或抑制不同的生物机制。因为个别实体,诸如基因,会在多个生物过程(例如,发炎和细胞增殖)中涉及到,所以基因的活性测量不足以识别触发该活性的底层生物过程。

发明内容

[0006] 这里描述了基于从生物系统中的实体子集测出的活性数据来量化生物系统对一个或多个扰动的响应的系统与方法。当前技术既不适于识别可以在微观级导致生物实体的活性的底层机制,也不提供对不同生物机制的激活的量化评估,其中这些实体响应于可能

有害的制剂和实验条件而起作用。因此，需要改进的系统和方法来鉴于生物机制而分析全系统的生物数据，并且在生物系统对制剂或者环境中的变化作出响应时量化生物系统中的变化。描述了基于测出的活性数据以及描述测量和未测量的实体之间关系的生物系统的网络模型来推断未测量的实体的活性的系统与方法。

[0007] 一方面，在此所述的系统与方法针对用于量化生物系统扰动(例如，响应于诸如制剂暴露的治疗条件(treatment condition)，或者响应于多种治疗条件)的计算机化方法和一个或多个计算机处理器。计算机化方法可以包括在第一处理器接收与第一组生物实体对第一治疗的响应对应的第一组治疗数据。第一生物系统包括第一组生物实体和第二组生物实体。第一生物系统中的每个生物实体与第一生物系统中的至少一个其它生物实体相互作用。计算机化方法还可以包括在第二处理器接收与第一组生物实体对与第一治疗不同的第二治疗的响应对应的第二组治疗数据。在有些实现中，第一组治疗数据代表向制剂的暴露，而第二组治疗数据是控制数据。计算机化方法还可以包括在第三处理器提供第一计算因果网络模型，所述第一计算因果网络模型代表第一生物系统。所述第一计算因果网络模型包括：代表第一组生物实体的第一组节点，代表第二组生物实体的第二组节点，连接节点并且代表生物实体之间的关系的边，及用于节点或边的方向值，代表第一控制数据与第一治疗数据之间预期的变化方向。在有些实现中，边和方向值代表节点之间的因果激活关系。

[0008] 计算机化方法还可以包括利用第四处理器为第一组节点中的对应节点计算代表第一治疗数据与第二治疗数据之间的差异的第一组活性测量。

[0009] 计算机化方法还可以包括基于第一计算因果网络模型和第一组活性测量，利用第五处理器为第二组节点中的对应节点生成第二组活性值。在有些实现中，生成第二组活性值包括为第二组节点中的每个特定节点选择最小化差异声明(difference statement)的活性值，所述差异声明表示特定节点的活性值与该特定节点利用第一计算因果网络模型中的边连接到的节点的活性值或活性测量之间的差异，其中差异声明依赖于第二组节点中每个节点的活性值。该差异声明可以进一步依赖于第二组节点中每个节点的方向值。在有些实现中，第二组活性值中的每个活性值是第一组活性测量中的活性测量的线性组合。特别地，该线性组合可以依赖于第一计算因果网络模型中第一组节点中的节点与第二组节点中的节点之间的边，而且还依赖于第一计算因果网络模型中第二组节点中的节点之间的边，并且可以不依赖于第一计算因果网络模型中第一组节点中的节点之间的边。

[0010] 最后，计算机化方法可以包括基于第一计算因果网络模型和第二组活性值，利用第六处理器为第一计算模型生成代表由第一制剂造成的第一生物系统的扰动的得分。在有些实现中，该得分对第二组活性值具有二次依赖性(quadratic dependence)。该计算机化方法还可以包括，通过为第一组活性测量的每个活性测量形成变差估计(variation estimate)的线性组合，为第二组活性值的每个活性值提供变差估计。例如，用于第二组活性值的每个活性值的变差估计可以是用于第一组活性测量的每个活性测量的变差估计的线性组合。用于得分的变差估计可以对第二组活性值具有二次依赖性。

[0011] 在有些实现中，第二组活性值被表示为第一活性值向量而且该第一活性值向量被分解成第一起作用向量和第一不起作用向量，使得第一起作用和不起作用向量之和是第一活性值向量。所述得分可以不依赖于第一不起作用向量，而且可以作为第二组活性值的二次函数来计算。在这种实现中，第一不起作用向量可以是该二次函数的内核。在有些实现

中，基于与计算因果网络模型(诸如第一计算因果网络模型)关联的带符号拉普拉斯算子(signed Laplacian)，第一不起作用向量是二次函数的内核。

[0012] 上述活性测量(activity measure)和活性值(activity value)可以用于提供反映应用到同一生物系统的不同制剂与治疗条件之间一致或不一致的可比性信息。为此，计算机化方法还可以包括：在第一处理器接收与第一组生物实体对第三治疗的响应对应的第三组治疗数据；在第二处理器接收与第一组生物实体对第四治疗的响应对应的第四组治疗数据；利用第四处理器计算对应于第一组节点的第三组活性测量，第三组活性测量中的每个活性测量代表用于第一组节点中的对应节点的第三组治疗数据与第四组治疗数据之间的差异。计算机化方法还可以包括：基于该计算因果网络模型和第三组活性测量，利用第五处理器生成第四组活性值，第四组活性值中的每个活性值代表用于第二组节点中的对应节点的活性值；以及把第四组活性值表示为第二活性值向量。

[0013] 计算机化方法还可以包括：把第二活性值向量分解成第二起作用向量和第二不起作用向量，使得第二起作用和不起作用向量之和是第二活性值向量；及比较第一和第二起作用向量。在有些实现中，比较第一和第二起作用向量包括计算第一和第二起作用向量之间的相关性，以指示第一和第二组治疗数据的可比性(comparability)。在有些实施例中，比较第一和第二起作用向量包括把第一和第二起作用向量投影到计算网络模型的带符号拉普拉斯算子的图像空间上。在有些实现中，第二组治疗数据包含与第四组治疗数据相同的信息。

[0014] 上述活性测量和活性值可以用于提供反映两个不同生物系统对由相同制剂或治疗条件造成的扰动相似地响应到什么程度的可译性信息(translatability information)。为此，计算机化方法还可以包括：在第一处理器接收与第三组生物实体对与第一治疗不同的第三治疗的响应对应的第三组治疗数据，其中第二生物系统包括多个生物实体，所述生物实体包括第三组生物实体和第四组生物实体，第二生物系统中的每个生物实体都与第二生物系统中的至少一个其它生物实体相互作用。计算机化方法还可以包括：在第二处理器接收与第三组生物实体对与第三治疗不同的第四治疗的响应对应的第四组治疗数据。此外，计算机化方法还可以包括：在第三处理器提供第二计算因果网络模型，所述第二计算因果网络模型代表第二生物系统。所述第二计算因果网络模型包括：代表第三组生物实体的第三组节点，代表第四组生物实体的第四组节点，连接节点并且代表生物实体之间的关系的边，及用于节点的方向值，代表第二控制数据与第二治疗数据之间预期的变化方向。

[0015] 计算机化方法还可以包括：利用第四处理器计算对应于第三组节点的第三组活性测量，第三组活性测量中的每个活性测量代表用于第三组节点中的对应节点的第三组治疗数据与第四组治疗数据之间的差异；以及基于第二计算因果网络模型和第三组活性测量，利用第五处理器生成第四组活性值，该第四组活性值中的每个活性值用于第四组节点中的对应节点。最后，计算机化方法还可以包括：比较第四组活性值与第二组活性值。在有些实现中，比较第四组活性值与第二组活性值包括：应用基于与第一计算因果网络模型关联的带符号拉普拉斯算子和与第二计算因果网络模型关联的带符号拉普拉斯算子的内核规范相关性分析(kernel canonical correlation analysis)。

[0016] 在某些实现中，第一至第六处理器中的每一个都包括在单个处理器或者单个计算

设备中。在其它实现中，第一至第六处理器中的一个或多个跨多个处理器或计算设备分布。

[0017] 在某些实现中，计算因果网络模型包括存在于代表可能原因的节点和代表测量量的节点之间的一组因果关系。在这种实现中，活性测量可以包括倍数变化(fold-change)。倍数变化可以是描述一个节点测量在控制数据与治疗数据之间，或者在代表不同治疗条件的两组数据之间，从初始值到最终值变化了多少的数字。倍数变化数字可以代表两种条件之间生物实体活性的倍数变化的对数。对于由对应节点代表的生物实体，用于每个节点的活性测量可以包括治疗数据与控制数据之间差异的对数。在某些实现中，计算机化方法包括利用处理器为每个生成的得分生成置信间隔(confidence interval)。

[0018] 在某些实现中，生物系统的子集包括，但不限于，细胞增殖机制、细胞应激机制、细胞发炎机制和DNA修复机制中的至少一个。制剂可以包括，但不限于，异类物质，包括生物系统中不存在或者不能从其得出的分子或实体。制剂还可以包括，但不限于，毒素、治疗用化合物、兴奋剂、弛缓剂、天然产物、制造产物和食品物质。制剂可以包括，但不限于，由加热烟草生成的浮质(aerosol)、由燃烧烟草生成的浮质、烟草烟雾和香烟烟雾中至少一种。制剂可以包括，但不限于，镉、汞、铬、尼古丁、特定于烟草的亚硝胺及其代谢物(4-(甲基亚硝氨基)-(3-吡啶)-1-丁酮(NNK)、N'-亚硝基降烟碱(NNN)、N-亚硝基新烟草碱(NAT)、N-亚硝基假木贼碱(NAB)及4-(甲基亚硝氨基)-1-(3-吡啶)-1-丁醇(NNAL))。在某些实现中，制剂包括用于尼古丁替代疗法的产品。

[0019] 在此所述的计算机化方法可以在具有一个或多个计算设备的计算机化系统中实现，每个计算设备都包括一个或多个处理器。一般而言，在此所述的计算机化系统可以包括一个或多个引擎，所述引擎包括一个或多个处理设备，诸如计算机、微处理器、逻辑设备或者配置成具有执行在此所述计算机化方法中一个或多个的硬件、固件和软件的其它设备或处理器。在某些实现中，计算机化系统包括系统响应剖面引擎、网络建模引擎和网络评分引擎。引擎可以不时地互连，并且进一步不时地连接到一个或多个数据库，包括扰动数据库(perturbations database)、可测量值数据库(measurables database)、实验数据数据库和文献数据库。在此所述的计算机化系统可以包括具有通过网络接口通信的一个或多个处理器和引擎的分布式计算机化系统。这种实现对于经多个通信系统的分布式计算可能是合适的。

附图说明

[0020] 在结合附图考虑以下具体描述之后，本公开内容的更多特征、其本质及各种优点将变得显然，贯穿所有附图相同的标号都指相同的部分，其中：

[0021] 图1是用于量化生物网络对扰动的响应的说明性计算机化系统的框图。

[0022] 图2是用于通过计算网络扰动量值(NPA)得分来量化生物网络对扰动的响应的说明性过程的流程图。

[0023] 图3是作为系统响应剖面(systems response profile)基础的数据的图形表示，包括两个制剂、两个参数和N个生物实体的数据。

[0024] 图4是具有若干生物实体及其关系的生物网络的计算模型的说明。

[0025] 图5是用于量化生物系统的扰动的说明性过程的流程图。

[0026] 图6是用于为一组节点生成活性值的说明性过程的流程图。

[0027] 图7是用于提供可比性信息的说明性过程的流程图。

[0028] 图8是用于提供可译性信息的说明性过程的流程图。

[0029] 图9是用于为活性值和NPA得分计算置信间隔的说明性过程的流程图。

[0030] 图10说明了具有骨干节点和支持节点的因果生物网络模型。

[0031] 图11-12是用于确定NPA得分的统计显著性的说明性过程的流程图。

[0032] 图13是用于识别前导骨干与基因节点的说明性过程的流程图。

[0033] 图14是用于量化生物扰动的影响的示例性分布式计算机化系统的框图。

[0034] 图15是可以用于实现在此所述的任何计算机化系统中任何组件的示例性计算设备的框图。

[0035] 图16说明了来自利用相似(顶部)和不相似生物(底部)的两个实验的示例结果。

[0036] 图17-18说明了来自用于量化生物系统的扰动的细胞培养实验的示例结果。

具体实施方式

[0037] 在此所述的是当生物系统被制剂扰动时量化地评估生物系统内变化的量级的计算机化系统与方法。某些实现包括用于计算表示生物系统的一部分当中变化量级的数值的方法。该计算把从一组受控实验获得的一组数据用作输入，在这组受控实验当中，生物系统被制剂扰动。然后，数据被应用到生物系统的特征的网络模型。该网络模型用作模拟和分析的基础，并且代表启用生物系统中感兴趣的特征的生物机制和过程。该特征或者其一些机制和过程会导致疾病的病状和生物系统的不利影响。数据库中所表示的对生物系统的先前知识用于构建网络模型，该网络模型是由关于各种条件下，包括正常条件下和被制剂扰动条件下，各种生物实体的状态的数据填充的。所使用的网络模型是动态的，因为它代表各种生物实体响应于扰动的状态变化，并且可以产生制剂对生物系统的影响的量化和客观评估。还提供了用于操作这些计算机化方法的计算机系统。

[0038] 除其它的之外，由本公开内容的计算机化方法生成的数值可以用于确定由制造产物(为了安全性评估或比较)、包括营养补充的治疗用化合物(为了疗效或健康益处的确定)及环境活性物质(为了对长期暴露的风险及与不利效果和疾病发作的关系的预测)造成的期望或不利生物效果的量级。

[0039] 一方面，在此所述的系统与方法基于被扰动生物机制的网络模型提供了代表被扰动生物系统中变化量级的计算数值。在此被称为网络扰动量值(NPA)得分的数值可以用于概要地表示既定生物机制中各种实体的状态变化。为不同制剂或不同类型扰动获得的数值可以用于相对地比较不同制剂或扰动对生物机制的影响，作为生物系统的特征，该生物机制启用或显现它自己。因而，NPA得分可以用于测量生物机制对不同扰动的响应。术语“得分”在这里一般用于指提供生物系统中变化量级的量化测量的一个值或一组值。这种得分是通过使用本领域中已知的各种数学与计算算法中任意一种并且根据在此所公开的方法，采用从样本或主体获得的一个或多个数据集计算的。

[0040] NPA得分可以帮助研究人员和临床医生改进诊断、实验设计、治疗决定及风险评估。例如，NPA得分可以用于在毒性分析中筛查一组候选的生物机制，以识别最有可能被暴露给潜在有害的制剂所影响到那些生物机制。通过提供对扰动的网络响应的测量，这些NPA得分可以允许分子事件(如通过实验数据测出的)与在细胞、组织、器官或有机体级发生的

表型或生物结果的相关。临床医生可以使用NPA值来比较受制剂影响的生物机制与患者的生理条件，以确定在暴露给该制剂时，患者最有可能经历的健康风险或益处(例如，免疫力低下的患者尤其易受造成强免疫抑制响应的制剂影响)。

[0041] 在这里还描述了用于量化实验数据和生物机制的网络模型的系统与方法，以便启用对相同生物网络的不同实验之间的比较，在此被称为“可比性”。在有些实现中，可比性是通过跨实验数据集比较NPA或其它扰动量化的统计度量来量化的。可比性度量可以帮助识别，例如，两种刺激(诸如TNF和IL1a)对特定生物网络(诸如NFKB)的激活的效果是否被相同的底层生物支持。图16说明了利用相似(顶部)和不相似生物(底部)的两个实验的示例结果。跨所有的测量节点，在顶部的结果中，实验1导致实验系统大约2倍于实验2的响应，这指示实验2诱发与实验1相同的底层生物，尽管程度比较小。在底部的结果中，实验1和实验2之间的每个测量的实验系统响应之间没有相关性，这意味着(除两个实验都得出相同的平均实验响应的事实之外)被两个实验诱发的生物不可比。当比较不同的暴露或者跨不同剂量的相同暴露时，在此所述的可比性测量可以用于识别网络中相似或不相似的生物。这种测量可以把生物学家指引到网络中为了正确理解实验结果或生物响应的其它量化，诸如NPA得分，而需要更深入分析的区域。

[0042] 在此还描述了用于量化实验数据和生物机制的网络模型的系统与方法，以便使得可以在物种、系统或机制之间相似的生物网络之间进行比较，在此被称为“可译性”。可译性测量提供这种物种、系统或机制之间实验扰动数据与得分(诸如NPA得分)的适用性的指示。例如，在此所述的可译性测量可以用于比较活体实验与试管实验、小鼠实验与人体试验、大鼠实验与人体实验，小鼠实验与大鼠实验、、非人灵长类实验与人体实验以及暴露给不同治疗(诸如暴露给制剂)的其它可比物种、系统或机制。

[0043] 图1是用于量化网络模型对扰动的响应的计算机化系统100的框图。具体地，系统100包括系统响应剖面引擎110、网络建模引擎112和网络评分引擎114。引擎110、112和114不时地互连，并且进一步不时地连接到一个或多个数据库，包括扰动数据库102、可测量值数据库104、实验数据数据库106和文献数据库108。如在此所使用的，引擎包括一个或多个处理设备，诸如计算机、微处理器、逻辑设备或者如关于图14所描述的、配置成具有执行一个或多个计算机化操作的硬件、固件和软件的一个或多个其它设备。

[0044] 图2是根据一种实现、用于通过计算网络扰动量值(NPA)得分来量化生物网络对扰动的响应的过程200的流程图。过程200的步骤将描述为由图1的系统100的各种组件来执行，但是这些步骤中任意步骤都可以由任何合适的硬件或软件组件、本地或远程执行，并且可以任何适当的次序安排或者并行执行。在步骤210，系统响应剖面(SRP)引擎110从多个不同的源接收生物数据，而且数据本身可以是多种不同类型。数据包括来自其中生物系统被扰动的实验的数据，以及控制数据。在步骤212，SRP引擎110生成系统响应剖面(SRP)，SRP是生物系统中一个或多个实体响应于制剂对生物系统的提供而变化的程度的表示。在步骤214，网络建模引擎112提供包含多个网络模型的一个或多个数据库，其中一个模型被选择为与感兴趣的制剂或特征相关。选择可以基于对作为该系统的生物功能基础的机制的现有知识来进行。在某些实现中，网络建模引擎112可以利用系统响应剖面、数据库中的网络和先前在文献中描述过的网络来提取系统中实体之间的因果关系，由此生成、精细化或扩展网络模型。在步骤216，网络评分引擎114利用在步骤214中被网络建模引擎112识别出的网

络和在步骤212由SRP引擎110生成的SRP为每个扰动生成NPA得分。在(由网络表示的)生物实体之间的底层关系的背景下,NPA得分量化对扰动或治疗的生物响应(由SRP表示)。为了公开内容的清晰但不是作为限制,以下描述被分成子部分。

[0045] 本公开内容背景下的生物系统是有机体或者有机体的一部分,包括功能部分,有机体在这里被称为主体。主体通常是哺乳动物,包括人。主体可以是人类总体当中个别的人。如在此所使用的,术语"哺乳动物"包括但不限于人、非人的灵长类动物、小鼠、大鼠、狗、猫、牛、羊、马和猪。除人以外的哺乳动物可以有利地用作可以用于提供人类疾病模型的主体。非人主体可以是未修改的,或者是基因修改的动物(例如,转基因动物,或者携带一个或多个基因突变或者沉默基因的动物)。主体可以是雄性或雌性。依赖于操作的目标,主体可以是已经暴露给感兴趣的制剂的主体。主体可以是已经在延长的时间段上暴露给一种制剂,可选地包括研究之前的时间,的主体。主体可以是暴露给一种制剂一段时间但不再与该制剂接触的主体。主体可以是已经被诊断或识别出有一种疾病的主体。主体可以是已经接受过或者正在接受疾病或不利健康状况治疗的主体。主体还可以是呈现具体健康状况或疾病的一个或多个症状或风险因素的主体。主体可以是易感染一种疾病的主体,而且可以是有征兆的或者无征兆的。在某些实现中,所讨论的疾病或健康状况与在延长的时间段上暴露给一种制剂或者使用一种制剂有关。根据有些实现,系统100(图1)包含或生成与感兴趣的一种类型的扰动或结果相关的一个或多个生物系统及其功能机制(统称为"生物网络"或"网络模型")的计算机化模型。

[0046] 依赖于操作的背景,生物系统可以在不同层次定义,这是它涉及群体中个别有机体,一般是一个有机体,器官、组织、细胞类型、细胞器官、细胞成分或者具体个体细胞的功能。每个生物系统都包括一个或多个生物机制或过程,其操作表现为系统的功能特征。再现人类健康状况的既定特征并且适于暴露给感兴趣的制剂的动物系统是优选的生物系统。反映疾病病因学或病理学中所涉及的细胞类型和组织的细胞和器官系统也是优选的生物系统。对于概括尽可能多活体人类生物的主要细胞或器官培养可以给予优先级。匹配试管人类细胞培养与得自活体动物模型的最等效培养也是重要的。这确保利用匹配的试管系统作为参考系统来产生从动物模型到人类生物的翻译连续(translational continuum)。因此,预期供在此所述的系统与方法使用的生物系统可以通过但不限于由功能特征(生物功能、生理功能或者细胞功能)、细胞器官、细胞类型、组织类型、器官、发育阶段或者以上所述的组合来定义。生物系统的例子包括,但不限于,肺、外皮、骨骼、肌肉、神经(中枢和外围)、内分泌、心血管、免疫、循环、呼吸、泌尿、肾脏、肠胃、结肠直肠、肝脏和生殖系统。生物系统的其它例子包括,但不限于,上皮细胞、神经细胞、血细胞、结缔组织细胞、平滑肌细胞、骨骼肌肉细胞、脂肪细胞、卵细胞、精子细胞、干细胞、肺细胞、脑细胞、心肌细胞、喉细胞、咽细胞、食道细胞、胃细胞、肾细胞、肝细胞、乳腺细胞、前列腺细胞、胰腺细胞、岛细胞、睾丸细胞、膀胱细胞、宫颈细胞、子宫细胞、结肠细胞及直肠细胞中的各种细胞功能。有些细胞可以是在适当的培养条件下在试管中培养或者无限地在试管中维持的细胞系的细胞。细胞功能的例子包括,但不限于,细胞增殖(例如,细胞分裂)、退化、再生、老化、由细胞核对细胞活性的控制、细胞到细胞的信令、细胞分化、细胞去分化、分泌、迁移、吞噬作用、修复、细胞凋亡及发育规划(developmental programming)。可以被当作生物系统的细胞成分的例子包括,但不限于,细胞质、细胞骨架、隔膜、核糖体、线粒体、核子、内质网(ER)、高尔基体、溶酶体、DNA、

RNA、蛋白质、肽及抗体。

[0047] 生物系统中的扰动会由于一个或多个制剂在一段时间上通过暴露或者与生物系统的一个或多个部分接触而造成。制剂可以是单一的物质或者是物质的混合，包括其中不是所有组成成分都被识别或特征化的混合物。制剂或者其组成成分的化学与物理属性可能没有被完全特征化。制剂可以通过其结构、其组成成分或者在某种条件下产生该制剂的源来定义。制剂的一个例子是异类物质，即，生物系统中不存在或者不能从其得到的分子或实体，以及在接触生物系统之后从其产生的任何中间物或代谢物。制剂可以是碳水化合物、蛋白质、脂质、核酸、生物碱、维生素、金属、重金属、矿物质、氧、离子、酶、激素、神经传递素、无机化学化合物、有机化学化合物、环境制剂、微生物、颗粒、环境条件、环境力或者物理力。制剂的非限制性例子包括，但不限于，养分、代谢废物、毒药、麻醉剂、毒素、治疗用化合物、兴奋剂、弛缓剂、天然产物、制造产物、食品物质、病原体(朊病毒、病毒、细菌、真菌、原生动物)、其尺寸处于微米范围或者更小的颗粒或实体，以上所述的副产品及以上所述的混合物。物理制剂的非限制性例子包括辐射、电磁波(包括太阳光)、温度的增加或降低、剪切力、流体压力、放电或者一系列放电，或者外伤。

[0048] 有些制剂不会扰动生物系统，除非它以阈值浓度存在或者它与生物系统接触一段时间，或者这二者的组合。可以根据剂量来量化导致扰动的制剂的暴露或接触。因而，扰动会由于长期暴露给制剂而产生。暴露周期可以通过时间单位、通过暴露频率或者通过主体实际或估计的生命周期内的时间百分比来表示。扰动还会由于停止制剂(如上所述)给生物系统的一个或多个部分的供给或者限制制剂对其的供给而造成。例如，扰动会由于养分、水、碳水化合物、蛋白质、脂质、生物碱、维生素、矿物质、氧气、离子、酶、激素、神经传递素、抗体、细胞因子、光的减少供给或缺乏或者由于约束有机体某些部分的运动或者由于强迫或要求锻炼而造成。

[0049] 依赖于生物系统的哪个(哪些)部分被暴露以及暴露状况，制剂会造成不同的扰动。制剂的非限制性例子可以包括由于加热烟草生成的浮质、燃烧烟草生成的浮质、烟草烟雾和香烟烟雾及其任何气体成分或颗粒成分中的任一种。制剂的更多非限制例子包括镉、汞、铬、尼古丁、特定于烟草的亚硝胺及其代谢物(4-(甲基亚硝氨基)-(3-吡啶)-1-丁酮(NNK)、N'-亚硝基降烟碱(NNN)、N-亚硝基新烟草碱(NAT)、N-亚硝基假木贼碱(NAB)及4-(甲基亚硝氨基)-1-(3-吡啶)-1-丁醇(NNAL))，以及用于尼古丁替代疗法的任何产品。制剂或复杂刺激的暴露方式应当反映日常设置中暴露的范围和条件。一组标准的暴露方式可以设计成系统地应用到同等地良好定义的实验系统。每个试验可以设计成收集依赖时间与剂量的数据，以便捕捉早期和晚期事件并且确保代表性的剂量范围被覆盖。但是，本领域普通技术人员将理解，在此所述的系统与方法可以被适配和修改以便适用于所针对的应用，而且在此设计的系统与方法可以在其它合适的应用中采用，而且此类其它的添加与修改将不背离本发明的范围。

[0050] 在各种实现中，在包括对应控制的各种条件下，为基因表达、蛋白质表达或周转(turnover)、微RNA表达或周转、翻译后修改、蛋白质修改、迁移、抗体产生代谢剖面或者以上所述两个或更多个的组合生成高吞吐量的全系统测量。功能结果测量在这里所述的方法中是期望的，因为它们可以总体上用作评估的依靠并且代表疾病病因学中的清晰步骤。

[0051] 如在此所使用的，“样本”指与主体或实验系统隔离的任何生物样本(例如，细胞、

组织、器官或者整个动物)。样本可以包括,但不限于,单个细胞或多个细胞、细胞片段、组织活检、被切除的组织、组织提取物、组织、组织培养提取物、组织培养基、呼出的气体、全血、血小板、血清、血浆、红血球、白血球、淋巴细胞、嗜中性白细胞、巨噬细胞、B细胞或者其子集、T细胞或者其子集、造血细胞的子集、内皮细胞、滑液、淋巴液、腹水、间质液、骨髓、脑脊髓液、胸腔积液、肿瘤浸润、唾液、黏液、痰、精液、汗、尿或者任何其它体液。样本可以通过包括,但不限于,静脉穿刺、排泄、活组织检查、针穿刺、灌洗、刮削、手术切除的手段或者本领域中已知的其它手段从主体获得。

[0052] 在操作过程中,对于给定的生物机制、结果、扰动或者以上所述的组合,系统100可以生成网络扰动量值(NPA)值,这是网络中生物实体响应于治疗条件的状态变化的量化测量。

[0053] 系统100(图1)包括一个或多个计算机化网络模型,这些网络模型与感兴趣的健康状况、疾病或者生物结果相关。这些网络模型中的一个或多个是基于现有的生物知识并且可以从外部源上载并且在系统100中产生。模型还可以基于测量在系统100中重新生成。通过现有知识的使用,可测量的元素有原因地集成到生物网络模型中。以下所述的是代表感兴趣生物系统中的变化或代表对扰动的响应的数据类型,其中所述变化可以用于生成或精细化网络模型。

[0054] 参考图2,在步骤210,系统响应剖面(SRP)引擎110接收生物数据。SRP引擎110可以从许多不同的源接收这种数据,而且数据本身可以是多种不同的类型。SRP引擎110所使用的生物数据可以从文献、数据库(包括来自药品或医疗设备在临床前、临床和临床后试验的数据)、基因组数据库(基因组序列和表达数据,例如,由国家生物技术信息中心进行的基因表达综合(Gene Expression Omnibus)或者由欧洲生物信息研究所进行的阵列实验(ArrayExpress)(Parkinson等人,2010,Nucl.Acids Res.,doi:10.1093/nar/gkq1040.Pubmed ID 21071405))、商业可用的数据库(例如,基因逻辑(Gene Logic),Gaithersburg,MD,USA)或者实验工作取得。数据可以包括来自一个或多个不同源的原始数据,诸如利用为研究特定治疗条件或暴露给特定制剂的效果而具体设计的一个或多个物种的试管、体外、体内实验。试管实验系统可以包括代表人类疾病的关键方面的组织培养或器官培养(三维培养)。在这种实现中,用于这些实验的制剂剂量与暴露方式可以基本上反映在日常使用或活性条件下或者在特殊使用或活性条件下可以对人预期的暴露范围与条件。实验参数与测试条件可以根据期望来选择,以反映制剂的本性与暴露条件、所讨论的生物系统的分子与过程、所涉及的细胞类型与组织、感兴趣的结果以及疾病病因学的各方面。特定的从动物模型导出的分子、细胞或组织可以与特定的人类分子、细胞或组织培养匹配,以改进基于动物的发现的可译性。

[0055] 除其它的之外,由SRP引擎110接收的数据包括,但不限于,某些条件下,涉及核酸(例如,具体DNA或RNA种属的绝对或相对量,DNA序列、RNA序列的变化,三级结构的变化,或者如通过序列化、杂交-尤其是对于微阵列上的核酸、量化的聚合酶链反应,或者本领域中已知的其它技术确定的甲基化模式)、蛋白质/缩氨酸(例如,蛋白质的绝对或相对量,蛋白质、缩氨酸的具体片段,二级或三级结构的变化,或者如由本领域中已知的方法确定的转译后的修改)及功能活性(例如,酶活性、蛋白水解活性、转录调节活性、运输活性、到某些绑定合作伙伴的绑定亲和力)的数据,这些数据中许多都是由高吞吐量实验技术生成的。包括蛋

白质或缩氨酸的转译后修改的修改可以包括，但不限于，甲基化、乙酰化、法尼基化、生物素化、硬脂化、甲酰化、豆蔻酰化、棕榈酰化、香叶酰化、聚乙二醇化、磷酸化、硫酸化、糖基化、糖修改、脂化、脂质修改、泛素化、蛋白质修饰化、二硫化物键合、胱氨酸化、氧化、谷胱甘肽化、羧化、葡萄糖醛酸化及脱氨基。此外，蛋白质可以在转译后通过一系列反应被修改，诸如阿马道里(Amadori)反应、希夫碱(Schiff base)反应及导致糖化的蛋白质产物的美拉德(Maillard)反应。

[0056] 数据还可以包括测出的功能结果，诸如，但不限于，在细胞级包括细胞增殖、发育命运及细胞死亡的那些功能结果，在生理级包括肺容量、血压、锻炼能力的那些功能结果。数据还可以包括疾病活性或严重性的测量，诸如但不限于在疾病某个阶段的肿瘤转移、肿瘤缓解、功能丧失以及生命期望值。疾病活性可以通过临床评估来测量，其结果是一个值或者一组值，这些值可以从既定条件下对来自一个或多个主体的样本(或者样本总体)的评价来获得。临床评估还可以基于由主体对采访或问卷调查提供的响应。

[0057] 这种数据可能已经明确地生成，用于确定系统响应剖面，或者可能已经在先前的实验中产生或者在文献中发表。一般来说，数据包括关于分子、生物结构、生理条件、遗传性状或表型的信息。在有些实现中，数据包括分子、生物结构、生理条件、遗传性状或表型的条件、位置、数量、活性或子结构的描述。如随后将描述的，在临床设置中，数据可以包括从对样本执行的试验或者关于人类主体的观察获得的原始或处理过的数据，其中样本是从暴露给制剂的人类主体获得的。

[0058] 在步骤212，系统响应剖面(SRP)引擎110基于在步骤212接收到的生物数据生成系统响应剖面(SRP)。这个步骤可以包括本底校正、正规化、倍数变化计算、显著性确定及差异响应(例如，差异表达的基因)识别中的一个或多个。SRP是表达生物系统内一个或多个被测实体(例如，分子、核酸、缩氨酸、蛋白质、细胞等)响应于施加到该生物系统的扰动(例如，暴露给制剂)而个别变化的程度的表示。在一个例子中，为了生成SRP，SRP引擎110收集用于应到给定实验系统(例如，“系统-治疗”对)的一组给定参数(例如，治疗或扰动条件)的一组测量。图3说明了两个SRP：包括用于利用变化的参数(例如，暴露给第一治疗制剂的剂量和时间)接受第一治疗306的N个不同生物实体的生物活性数据的SRP 302，和包括用于接受第二治疗308的这N个不同生物实体的生物活性数据的类似的SRP 304。SRP中所包括的数据可以是原始实验数据、处理过的实验数据(例如，被过滤以除去离群值、利用置信估计做标记、对多次试验求平均)、通过计算生物模型生成的数据，或者取自科学文献的数据。SRP可以用任意数量的途径表示数据，诸如绝对值、绝对变化、倍数变化、对数变化、函数和表。SRP引擎110把SRP传递到网络建模引擎112。

[0059] 虽然在前面步骤中得出的SRP代表将从其确定网络扰动量级的实验数据，但生物网络模型才是用于计算和分析的基础。这种分析需要开发与生物系统特征相关的机制与过程的具体网络模型。这种框架提供了超出已经在更经典基因表达分析中所使用的基因列表检查的一层机制理解。生物系统的网络模型是代表动态生物系统并且通过组装关于生物系统的各种基本属性的量化信息来建立的数学结构。

[0060] 这种网络的构造是一个迭代过程。网络边界的勾勒是通过与感兴趣的过程(例如，肺中的细胞增殖)相关的机制与过程的文献调查指导的。描述这些过程的因果关系从先前的知识提取，以便使网络成核。基于文献的网络可以利用包含相关表型端点的高吞吐量数

据集来验证。SRP引擎110可以用于分析数据集，其结果可以用于确认、精细化或生成网络模型。

[0061] 参考图2，在步骤214，利用基于作为感兴趣生物系统特征基础的机制或过程的网络模型，网络建模引擎112使用来自SRP引擎110的系统响应剖面。在某些方面，网络建模引擎112用于识别已经基于SRP生成的网络。网络建模引擎112可以包括用于接收对模型的更新和变化的组件。网络建模引擎112还可以通过结合新的数据并且生成附加的或者精细化的网络模型，迭代网络生成过程。网络建模引擎112还可以方便一个或多个数据集的融合或者一个或多个网络的融合。取自数据库的网络集合可以通过附加的节点、边或者全新的网络手动补充(例如，通过挖掘文献文字以获得对直接被特定生物实体调节的附加基因的描述)。这些网络包含可以启用过程打分的特征。网络拓扑结构被维持；因果关系的网络可以从网络中任何点被跟踪到可测量的实体。另外，模型是动态的而且用于建立它们的假设可以被修改或重新声明并且启用区分组织上下文与物种的适应性。这在新知识变得可用时允许迭代测量和改进。网络建模引擎112可以除去具有低可信度或者是科学文献中冲突的实验结果的主体的节点或边。网络建模引擎112还可以包括可以利用被监督或不被监督的学习方法(例如，度量学习、矩阵完成(matrix completion)、模式识别)推断的附加节点或边。

[0062] 在某些方面，生物系统被建模为包含顶点(或节点)和连接节点的边的数学图。例如，图4说明了具有9个节点(包括节点402和404)和边(406和408)的简单网络400。节点可以代表生物系统中的生物实体，诸如但不限于化合物、DNA、RNA、蛋白质、缩氨酸、抗体、细胞、组织和器官。边可以代表节点之间的关系。图中的边可以代表节点之间的各种关系。例如，边可以代表“绑定到”关系、“在…中表达”关系、“基于表达成型(expression profiling)共同调节的”关系、“抑制”关系、“在手稿中共存”关系或者“共享结构性元素”关系。一般而言，这些类型的关系描述了一对节点之间的关系。图中的节点还可以代表节点之间的关系。因而，有可能表示关系之间的关系，或者一个关系和图中所表示的另一种类型的生物实体之间的关系。例如，代表化学物的两个节点之间的关系可以代表反应。这种反应可以是该反应与抑制该反应的化学物之间的关系中的一个节点。

[0063] 图可以是无向的，意味着在与每条边关联的两个顶点之间没有区别。作为替代，图的边可以从一个顶点指向另一个顶点。例如，在生物背景下，转译调节网络和代谢网络可以建模为有向图。在转译调节网络的图模型中，节点将代表基因，边表示它们之间的转译关系。作为另一个例子，蛋白质-蛋白质相互作用网络描述了有机体蛋白质组中蛋白质之间的直接物理相互作用并且在这种网络中常常没有与相互作用关联的方向。因而，这些网络可以建模为无向图。某些网络可以既有有向的边又有无向的边。构成图的实体和关系(即，节点和边)可以作为相互关连的节点的网存储在系统100中的数据库中。

[0064] 数据库中所代表的知识可以是取自各种不同源的各种不同类型。例如，某些数据可以代表基因组数据库，包括关于基因的信息以及它们之间的关系。在这种例子中，一个节点可以代表致癌基因，而连接到该致癌基因节点的另一个节点可以代表抑制该致癌基因的基因。数据可以代表蛋白质以及它们之间的关系、疾病和它们的相互关连，以及各种疾病状态。有许多不同类型的数据可以在图形化的表示中组合。计算模型可以代表节点之间的关系网，其中节点代表例如DNA数据集、RNA数据集、蛋白质数据集、抗体数据集、细胞数据集、组织数据集、器官数据集、医疗数据集、流行病数据集、化学物数据集、毒物数据集、患者数

据集和人口数据集中的知识。如在此所使用的，数据集是在既定条件下从对一个样本(或一组样本)进行评价所得到的数值的集合。数据集可以通过例如实验测量样本的可量化实体来获得；或者作为替代，或者从诸如实验室的服务提供者、临床研究机构或者从公共或私人数据库获得。数据集可以包含数据和由节点表示的生物实体，而且每个数据集中的节点可以关连到同一数据集中或者其它数据集中的其它节点。而且，网络建模引擎112可以生成代表从例如DNA、RNA、蛋白质或抗体数据集中的基因信息到医疗数据集中的医疗信息再到患者数据集中关于个别患者的信息再到流行病数据集中关于整个人口的信息的计算模型。除了以上所述的各种数据集，还可以有许多其它数据集或者在生成计算模型时可以包括的生物信息类型。例如，数据库可以进一步包括医疗记录数据、结构/活性关系数据、关于感染性病变的信息、关于临床试验的信息、暴露模式数据、关于一种产品的使用历史的数据，以及任何其它类型的生命科学相关的信息。

[0065] 网络建模引擎112可以生成代表例如基因之间调节相互作用、蛋白质之间相互作用或者细胞或组织内部复杂的生化相互作用的一个或多个网络模型。由网络建模引擎112生成的网络可以包括静态与动态模型。网络建模引擎112可以采用任何适用的数学方案来表示系统，诸如超图和加权二分图(bipartite graph)，其中两种类型的节点用于表示反应和化合物。网络建模引擎112还可以使用其它推理技术来生成网络模型，诸如基于对差异表达基因中功能相关的基因的过表示的分析，贝叶斯网络分析、图形高斯模型技术或者基因相关性网络技术，来基于一组实验数据(例如，基因表达、代谢浓度、细胞反应等)识别相关的生物网络。

[0066] 如上所述，网络模型是基于作为生物系统功能特征的基础的机制与过程。网络建模引擎112可以生成或包含代表关于与制剂的长期健康风险或健康益处的研究相关的生物系统特征的结果的模型。因此，网络建模引擎112可以生成或包含用于细胞功能的各种机制的网络模型，尤其是关于生物系统中感兴趣的特征或者对其起作用的那些细胞功能，包括但不限于细胞增殖、细胞应激、细胞再生、细胞凋亡、DNA损坏/修复或者炎症反应。在其它实施例中，网络建模引擎112可以包含或生成与急性全身毒性、致癌性、皮肤渗透、心血管疾病、肺病、生态毒性、眼睛冲洗/腐蚀、基因毒性、免疫毒性、神经毒性、药代动力学、药物代谢、器官毒性、生殖与发育毒性、皮肤刺激/腐蚀或者皮肤致敏相关的计算机化模型。一般而言，网络建模引擎112可以包含或生成用于核酸(DNA、RNA、SNP、siRNA、miRNA、RNAi)、蛋白质、缩氨酸、抗体、细胞、组织、器官以及任何其它生物实体的状态，及其对应相互作用的计算机化模型。在一个例子中，计算网络模型可以用于表示免疫系统的状态及免疫响应或炎症反应过程中各种类型白血细胞的机能。在其它例子中，计算网络模型可以用于表示心血管系统的性能及内皮细胞的机能与代谢。

[0067] 在本公开内容的有些实现中，网络取自因果关系生物知识的数据库。这种数据库可以通过对不同生物机制执行实验研究以提取机制之间的关系(例如，激活或抑制关系)来生成，其中一些关系可以是因果关系，而且可以与商业可用的数据库，诸如由位于美国麻省剑桥市的Selventa公司产生的Genstruct技术平台或者Selventa知识库，组合。利用因果生物知识数据库，网络建模引擎112可以识别链接扰动102和可测量值104的网络。在某些实现中，网络建模引擎112利用来自SRP引擎110的系统响应剖面和之前在文献中生成的网络提取生物实体之间的因果关系。除其它的处理步骤之外，数据库可以被进一步处理，以除去逻

辑不一致性并通过在不同的生物实体集合之间应用同源推理(homologous reasoning)来生成新生物知识。

[0068] 在某些实现中,从数据库提取出的网络模型是基于反向因果推理(RCR),一种自动化的推理技术,这种技术处理因果关系网络来用公式表达机制假说,然后对照差异测量数据集评价那些机制假说。每个机制假说都把一个生物实体链接到它可以影响的可测量量。例如,除其它的之外,可测量量可以包括浓度的增加或减小、生物实体的个数或相对充裕度(abundance)、生物实体的激活或抑制,或者生物实体结构、功能或逻辑的变化。RCR使用生物实体之间有向的实验观察因果相互作用网络作为计算的基础。有向网络可以用生物表达语言(Biological Expression Language™)(BEL™)表达,这是用于记录生物实体之间相互关系的一种语法。RCR计算规定用于网络模型生成的某些约束,诸如但不限于路径长度(连接上游节点和下游节点的边的最大条数),以及把上游节点连接到下游节点的可能因果路径。RCR的输出是代表实验测量中差异的上游控制器的一组机制假说,通过评价相关性与准确性的统计数据给这组假说分级。机制假说输出可以组装成因果链并且越大的网络以越高的互连机制与过程级解释数据集。

[0069] 一种类型的机制假说包括在代表可能原因的节点(上游节点或控制器)与代表测量量的节点(下游节点)之间的一组因果关系。这种类型的机制假说可以用于进行预测,诸如,如果上游节点表示的实体的充裕度增加,则推断出由因果增加关系链接的下游节点将增加,而且推断出由因果减少关系链接的下游节点将减少。

[0070] 机制假说表示一组测量数据,例如基因表达数据,与作为那些基因的已知控制器的生物实体之间的关系。此外,这些关系包括上游实体之间的影响的符号(正或负)及下游实体(例如,下游基因)的差异表达。机制假说的下游实体可以取自文献产生的因果生物知识的数据库。在某些实现中,以可计算的因果网络模型的形式,把上游实体链接到下游实体的机制假说的因果关系是由NPA打分方法用于计算网络变化的基础。

[0071] 在某些实现中,通过收集模型中表示生物系统中各种特征的个别机制假说并且把所有下游实体(例如,下游基因)的连接重新分组到单个上游实体或过程,生物实体的复杂因果网络模型可以变换成单个因果网络模型,由此来表示整个复杂因果网络模型;这本质上是底层图结构的扁平化。因而,如在网络模型中所表示的生物系统的特征与实体的变化可以通过组合个别的机制假说来评估。在有些实现中,因果网络模型中的一个节点子集(在这里被称为“骨干节点”)代表对应于不测量或者不能常规或经济地测量的实体的第一组生物实体,例如,生物系统中关键角色的生物机制或活性;还有另一个节点子集(在这里被称为“支持节点”)代表生物系统中可以被测量并且其值是实验确定并且在数据集中给出以便计算例如生物系统中多个基因的表达级别的第二组生物实体。图10绘出了一个示例性网络,该网络包括四个骨干节点1002、1004、1006和1008,及骨干节点之间和从骨干节点到支持基因表达节点组1010、1012和1014的边。图10中的每条边都是有向的(即,代表原因和效果关系的方向)和带符号的(即,代表正或负调节)。这种类型的网络可以代表在某些生物实体或机制(例如,范围从具体到如特定酶的充裕度或激活的增加的量到复杂到如反映生长因素信令过程状态的量)与被正或负调节的其它下游实体(例如,基因表达级别)之间存在的一组因果关系。

[0072] 在某些实现中,当细胞暴露给香烟烟雾时,系统100可以包含或者生成用于细胞增

殖机制的计算机化模型。在这种例子中，系统100还可以包含或生成代表与香烟烟雾暴露相关的各种健康状况的一个或多个网络模型，其中的健康状况包括但不限于癌症、肺病和心血管疾病。在某些方面，这些网络模型是基于所施加的扰动(例如，暴露给制剂)、各种条件下的响应、感兴趣的可测量量、所研究的结果(例如，细胞增殖、细胞应激、发炎、DNA修复)、实验数据、临床数据、流行病数据和文献中的至少一个。

[0073] 作为一个说明性例子，网络建模引擎112可以配置为用于生成细胞应激的网络模型。网络建模引擎112可以接收描述从文献数据库已知的应激响应中所涉及的相关机制的网络。网络建模引擎112可以基于已知响应于肺和心血管背景下应激而操作的生物机制选择一个或多个网络。在某些实现中，网络建模引擎112识别生物系统中的一个或多个功能单元并且通过基于它们的功能性组合较小的网络来建立更大的网络。特别地，对于细胞应激模型，网络建模引擎112可以考虑与对氧化、基因毒性、缺氧、渗透、异型生物质和剪切力的响应相关的功能单元。因此，用于细胞应激模型的网络组件可以包括异型生物质代谢响应、基因毒性应激、内皮剪切力、缺氧响应、渗透应力和氧化应力。网络建模引擎112还可以从公共可用的转录数据的计算分析接收内容，这些转录数据是来自在特定细胞组中执行的应力相关实验。

[0074] 当生成生物机制的网络模型时，网络建模引擎112可以包括一条或多条规则。这种规则可以包括用于选择网络内容、节点类型等的规则。网络建模引擎112可以从实验数据数据库106选择一个或多个数据集，包括试管和活体实验结果的组合。网络建模引擎112可以利用实验数据来验证在文献中识别出的节点与边。在建模细胞应激的例子中，网络建模引擎112可以基于实验多好地表示无病的肺或心血管组织中生理相关应力来选择用于实验的数据集。数据集的选择可以基于，例如，表型应力端点数据的可用性、基因表达概述实验的统计严密性以及实验背景与正常无病的肺或心血管生物学的相关性。

[0075] 在识别出相关网络的集合之后，网络建模引擎112可以进一步处理并精细化那些网络。例如，在有些实现中，多个生物实体及它们的连接可以分组并且被新的一个或多个节点表示(例如，利用群集或者其它技术)。

[0076] 网络建模引擎112还可以包括关于所识别出的网络中的节点与边的描述性信息。如以上所讨论的，例如，节点可以通过其关联的生物实体、相关联的生物实体是否是可测量量的指示或者生物实体的任何其它描述符来描述，而边可以通过它所代表的关系类型(例如，诸如上调节或下调节的因果关系、相关、条件依赖或独立)、那个关系的强度或者那个关系中的统计置信度来描述。在有些实现中，对于每种治疗，表示可测量实体的每个节点都与响应于该治疗而活性变化的预期方向(即，增加或减少)关联。例如，当支气管上皮细胞暴露给诸如肿瘤坏死因子(TNF)的制剂时，特定基因的活性可以增加。这种增加可能由于从文献知道(并且在由网络建模引擎112识别出的一个网络中表示)或者由通过由网络建模引擎112识别出的一个或多个网络的边来跟踪多个调节关系(例如，自分泌信令)知道的直接调节关系所导致。在有些情况下，响应于特定的扰动，网络建模引擎112可以为每个可测量的实体识别预期的变化方向。当网络中的不同过程对于一个特定实体指示矛盾的预期变化方向时，可以更具体地检查这两个过程，以确定净变化方向，或者那个特定实体的测量可以被丢弃。

[0077] 这里提供的计算方法与系统基于实验数据与计算网络模型计算NPA得分。计算网

络模型可以由系统100生成、导入系统100中或者在系统100中识别(例如,从生物知识的数据库)。被识别为网络模型中扰动的下游效果的实验测量在生成特定于网络的响应得分时组合。因此,在步骤216,网络评分引擎114利用在步骤214由网络建模引擎112识别出的网络和在步骤212由SRP引擎100生成的SRP为每个扰动生成NPA得分。在生物实体的底层关系(用识别出的网络标识)的背景下,NPA得分量化对治疗的生物响应(用SRP表示)。网络评分引擎114可以包括用于为网络建模引擎112中所包含的或者被其识别出的每个网络生成NPA得分的硬件与软件组件。

[0078] 网络评分引擎114可以配置为实现多种技术中任意一种,包括生成指示网络对扰动的响应的量级与拓扑分布的标量-或向量-值得分的技术。

[0079] 在某些应用中,额外的评分技术可以有利地应用,并且可以扩展成启用关于相同生物网络的不同实验之间的比较(在这里被称为“可比性”)或者物种、系统或机制之间类似生物网络之间的比较(在这里被称为“可译性”)。现在描述多种评分技术及用于评估可比性和可译性的技术。

[0080] 图5是用于量化响应于制剂的生物系统扰动的说明性过程500的流程图。例如,过程500可以由网络评分引擎114或者系统100的任何其它合适配置的一个或多个组件实现。特别地,第一组生物实体可以被测量(即,对第一组生物实体测量治疗数据和控制数据),而第二组生物实体不能测量(即,不对第二组生物实体测量治疗数据和控制数据)。出于任何数量的原因,可能不能为第二组生物实体容易地获得数据(或者只能获得有限的量)。作为例子,对应于第二组生物实体的数据可能是特别难获得的,或者第二组生物实体可能与另一组容易测量的生物实体相关,使得数据可以从可测量的集合合理地推断。

[0081] 为了量化响应于制剂的生物系统的扰动,网络评分引擎114可以计算NPA得分,这是代表生物机制对扰动的响应的数值。计算NPA得分的一条途径是只使用直接测出的数据(即,对应于以上例子中第一组生物实体)。但是,这种方法局限于有可能用于确定扰动对生物机制影响的数据子集。具体地,可以存在不直接测量但是可以提供用于NPA得分的信息的另一组生物实体(即,对应于以上例子中第二组生物实体)。在这种情况下,未测量的生物实体集合可以与测出的集合相关,使得网络评分引擎114可以从可测量的集合推断与未测量集合相关的数据。因而,NPA得分可以基于测出的数据、推断出的数据或者这二者的组合。图5中的过程500描述了基于推断出的数据计算NPA得分的方法。

[0082] 在步骤502,网络评分引擎114接收用于生物系统中第一组生物实体的治疗与控制数据。治疗数据对应于第一组生物实体对制剂的响应,而控制数据对应于第一组生物实体对缺乏该制剂的响应。生物系统包括第一组生物实体(在步骤502中为其接收了治疗和控制数据),以及第二组生物实体(没有为其接收治疗和控制数据)。生物系统中的每个生物实体与该生物系统中的至少一个其它生物实体相互作用,并且特别地,第一组中的至少一个生物实体与第二组中的至少一个生物实体相互作用。生物系统中生物实体之间的关系可以由一个计算网络模型来表示,这个模型包括代表第一组生物实体的第一组节点、代表第二组生物实体的第二组节点、以及连接节点并且代表生物实体之间关系的边。该计算网络模型还可以包括用于节点的方向值,这代表控制与治疗数据之间预期的变化方向(例如,激活或抑制)。以上具体描述了这种网络模型的例子。

[0083] 在步骤504,网络评分引擎114为第一组生物实体中的生物实体计算活性测量。第

一组活性测量中的每个活性测量代表对于第一组中一个特定生物实体的治疗数据与控制数据之间的差异。因为第一组生物实体与计算网络模型中第一组节点之间的对应性，所以步骤504也为计算网络模型中的第一组节点计算活性测量。在有些实现中，活性测量可以包括倍数变化。倍数变化可以是描述一个节点测量在控制数据与治疗数据之间，或者在代表不同治疗条件的两组数据之间，从初始值到最终值变化了多少的一个数字。倍数变化数字可以代表两种条件之间生物实体活性的倍数变化的对数。用于每个节点的活性测量可以包括，对于由对应节点表示的生物实体，治疗数据与控制数据之间的差异的对数。在某些实现中，计算机化方法包括利用处理器生成用于每个所生成得分的置信间隔。

[0084] 在步骤506，网络评分引擎114为第二组生物实体中的生物实体生成活性值。因为没有为第二组中的生物实体接收治疗和控制数据，所以在步骤506生成的活性值代表推断出的活性值，并且是基于第一组活性测量和计算网络模型。为第二组生物实体(对应于计算网络模型中的第二组节点)推断出的活性值可以根据多种推断技术中的任何一种生成；以下参考图6描述几种实现。在步骤506为非测量实体生成的活性值利用由网络模型提供的实体之间的关系阐明了不直接测量的生物实体的行为。

[0085] 在步骤508，网络评分引擎114基于在步骤506生成的活性值计算NPA得分。NPA得分代表由制剂造成的生物系统扰动(如在控制与治疗数据之间的差异中所反映的)，并且是基于在步骤506生成的活性值及计算网络模型。在有些实现中，在步骤508计算的NPA得分可以根据下式计算：

[0086] $$NPA(G,\beta)=\frac{1}{|\{x\to y\}\ s.t.\ x,y\in V_0|}\sum_{\substack{x\to y\\ s.t.\ x,y\in V_0}}(f(x)+sign(x\to y)f(y))^2, \quad (1)$$

[0087] 其中，V_0表示第一组生物实体(即，在步骤502为其接收治疗和控制数据的那些实体)，f(x)表示在步骤508为生物实体x生成的活性值，而sign(x→y)表示计算网络模型中把表示生物实体x的节点连接到表示生物实体y的节点的边的方向值。如果与第二组生物实体关联的活性值的向量表示为f2，则网络评分引擎114可以配置为经二次形式计算NPA得分：

[0088] $$NPA=f_2^TQf_2, \quad (2)$$

[0089] 其中

[0090] $$Q=\frac{1}{|\{x\to y\}\ s.t.\ x,y\in V_0|}\left[\left(diag(out|_{l^2(V\setminus V_0)})+diag(in|_{l^2(V\setminus V_0)})-(-A-A^T)\right)|_{l^2(V\setminus V_0)}\right]\in l^2(V\setminus V_0) \quad (3)$$

[0091] diag(out)表示对角矩阵，其具有第二组节点中每个节点的出度(out-degree)，diag(in)表示对角矩阵，其具有第二组节点中每个节点的入度(in-degree)，而A表示仅限于第二组中根据下式定义的那些节点的计算网络模型的邻接矩阵：

[0092] $$A_{xy}=\begin{cases}sign(x\to y) & if\ x\to y\\ 0 & else\end{cases}. \quad (4)$$

[0093] 如果A是加权的邻接矩阵，则A的元素(x,y)可以用权重因子w(x→y)去乘。

[0094] 步骤508还可以包括为NPS得分计算置信间隔。在有些实现中，假定活性值f2遵循多元正态分布N(μ，Σ)，于是根据等式2计算出的NPA得分将具有可以根据下式计算的相关联的变差(variance)：

[0095] $$var(f^TQf)=2tr(Q\Sigma Q\Sigma)+4\mu^TQ\Sigma Q\mu. \quad (5)$$

在有些实现中，诸如根据等式5操作的那些实现，NPA得分具有对活性值的二次依赖性。网络评分引擎114可以进一步配置为，除其它方法之外，通过采用切比雪夫(Chebyshev)不等式或者依赖于中心极限定理，使用根据等式5计算出的变差来生成保守置信间隔。

[0096] 图6是用于为一组节点生成活性值的说明性过程600的流程图。例如，过程600可以在图5过程500的步骤506执行，并且为了容易说明而描述为由网络评分引擎114执行。在步骤602，网络评分引擎114识别差异声明(difference statement)。差异声明可以是代表特定生物实体的活性测量或值与该特定生物实体连接到的生物实体的活性测量或值之间差异的表达或其它可执行声明。在表示感兴趣的生物系统的计算网络模型的语言中，差异声明代表网络模型中特定节点的活性测量或值与该特定节点经边连接到的节点的活性测量或值之间的差异。差异声明可以依赖于计算网络模型中的任何一个或多个节点。在有些实施例中，差异声明依赖于以上关于图5步骤506描述的第二组节点中每个节点(即，没有治疗或控制数据对其可用的那些节点，而且其活性值是从与其它节点关联的治疗或控制数据及计算网络模型推断出来的)的活性值。

[0097] 在有些实现中，网络评分引擎114在步骤602识别以下差异声明：

[0098] $$\sum_{x\to y}(f(x)-sign(x\to y)f(y))^2 w(x\to y), \tag{6}$$

[0099] 其中f(x)表示活性值(对于第二组节点中的节点x)或者测量(对于第一组节点中的节点x)，sign(x→y)表示计算网络模型中把代表生物实体x的节点连接到代表生物实体y的节点的边的方向值，而w(x→y)表示与连接代表实体x和y的节点的边关联的权重。为了容易说明，剩余的讨论将假定w(x→y)等于一，但本领域普通技术人员将通过对等式6的差异声明的讨论容易地跟踪非单位权重(即，通过使用如以上参考等式4描述的加权邻接矩阵)。

[0100] 网络评分引擎114可以许多不同的途径实现等式6的差异声明，包括以下等价声明中任何一个：

[0101] $$\begin{aligned}&\textstyle\sum_{x\to y}(f(x)-sign(x\to y)f(y))^2\\&=\textstyle\sum_x\sum_{y:x\to y}f(x)^2+f(y)^2-2sign(x\to y)f(x)f(y)\\&=\textstyle\sum_x f(x)^2\cdot out(x)+\sum_y f(y)^2\cdot in(y)-2\sum_{x,y}sign(x\to y)f(x)f(y)\\&=f^T\big(diag(out)+diag(in)\big)f-f^T(A+A^T)f.\end{aligned}\tag{7}$$

[0102] 在步骤604，网络评分引擎114识别差异目标。差异目标代表网络评分引擎114将朝着其选择用于第二组生物实体的活性值的差异声明的值的优化目标。差异目标可以规定差异声明要最大化、最小化或者尽可能接近目标值。差异目标可以规定要为其选择活性值的生物实体，并且可以确立对为每个实体所允许的活性值范围的约束。在有些实现中，在第一组生物实体(即，治疗和控制数据对其可用的那些实体)的活性等于在图5步骤504计算出的活性测量的约束下，差异目标要对以上参考图5步骤506讨论的第二组节点中的所有生物实体最小化等式6的差异声明。这个差异目标可以写成以下计算优化问题：

[0103] $$\operatorname{argmin}_{f\in l^2(V)}\textstyle\sum_{x\to y}(f(x)-sign(x\to y)f(y))^2\cdot w(x\to y)\ \ such\ that\ f|_{V_0}=\beta, \tag{8}$$

[0104] 其中β表示为第一组中每个实体在图5步骤504计算出的活性测量。

[0105] 为了解决在步骤604中识别出的差异目标，网络评分引擎114配置为前进到步骤

606,以基于该差异目标计算性地特征化网络模型。表示生物系统的计算网络模型可以任何多种途径特征化(例如,经如上讨论的加权或不加权的邻接矩阵A)。不同的特征化可以更好地适合不同的差异目标,从而改善网络评分引擎114在计算NPA得分过程中的性能。例如,当差异目标根据以上等式8用公式表示时,网络评分引擎114可以配置为利用根据下式定义的带符号拉普拉斯矩阵特征化计算网络模型:

[0106] $L=\mathrm{diag}(\mathrm{out})+\mathrm{diag}(\mathrm{in})-(A+A^T)$. (9)

给定这种特征化,等式8的差异目标可以表示为:

[0107] $$\mathrm{argmin}_{f\in l^2(V)} f^T L f \ \ such\ that\ f|_{V_0}=\beta. \tag{10}$$

[0108] 网络评分引擎114可以配置为通过把网络模型分成四个组成部分在第二级特征化计算网络模型:第一组节点内部的连接、从第一组节点到第二组节点的连接、从第二组节点到第一组节点的连接以及第二组节点内部的连接。从计算上来说,网络评分引擎114可以通过把拉普拉斯矩阵分成四个子矩阵(这些组成部分中每个部分一个子矩阵)并且把活性向量f分成两个子向量(第一组节点的活性f_1一个子向量并且第二组节点的活性f_2一个子向量)实现这个附加特征化。等式10的差异声明的这种重新特征化可以写成:

[0109] $$f^T\begin{pmatrix}L_1 & L_2\\ L_2^T & L_3\end{pmatrix}f=(f_1^T\ f_2^T)\begin{pmatrix}L_1 & L_2\\ L_2^T & L_3\end{pmatrix}\begin{pmatrix}f_1\\ f_2\end{pmatrix}=f_1^TL_1f_1+f_1^TL_2f_2+f_2^TL_2^Tf_1+f_2^TL_3f_2\,. \tag{11}$$

[0110] 在步骤606,网络评分引擎114选择实现或近似差异目标的活性值。许多不同的计算优化例程在本领域中是已知的,并且可以适用于在步骤604中识别出的任何差异目标。在其中等式10的差异目标在步骤604中识别出的实现中,网络评分引擎114可以配置为通过取等式11对f2的导数(数值的或者分析的)、把该导数设置成等于零并且重排以便隔离用于f2的表达来选择最小化等式11表达的f2的值。由于

[0111] $$\frac{\partial}{\partial f_2}(f^TLf)=2L_2^Tf_1+2L_3f_2\,, \tag{12}$$

[0112] 因此网络评分引擎114可以配置为根据下式计算f2:

[0113] $$f_2=-L_3^{-1}L_2^Tf_1\equiv Kf_1. \tag{13}$$

[0114] 由于f1是为第一组生物实体(治疗和控制数据对其可用的实体)计算出的活性测量的向量,因此用于第二组生物实体的活性值可以表示为根据等式13计算出的活性测量的线性组合。就像在等式13中一样,活性值可以依赖于第一计算网络模型(即,L_2)中第一组节点中的节点与第二组节点中的节点之间的边,并且还可以依赖于计算因果网络模型(即,L_3)中第二组节点中节点之间的边。在有些实现中(诸如根据等式13操作的那些实现),活性值不依赖于计算网络模型中第一组节点中节点之间的边。

[0115] 在步骤608,网络评分引擎114提供在步骤606生成的活性值。在有些实现中,活性值显示给用户。在有些实现中,活性值在图5的步骤508中用于计算NPA得分,如上所述。在有些实现中,用于活性值的变差和置信信息也可以在步骤608生成。例如,如果活性值和测量可以假定为大致遵循多元正态分布,$N(\mu,\Sigma)$,则Af也将遵循多元正态分布,其中:

[0116] $\mathrm{var}(Af)=A\Sigma A^T$. (14)

在这种情况下,用于推导出的活性值的置信间隔可以利用其中 $A=-L_3^{-1}L_2^T$ 和$\Sigma=\mathrm{diag}(\mathrm{var}(\beta))$的标准统计技术来计算。

[0117] 在图5步骤504计算出的活性测量和在图5步骤506生成的活性值(例如,根据图6的过程600)可以用于提供反映应用到相同生物系统的不同制剂与治疗条件之间一致性或不一致性的可比性信息。图7是用于提供可比性信息的说明性过程700的流程图。例如,在图5的步骤506中生成用于第二组节点的活性值之后,过程700可以由网络评分引擎114或系统100的任何其它合适配置的一个或多个组件执行。

[0118] 在步骤702,网络评分引擎114把第一组活性值表示为第一活性值向量。这种类型的表示在上面参考等式11讨论过了,其中一组活性值表示为向量f2。在步骤704,网络评分引擎114把第一活性值向量分解成第一起作用向量和第一不起作用向量。第一起作用向量和第一不起作用向量依赖于活性值向量与NPA得分之间的关系。如果NPA得分表示为第一活性值向量v1的变换g,使得

[0119] NPA＝g(h(v1)), (15)

则v1可以在步骤704分成两个向量v1c和v1nc之和,使得

[0120] v1＝v1c+v1nc (16)

[0121] 并且

[0122] g(v1nc)＝0. (17)

数学上,当g严格正定时,不起作用的向量v1nc被说成是变换h的内核,而起作用的向量v1c被说成是在变换h的图像空间中。标准计算技术可以适用于确定各种类型变换的内核与图像空间。如果网络评分引擎114根据等式5和13从活性值向量v1计算出一个NPA得分,则那个NPA得分变换的内核是矩阵积$(L_3^{-1}L_2^T)$的内核,而且那个NPA得分变换的图像空间是矩阵积$(L_3^{-1}L_2^T)$的图像空间。因而,活性值向量可以利用标准的计算投影技术分解成处于矩阵积$(L_3^{-1}L_2^T)$的图像空间中的起作用向量v1c和矩阵积$(L_3^{-1}L_2^T)$的内核中的不起作用向量v1nc,并且NPA得分不会依赖于不起作用的向量v1nc。

[0123] 由于NPA得分可以计算为二次形式(如上所示),因此,即使输入数据不反映模型中机制的实际扰动,网络评分引擎114也会生成显著的(关于生物变异)得分。为了评估网络是否实际被扰动(即,模型中所描述的生物是否在数据中反映出来),同伴统计可以用于帮助确定提取出的信号是否特定于该网络结构还是所收集到的数据中固有的。几种类型的置换测试对于评估观察到的信息是否更代表数据固有的属性还是由因果生物网络模型给出的结构可能是特别有用的。

[0124] 图11和12说明了给定因果网络模型和具体的数据集,可以由网络评分引擎114用于确定推荐的NPA得分的统计显著性的过程1100和1200。确定推荐的NPA得分的统计显著性对于指示由该网络建模的生物系统是否被扰动会是有用的。为了确定所推荐的NPA得分的统计显著性,网络评分引擎114可以让数据接受如下所述的一个或两个测试。

[0125] 这两个测试(在此被称为置换测试)都基于生成因果网络模型的一个或多个方面的随机置换,利用结果产生的测试模型基于与生成所推荐NPA得分的相同的数据集和算法来计算测试NPA得分,并且比较测试NPA得分与推荐的NPA得分或者给测试NPA得分和推荐的NPA得分分级,以确定所推荐的NPA得分的统计显著性。可以随机分类以生成测试模型的因果网络模型的各方面包括支持节点的标签、把骨干节点连接到支持节点的边,或者把骨干

节点彼此连接的边。

[0126] 在一种实现中，在此被称为“O-统计”测试的置换测试评估因果网络模型中支持节点位置的重要性。过程1100包括评估计算出的NPA得分的统计显著性的方法。特别地，在步骤1102，基于对生物系统中实体的因果关系的知识，第一推荐NPA得分基于网络来计算，该网络也称为未修改的网络。在步骤1106，基因标签并且因此每个支持节点的对应值在网络模型中的支持节点当中随机地重新分配。随机重新分配重复多次，例如，C次，并且在步骤1112，测试NPA得分基于该随机重新分配来计算，导致C个测试NPA得分的分布。网络评分引擎114可以根据以上为基于网络计算NPA得分所述的任何方法计算推荐的和测试NPA得分。在步骤1114，推荐的NPA得分与测试NPA得分的分布进行比较或者对照其进行分级，以确定推荐的NPA得分的统计显著性。

[0127] 在某些实现中，量化生物系统的扰动的方法包括基于因果网络模型计算推荐的NPA得分，并且确定该得分的统计显著性。显著性可以由一种方法来计算，该方法包括随机地重新分配因果网络模型的支持节点的标签以创建测试模型、基于该测试模型计算测试NPA得分并且比较推荐的NPA得分与测试NPA得分以确定生物系统是否被扰动。支持节点的标签与活性测量关联。

[0128] 整数C可以是由网络评分引擎确定的任何数字并且可以基于用户输入。整数C可以足够大，使得基于随机重新分配产生的NPA得分的分布大致是平滑的。整数C可以是固定的，使得重新分配执行预定多次。作为替代，整数C可以依赖结果产生的NPA得分变化。例如，整数C可以迭代增加，并且，如果结果产生的NPA分布不平滑，则可以执行附加的重新分配。此外，任何其它对分布的附加需求都可以使用，诸如增加C直到分布像某种形式，诸如高斯或任何其它合适的分布。在某些实现中，整数C从大约500到大约1000变动。

[0129] 在步骤1110，网络评分引擎114基于在步骤1106生成的随机重新分配计算C个NPA得分。特别地，NPA得分为在步骤1106生成的每个重新分配计算。在某些实现中，所有C次重新分配都首先在步骤1106生成，然后对应的NPA得分在步骤1110基于C次重新分配计算。在其它实现中，对应的NPA得分在生成每个重新分配集合之后计算，并且这个过程重复C次。如果C的值依赖于前面计算出的N个值，则后一种场景可以节省存储器成本并且可能是期望的。在步骤1112，网络评分引擎114汇聚结果产生的C个NPA得分，以形成或生成NPA值的分布，对应于在步骤1106生成的随机重新分配。该分布可以对应于NPA值的柱状图或者该柱状图的正规化版本。

[0130] 在步骤1114，网络评分引擎114比较第一NPA得分与在步骤1112生成的NPA得分的分布。作为一个例子，该比较可以包括确定代表推荐的NPA得分与分布之间关系的“p-值”。特别地，p-值可以对应于所述分布高于或低于所推荐NPA得分值的百分比。小的，例如小于0.5％、小于1％、小于5％或者任何其它分数的，p-值指示推荐的NPA得分是统计显著的。例如，在步骤1114计算出的具有低p-值(<0.05或者低于5％，例如)的所推荐NPA得分指示推荐的NPA得分与从随机基因标签重新分配产生的显著数量的测试NPA得分高度相关。

[0131] 在某些实现中，在此被称为“K-统计”测试的另一种置换测试评估因果网络模型中骨干节点结构的重要性。过程1200包括评估所推荐的NPA得分的统计显著性的方法。过程1200与过程1100类似，因为因果网络模型的一方面被随机分类，以创建多个测试模型，在这些模型之上计算多个测试NPA得分。建立于生物系统中实体的因果关系的知识之上的因果

网络模型也被称为未修改的网络。在这种模型当中，边可以带符号，因而边可以代表两个骨干节点之间的正或负关系。因此，因果网络模型包括n条连接导致正影响的骨干节点的边，和m条连接导致负影响的骨干节点的边。

[0132] 在步骤1202，推荐的NPA得分基于建立在生物系统中实体的因果关系的知识之上的网络来计算。然后，在步骤1204中，确定负边的条数n和正边的条数m。在步骤1206，骨干节点对每个都随机地与n条负边中的一条或者m条正边中的一条连接。这个生成与n+m条边的随机连接的过程重复C次。如前面所描述的，迭代次数C可以由用户输入或者由测试NPA得分的分布的平滑性来确定。在步骤1212，基于包括随机连接到其它骨干节点的骨干节点的多个测试模型计算多个测试NPA得分。网络评分引擎114可以根据以上为基于网络计算NPA得分所述的任何方法来计算推荐和测试NPA得分。在步骤1214，推荐的NPA得分与测试NPA得分进行比较或者对照其分级，以确定推荐的NPA得分的统计显著性。

[0133] 在步骤1210，网络评分引擎114基于在步骤1206形成的随机重新连接计算C个NPA得分。在步骤1212，基于从在步骤1106生成的随机重新连接产生的测试模型，网络评分引擎114汇聚结果产生的C个NPA得分，以生成测试NPA值的分布。这种分布可以对应于NPA值的柱状图或者该柱状图的正规化版本。

[0134] 在步骤1214，网络评分引擎114比较推荐的NPA得分与在步骤1212生成的NPA得分的分布。作为一个例子，该比较可以包括确定代表所推荐的NPA得分与所述分布之间关系的“p-值”。特别地，p-值可以对应于该分布高于或低于所推荐NPA得分值的百分比。小的，例如小于0.1％、小于0.5％、小于1％、小于5％或者任何中间分数，的p-值指示推荐的NPA得分是统计显著的。例如，在步骤1214计算出的具有低p-值(<0.05或者低于5％，例如)的推荐NPA得分指示推荐的NPA得分与从骨干节点的随机重新连接产生的显著数量的测试NPA得分高度相关。

[0135] 在某些实现中，可能需要(在图11和12中计算出的)两个p-值对于被认为统计显著的所推荐NPA得分是低的。在其它实现中，为了找出显著的所推荐NPA得分，网络评分引擎114可能需要一个或多个p-值为低。

[0136] 图13是用于识别前导骨干与基因节点的说明性过程1300的流程图。在步骤1302，网络评分引擎114基于识别出的网络模型生成骨干算子。骨干算子作用于支持节点的活性测量向量并且输出用于骨干节点的活性值向量。在有些实现中，合适的骨干算子是在以上等式13中定义的算子K。

[0137] 在步骤1304，网络评分引擎114利用在步骤1302生成的骨干算子生成前导骨干节点列表。前导骨干节点可以代表在治疗与控制数据和因果生物网络模型的分析过程中识别出的最显著的骨干节点。为了生成这个列表，网络评分引擎114可以使用骨干算子形成内核，然后该内核可以在用于骨干节点的活性值的向量与其自己之间的内积中使用。在有些实现中，通过按降序给从这种内积产生的和中的项排序，并且选择对应于对这个和起最大作用者的固定数量的节点或者获得总和的指定百分比(例如，60％)所需的多个最显著起作用节点，网络评分引擎114生成前导骨干节点列表。等价地，通过经计算等式1中有序项的累加和来包括构成NPA得分80％的骨干节点，网络评分引擎114可以生成前导骨干节点列表。如以上所讨论的，这个累加和可以作为以下内积的项的累加和来计算(利用骨干算子K)：

[0138] $$f_1^T K^T K f_1. \qquad (18)$$

[0139] 因而，前导节点的识别既依赖于活性测量，又依赖于网络拓扑结构。

[0140] 在步骤1306，网络评分引擎114利用在步骤1302生成的骨干算子生成前导基因节点列表。如由等式2所示的，NPA得分可以以倍数变化表示为二次形式。因而，在有些实现中，前导基因列表是通过识别以下标量积的有序和的项生成的：

[0141] $$\langle f_1 | L_2 (L_3^{-1})^T L_3^{-1} L_2^T f_1 \rangle. \quad (19)$$

[0142] 前导基因列表的两端可能都是重要的，因为对NPA得分起负作用的基因也具有生物显著性。

[0143] 在有些实现中，网络评分引擎114还在步骤1306为每个基因生成结构重要性值。由于基因在模型中的位置，该结构重要性值独立于实验数据并且代表有些基因可能比其它基因对推断骨干节点值更重要的事实。结构重要性可以通过下式为基因j定义：

[0144] $$I_j = \sum_{i=1}^{N} \left| (L_3^{-1} L_2^T)_{ij} \right| \quad (20)$$

[0145] 前导骨干节点列表中的生物实体和前导基因节点列表中的基因是治疗条件(相对于控制条件)激活底层网络的生物标志的候选。这两个列表可以单独地或者一起用来为将来的研究识别目标，或者可以在其它生物标志识别过程中使用，如以下描述的。

[0146] 现在参考图7，在有些实现中，基于以下拉普拉斯矩阵的内核与图像空间，网络评分引擎114把步骤704中的第一活性向量分别分解成不起作用和起作用的分量：

[0147] $$L_{l^2(V\setminus V_0)} = \left(diag\left(out|_{l^2(V\setminus V_0)}\right) + diag\left(in|_{l^2(V\setminus V_0)}\right) - (A + A^T)\right)|_{l^2(V\setminus V_0)} \in l^2(V\setminus V_0) \quad (21)$$

[0148] 其中，计算网络模型已经被约束到对应于第二组生物实体中生物实体的节点，如以上参考图5的步骤506所讨论的。网络评分引擎114还可以配置为作为等式21拉普拉斯算子的矩阵指数计算“带符号的”扩散内核并且把第一活性值投影到空间分量上，以生成用于将来分析的至少一个起作用成分，如以下所描述的。

[0149] 在步骤706，网络评分引擎114比较(在步骤704确定的)第一起作用向量与从来自不同实验的第二组活性值确定的第二起作用向量。为了确定这个第二起作用向量，步骤702和704可以利用不同的治疗和控制数据对第一组节点重复(根据图5)。在有些实施例中，相同的治疗和/或控制数据可以用于确定第二起作用向量。第二起作用向量代表利用对用于不同实验的NPA得分起作用的不同治疗(和可选地，不同控制数据)从不同实验得到的活性值的分量。由于两个实验中感兴趣的生物系统是相同的，因此底层计算网络模型是相同的并且因此第二不起作用和起作用向量分别依赖于矩阵积$(L_3^{-1}L_2^T)$的内核和矩阵积$(L_3^{-1}L_2^T)$的图像空间。

[0150] 在步骤708，网络评分引擎114基于步骤706的比较提供可比性信息。在有些实现中，可比性信息是第一和第二起作用向量之间的相关性。在有些实现中，可比性信息是第一和第二起作用向量之间的距离。用于比较向量的多种技术中任何一种都可以用于在步骤708提供可比性信息。

[0151] 在图5步骤504计算出的活性测量和在图5步骤506生成的活性值(例如，根据图6的过程600)可以用于提供可译性信息，该可译性信息反映两个不同生物系统对由相同制剂或治疗条件造成的扰动类似地响应的程度。在一个例子中，两个不同的生物系统可以是试管系统、活体系统、小鼠系统、大鼠系统、非人灵长类动物系统及人体系统的任意组合。图8是

用于提供可译性信息的说明性过程800的流程图。例如，在图5的步骤506生成用于第二组节点的活性值之后，过程800可以由网络评分引擎114或者系统100的任何其它合适配置的一个或多个组件执行。在步骤802，网络评分引擎114确定用于第一生物系统中实体的第一组活性值，并且在步骤804，网络评分引擎114确定用于第二生物系统中实体的第二组活性值。第一和第二生物系统中每一个由对应的第一和第二计算网络模型表示。例如，活性值可以根据图5的步骤506或者图6的过程600来确定。

[0152] 在步骤806，网络评分引擎114比较在步骤802确定的第一组活性值与在步骤804确定的第二组活性值。在有些实现中，网络评分引擎114配置为分析用于第一生物系统的第一活性值($V^{(1)}$)与用于第二生物系统的第二活性值($V^{(2)}$)之间的以下关系：

[0153] $$l^2\left(V_0^{(1)}\right) \xrightarrow{h_1} l^2\left(V_0^{(2)}\right)$$

[0154] $$(L_3^{(1)})^{-1}(L_2^{(1)})^T \downarrow \circlearrowright \downarrow (L_3^{(2)})^{-1}(L_2^{(2)})^T$$

[0155] $$l^2\left(V^{(1)}\backslash V_0^{(1)}\right) \xrightarrow{h_2} l^2\left(V^{(2)}\backslash V_0^{(2)}\right), \qquad (22)$$

[0156] 其中h1和h2分别代表在活性测量级别的第一和第二生物系统之间的映射(例如，从用于对第一生物系统的实验的治疗与控制数据到用于对第二生物系统的实验的治疗与控制数据的映射)和在推断出的活性值级别的第一和第二生物系统之间的映射(例如，从用于第一生物系统的推断活性值到用于第二生物系统的推断活性值的映射)。虽然这些映射可能是未知的，但是网络评分引擎114可以配置为通过在活性测量级别和推断的活性值级别执行比较来确定关于这些映射的信息。例如，在有些实现中，网络评分引擎114配置为计算投影到对应矩阵积$(L_3^{(i)})^{-1}(L_2^{(i)})^T$的图像空间中或者投影到相关矩阵(例如以上参考等式21所讨论的拉普拉斯矩阵)的空间分量上的活性值之间的相关性。在有些实现中，网络评分引擎114可以通过应用内核规范相关分析(KCCA)技术来比较第一和第二组活性值，这些技术中许多是本领域中众所周知的。

[0157] 在步骤808，网络评分引擎114基于在步骤806的比较提供可译性信息。如以上参考在图7步骤708提供的可比性信息所讨论的，用于比较向量的多种技术中任何一种都可以用于在步骤808提供可比性信息。例如，在有些实现中，网络评分引擎114配置为计算投影到对应矩阵积$(L_3^{(i)})^{-1}(L_2^{(i)})^T$的图像空间中或者投影到相关矩阵(诸如以上参考等式21所讨论的拉普拉斯矩阵)的空间分量上的活性值之间的相关性。在有些实现中，网络评分引擎114可以比较第一和第二组活性值并且通过应用内核规范相关分析(KCCA)技术来提供可译性信息，这些技术中许多是本领域中众所周知的。

[0158] 图9是用于计算用于活性值和NPA得分的置信间隔的说明性过程900的流程图。在步骤902，网络评分引擎114计算如以上参考图5步骤504所描述的活性测量(在这里表示为β)。在有些实现中，活性测量可以是由Limma R统计分析包或者由另一种标准统计技术确定的倍数变化值或者加权的倍数变化值(例如，利用关联的错误未发现率加权)。在步骤904，网络评分引擎114计算与步骤902中计算出的活性测量(或者加权的活性测量)关联的变差。在有些实现中，在步骤904，矩阵Σ定义为Σ＝diag(var(β))。在步骤906，相关网络的结构

用于生成拉普拉斯矩阵(例如,如以下参考等式9所描述的)。网络可以是加权的、带符号的及有向的,或者其任意组合。在步骤908,网络评分引擎114通过令左手侧等于零来求解等式12的拉普拉斯表达式,以生成f_2(活性值的向量)。在步骤910,网络评分引擎114计算活性值向量的变差。在有些实现中,这个向量是根据下式计算的:

[0159] $$var(f_2) = L_3 L_2^T \sum L_2 L_3^T, \quad (23)$$

[0160] 其中,L_2和L_3如在等式11中定义的。在步骤912中,网络评分引擎114根据下式计算f_2的每个输入的置信间隔:

[0161] $$f_2(x) \pm z(1-\frac{\alpha}{2})\sqrt{var(f_2(x))}, \quad (24)$$

[0162] 其中 $z(1-\frac{\alpha}{2})$ 是关联的N(0,1)分位数(例如,如果a=0.05,则是1.96)。在步骤914,网络评分引擎114计算要在步骤916中用于计算NPA得分的二次形式矩阵。在有些实现中,二次形式矩阵是根据上面的等式3计算的。在步骤916,网络评分引擎114根据等式2利用二次形式矩阵Q计算NPA得分。在步骤918,网络评分引擎114计算在步骤916计算出的NPA得分的变差。在有些实现中,这个变差是根据下式计算的:

[0163] $$var(NPA) = var(f_2^T Q f_2) = 2tr(Q\sum^2 Q\sum^2) + 4f_2^T Q\sum^2 Q f_2, \quad (25)$$

[0164] 其中,Σ^2=var(f_2)。在步骤920,网络评分引擎114为在步骤916计算出的NPA得分计算置信间隔。在有些实现中,置信间隔是根据下式计算的:

[0165] $$NPA \pm \sqrt{(\frac{1}{1-\alpha})}\sqrt{var(NPA)}. \quad (26)$$

[0166] 或者

[0167] $$NPA \pm z(1-\frac{\alpha}{2})\sqrt{var(NPA)}. \quad (27)$$

[0168] 图14是用于量化生物扰动的影响的分布式计算机化系统1400的框图。系统1400的组件类似于图1系统100中的那些组件,但是系统100的布置使得每个组件通过网络接口1410通信。这种实现对于经多个通信系统——包括可以共享对公共网络资源的访问的无线通信系统,诸如“云计算”范例——的分布式计算可能是合适的。

[0169] 图15是计算设备的框图,诸如用于执行在此所述的过程的图1系统100或图11系统1100的任何组件。系统100的每个组件,包括系统响应剖面引擎110、网络建模引擎112、网络评分引擎114、汇聚引擎116及一个或多个数据库,可以在一个或多个计算设备1500上实现,其中的数据库包括结果数据库、扰动数据库和文献数据库。在某些方面,以上的组件和数据库中的多个可以包括在一个计算设备1500中。在某些实现中,组件和数据库可以跨几个计算设备1500实现。

[0170] 计算设备1500包括至少一个通信接口单元、输入/输出控制器1510、系统存储器及一个或多个数据存储设备。系统存储器包括至少一个随机存取存储器(RAM 1502)和至少一个只读存储器(ROM1504)。所有这些元件都与中央处理单元(CPU 1506)通信,以方便计算设备1500的操作。计算设备1500可以许多不同的途径配置。例如,计算设备1500可以是常规的独立计算机或者,作为替代,计算设备1500的功能可以跨多个计算机系统和体系结构分布。计算设备1500可以配置为执行建模、评分和汇聚操作中的一些或全部。在图15中,计算设备1500经网络或局部网络链接到其它服务器或系统。

[0171] 计算设备1500可以在分布式体系架构中配置，其中数据库和处理器放置在分开的单元或位置中。有些这种单元执行主要的处理功能并且至少包括一个通用控制器或处理器及一个系统存储器。在这一方面，这些单元中每一个都经通信接口单元1508附连到充当与其它服务器、客户端或用户计算机和其它相关设备的主要通信链路的通信集线器或端口(未示出)。通信集线器或端口自己可以具有最小化的处理能力，主要充当通信路由器。多种通信协议可以是系统的一部分，包括但不限于：以太网、SAP、SAS™、ATP、BLUETOOTH™、GSM和TCP/IP。

[0172] CPU 1506包括处理器，诸如一个或多个常规的微处理器，以及一个或多个补充的协处理器，诸如用于从CPU 1506卸载工作量的数学协处理器。CPU 1506与通信接口单元1508和输入/输出控制器1510通信，CPU 1506可以通过通信接口单元1508和输入/输出控制器1510与诸如其它服务器、用户终端或设备的其它设备通信。通信接口单元1508和输入/输出控制器1510可以包括用于与例如其它处理器、服务器或客户终端同时通信的多个通信通道。彼此通信的设备不需要持续地向彼此发送。相反，这种设备只需根据需要向彼此发送、实际上可以大部分时间避免交换数据，而且可能需要执行几个步骤来确立设备之间的通信链路。

[0173] CPU 1506还与数据存储设备通信。数据存储设备可以包括磁、光或半导体存储器的适当组合，而且可以包括例如RAM 1502、ROM 1504、闪存驱动器、诸如致密盘(compact disc)的光盘(optical disc)或者硬盘或驱动器。CPU 1506和数据存储设备每个都可以例如完全位于单个计算机或其它计算设备中；或者通过通信介质彼此连接，其中的通信介质诸如UBS端口、串口电缆、同轴电缆、以太网类型电缆、电话线、射频收发器或者其它类似的无线或有线介质或者以上所述的组合。例如，CPU 1506可以经通信接口单元1508连接到数据存储设备。CPU 1506可以配置为执行一个或多个特定的处理功能。

[0174] 数据存储设备可以存储例如(i)用于计算设备1500的操作系统1512；(ii)适于根据在此所述的系统和方法，尤其是根据关于CPU1506具体描述的过程，指引CPU 1506的一个或多个应用1514(例如，计算机程序代码或计算机程序产品)；或者(iii)适于存储程序所需信息的数据库1516。在有些方面，数据库包括存储实验数据及已发表的文献模型的数据库。

[0175] 操作系统1512和应用1514可以例如以压缩、未压缩和加密的格式存储，并且可以包括计算机程序代码。程序的指令可以从除数据存储设备之外的计算机可读介质，诸如从ROM 1504或者从RAM1502，读到处理器的主存储器中。在程序中指令序列的执行使CPU1506执行在此所述的过程步骤的同时，硬连线的电路系统可以代替，或者与软件指令结合来实现本公开内容的过程。因而，所述系统与方法不限于硬件与软件的任何具体组合。

[0176] 可以提供合适的计算机程序代码，用于执行与在此所述的建模、评分和汇聚相关的一个或多个函数。程序还可以包括诸如操作系统1512、数据库管理系统和“设备驱动器”的程序元素，其中“设备驱动器”允许处理器经输入/输出控制器1510与计算机外围设备(例如，视频显示器、键盘、计算机鼠标等)接口。

[0177] 如在此所使用的，术语“计算机可读介质”指向计算设备1500的处理器(或者在此所述的设备的任何其它处理器)提供或参与提供要执行的指令的任何非临时性介质。这种介质可以采取许多形式，包括但不限于非易失性介质和易失性介质。非易失性介质包括，例如，光、磁或光-磁盘，或者集成电路存储器，诸如闪存存储器。易失性介质包括一般构成主

存储器的动态随机存取存储器(DRAM)。计算机可读介质的常见形式包括，例如，软盘、软磁盘、硬盘、磁带、任何其它磁性介质、CD-ROM、DVD、任何其它光学介质、穿孔卡片、纸带、任何其它具有孔模式的物理介质、RAM、PROM、EPROM或EEPROM(电可擦除可编程只读存储器)、闪存EEPROM、任何其它存储器芯片或盒式磁带、或者计算机可以从其读的任何其它非临时性介质。

[0178] 各种形式的计算机可读介质可以用于把一条或多条指令的一个或多个序列携带到CPU 1506(或者在此所述的设备的任何其它处理器)以供执行。例如，指令可以最初在远程计算机(未示出)的磁盘上产生。该远程计算机可以把指令加载到其动态存储器中并且经以太网连接、电缆线或者甚至利用调制解调器经电话线发送指令。计算设备1500(例如，服务器)本地的通信设备可以在对应的通信线路上接收数据并且把数据放到用于处理器的系统总线上。系统总线把数据携带到主存储器，处理器从主存储器接收并执行指令。在被处理器执行之前或之后，由主存储器接收的指令可以可选地存储在存储器中。此外，指令可以作为电、电磁或光信号经通信端口接收，这些是可以携带各种类型信息的无线通信或数据流的示例性形式。

[0179] 虽然本公开内容的实现已经参考具体的例子特别示出并进行了描述，但是本领域技术人员应当理解，在不背离由所附权利要求定义的本公开内容范围的情况下可以对其形式与细节进行各种改变。因而，本公开内容的范围是由所附权利要求指示的，而且因此在权利要求等价意义与范围内的所有变化都要被涵盖。

[0180] 在此所述的系统与方法已经利用能很好理解的细胞培养实验进行了测试。正常的人体支气管上皮(NHBE)细胞通过暴露给PD-0332991、一种吸引G1中细胞的CDK4/6抑制剂(CDKI)，来治疗。然后，通过从介质除去CDKI并清洗，允许处理过的细胞重新进入细胞循环。通过在CDKI被除去并且细胞被清洗之后2、4、6和8小时以S-相位荧光标记细胞，实验上确认细胞循环的重新进入。获得在CDKI除去之后2、4、6和8小时采样的细胞的基因转录剖面。获得介质中持续地暴露给CDKI的细胞的剖面。为了识别在CDKI被除去时有区别地被激活的生物过程和机制，网络扰动量值得分利用在各个时间点所获得的清洗后的细胞的基因转录剖面来计算。对于用于与CDKI的去除关联的扰动的NPA得分的计算，使用包括127个节点和240条边的细胞循环子网络。这是在Schlage等人发表的细胞增殖网络模型(2011，“A computable cellular stress network model for non-diseased pulmonary and cardiovascular tissue”，BMC Syst Biol.Oct 19;5:158，在此引入其全部作为参考)的一个子网络。

[0181] 发现NPA得分(图18)在从2小时时间点到8小时时间点的时间点范围上增加，这与荧光激活的细胞分类(FACS)分析(图17)的结果一致，其中FACS分析示出了S-相位中细胞数量的对应增加。NPA得分在P-值<0.05接受两个置换测试，如以上所描述的，而且统计(“O”和“K”统计)都指示实验的NHBE细胞中这个特定的生物系统，即，细胞循环，实际上被扰动了。该分析还识别出细胞循环网络模型中的前导节点，该前导节点精确地对应于已知在S-相位的输入中所涉及的密钥机制:E2F蛋白质与RbP构成联合体，而RbP又在p53和CHEK1的(间接)控制下被Cdk磷酸化。而且还结合Cdk，G1/S-细胞周期蛋白是前导节点过程的一部分，如所预期的。由该方法识别出的前导节点是:taof(TFDP1)、taof(E2F2)、CHEK1、TFDP1、kaof(CHEK1)、taof(E2F3)、taof(E2F1)、taof(RB1)、有丝分裂细胞周期的G1/S过渡、CDC2、E2F2、

CCNA2、CCNE1、THAP1、CDKN1A、TP53P@S20、E2F3、kaof(CDK2)。Taof是“...的转录活性”的缩写，而kaof是“…的激酶活性”的缩写。TP53P@S20是TP53中位置20的丝氨酸被磷酸化的缩写。结果显示基因表达数据与充分利用包含在因果网络模型中的生物系统的知识的机制驱动方法的组合可以用于量化生物系统的扰动。

[0182] 本发明进一步在以下带标号的段落中进行了定义：

[0183] 一种用于量化生物系统的扰动的计算机化方法，包括：

[0184] 在第一处理器接收与第一组生物实体对第一治疗的响应对应的第一组治疗数据，其中第一生物系统包括生物实体，所述生物实体包括第一组生物实体和第二组生物实体，第一生物系统中的每个生物实体与第一生物系统中的至少一个其它生物实体相互作用；

[0185] 在第二处理器接收与第一组生物实体对与第一治疗不同的第二治疗的响应对应的第二组治疗数据；

[0186] 在第三处理器提供第一计算因果网络模型，所述第一计算因果网络模型代表第一生物系统并且包括：

[0187] 代表第一组生物实体的第一组节点，

[0188] 代表第二组生物实体的第二组节点，

[0189] 连接节点并且代表生物实体之间的关系的边，及

[0190] 方向值，代表第一治疗数据与第二治疗数据之间预期的变化方向；

[0191] 利用第四处理器为第一组节点中的对应节点计算代表第一治疗数据与第二治疗数据之间的差异的第一组活性测量；及

[0192] 基于第一计算因果网络模型和第一组活性测量，利用第五处理器为第二组节点中的对应节点生成第二组活性值。

[0193] 段落137的方法，还包括：

[0194] 基于第一计算因果网络模型和第二组活性值，利用第六处理器为第一计算因果网络模型生成代表由第一和第二治疗造成的第一生物系统的扰动的得分。

[0195] 段落137的方法，其中生成第二组活性值包括：为第二组节点中的每个特定节点识别最小化差异声明的活性值，所述差异声明表示特定节点的活性值与该特定节点利用第一计算因果网络模型中的边连接到的节点的活性值或活性测量之间的差异，其中差异声明依赖于第二组节点中每个节点的活性值。

[0196] 段落139的方法，其中差异声明还依赖于第二组节点中每个节点的方向值。

[0197] 段落137的方法，其中第二组活性值中的每个活性值是第一组活性测量的活性测量的线性组合。

[0198] 段落141的方法，其中线性组合依赖于第一计算因果网络模型中第一组节点中的节点与第二组节点中的节点之间的边，而且还依赖于第一计算因果网络模型中第二组节点中节点之间的边。

[0199] 段落141的方法，其中线性组合不依赖于第一计算因果网络模型中第一组节点中节点之间的边。

[0200] 段落138的方法，其中得分对第二组活性值具有二次依赖性。

[0201] 段落137的方法，还包括：通过为第一组活性测量的每个活性测量形成变差估计的线性组合，为第二组活性值中的每个活性值提供变差估计。

[0202] 段落138的方法，其中用于得分的变差估计对第二组活性值具有二次依赖性。

[0203] 段落138的方法，还包括：

[0204] 把第二组活性值表示为第一活性值向量；

[0205] 把第一活性值向量分解成第一起作用向量和第一不起作用向量，使得第一起作用和不起作用向量之和是第一活性值向量。

[0206] 段落147的方法，其中得分不依赖于第一不起作用向量。

[0207] 段落148的方法，其中得分是作为第二组活性值的二次函数计算的，而且第一不起作用向量是该二次函数的内核。

[0208] 段落147的方法，其中，第一不起作用向量在与第一计算因果网络模型关联的带符号拉普拉斯算子的二次函数的内核中。

[0209] 段落147的方法，还包括：

[0210] 在第一处理器接收与第一组生物实体对第三治疗的响应对应的第三组治疗数据；

[0211] 在第二处理器接收与第一组生物实体对第四治疗的响应对应的第四组治疗数据；

[0212] 利用第四处理器计算对应于第一组节点的第三组活性测量，第三组活性测量中的每个活性测量代表用于第一组节点中的对应节点的第三组治疗数据与第四组治疗数据之间的差异；

[0213] 基于第一计算因果网络模型和第三组活性测量，利用第五处理器生成第四组活性值，每个活性值代表用于第二组节点中的对应节点的活性值；

[0214] 把第四组活性值表示为第二活性值向量；

[0215] 把第二活性值向量分解成第二起作用向量和第二不起作用向量，使得第二起作用和不起作用向量之和是第二活性值向量；及

[0216] 比较第一和第二起作用向量。

[0217] 段落151的方法，其中比较第一和第二起作用向量包括：计算第一和第二起作用向量之间的相关性以指示第一组治疗数据和第三组治疗数据的可比性。

[0218] 段落151的方法，其中比较第一和第二起作用向量包括：把第一和第二起作用向量投影到计算网络模型的带符号拉普拉斯算子的图像空间上。

[0219] 段落151的方法，其中第二组治疗数据包含与第四组治疗数据相同的信息。

[0220] 段落137的方法，还包括：

[0221] 在第一处理器接收与第三组生物实体对与第一治疗不同的第三治疗的响应对应的第三组治疗数据，其中第二生物系统包括多个生物实体，所述生物实体包括第三组生物实体和第四组生物实体，第二生物系统中的每个生物实体都与第二生物系统中的至少一个其它生物实体相互作用；

[0222] 在第二处理器接收与第三组生物实体对与第三治疗不同的第四治疗的响应对应的第四组治疗数据；

[0223] 在第三处理器提供第二计算因果网络模型，所述第二计算因果网络模型代表第二生物系统并且包括：

[0224] 代表第三组生物实体的第三组节点，

[0225] 代表第四组生物实体的第四组节点，

[0226] 连接节点并且代表生物实体之间的关系的边，及

[0227] 方向值，代表第三治疗数据与第四治疗数据之间预期的变化方向；

[0228] 利用第四处理器计算对应于第三组节点的第三组活性测量，第三组活性测量中的每个活性测量代表用于第三组节点中的对应节点的第三组治疗数据与第四组治疗数据之间的差异；

[0229] 基于第二计算因果网络模型和第三组活性测量，利用第五处理器生成第四组活性值，每个活性值代表用于第四组节点中的对应节点的活性值；及

[0230] 比较第四组活性值与第二组活性值。

[0231] 段落155的方法，其中比较第四组活性值与第二组活性值包括：应用基于与第一计算因果网络模型关联的带符号拉普拉斯算子和与第二计算因果网络模型关联的带符号拉普拉斯算子的内核规范相关性分析。

[0232] 以上段落137-156任何一段的计算机化方法，其中活性测量是倍数变化值，且用于每个节点的倍数变化值包括用于由对应节点表示的生物实体的对应各组治疗数据之间的差异的对数。

[0233] 以上段落137-157任何一段的计算机化方法，其中生物系统包括细胞增殖机制、细胞应激机制、细胞发炎机制和DNA修复机制中的至少一个。

[0234] 以上段落137-158任何一段的计算机化方法，其中第一治疗包括暴露给通过加热烟草生成的浮质、暴露给通过燃烧烟草生成的浮质、暴露给烟草烟雾及暴露给香烟烟雾中的至少一个。

[0235] 以上段落137-159任何一段的计算机化方法，其中第一治疗包括暴露给异类物质(heterogeneous substance)，包括生物系统中不存在或者不能从其得到的分子或实体。

[0236] 以上段落137-160任何一段的计算机化方法，其中第一治疗包括暴露给毒素、治疗用化合物、刺激物、弛缓剂、天然产物、制造产物及食品物质。

[0237] 段落155和156任何一段的计算机化方法，其中第一生物系统和第二生物系统是包括试管系统、活体系统、小鼠系统、大鼠系统、非人灵长类系统和人体系统的组中的两个不同元素。

[0238] 段落137的计算机化方法，其中：

[0239] 第一治疗数据对应于暴露给制剂的第一生物系统；及

[0240] 第二治疗数据对应于不暴露给制剂的第一生物系统。

[0241] 段落138的计算机化方法，还包括确定得分的统计显著性，所述统计显著性指示生物系统的扰动。

[0242] 段落164的计算机化方法，其中得分的统计显著性是通过比较该得分与多个测试得分来确定的，其中每个测试得分是从多个随机生成的测试计算因果网络模型计算出来的。

[0243] 段落165的计算机化方法，其中随机生成的测试计算因果网络模型是通过随机分类第一计算因果网络模型的一个或多个方面生成的。

[0244] 段落166的计算机化方法，其中第一计算因果网络模型的一个或多个方面包括第一组节点的标签、把第二组节点连接到第一组节点的边或者把第二组节点彼此连接的边。

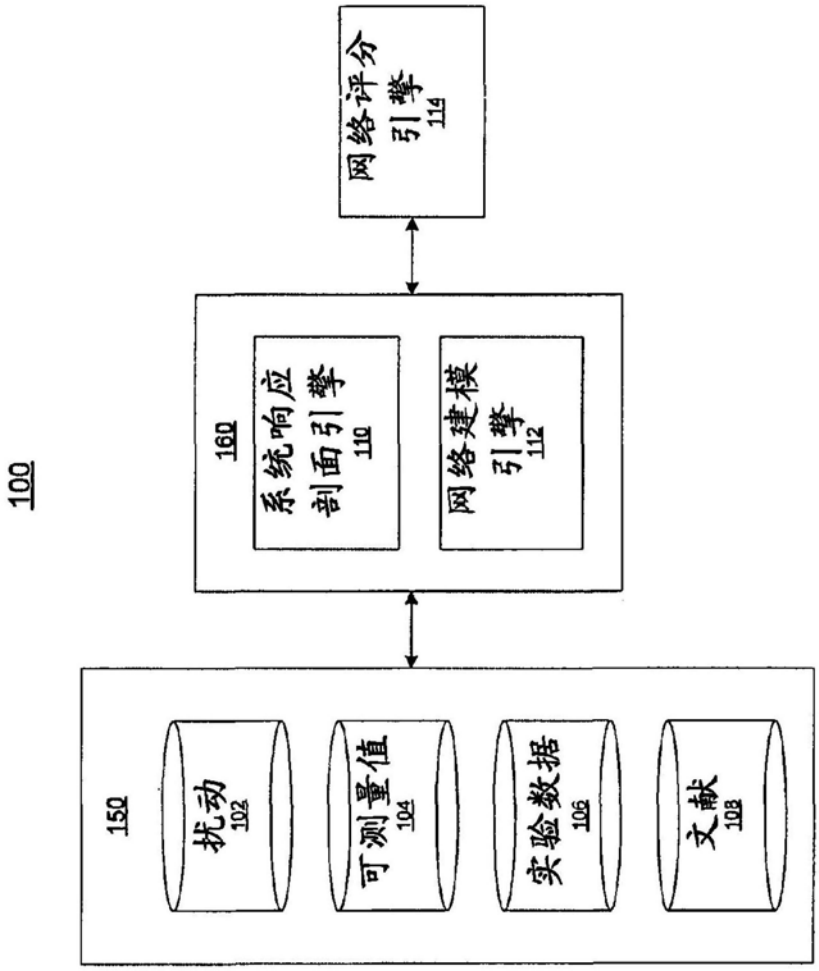

100
网络评分引擎 114
160
系统响应剖面引擎 110
网络建模引擎 112
150
扰动 102
可测量值 104
实验数据 106
文献 108

图1

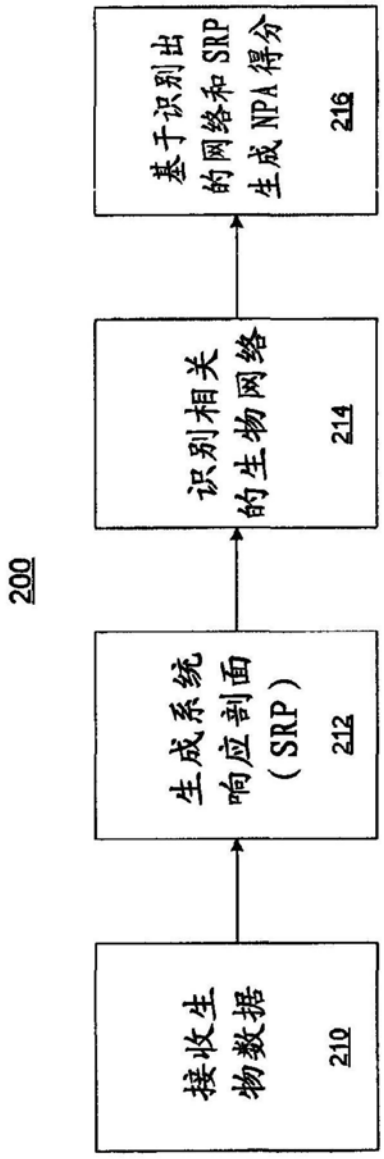

200
接收生物数据
210
生成系统响应剖面（SRP）
212
识别相关的生物网络
214
基于识别出的网络和SRP生成NPA得分
216

图2

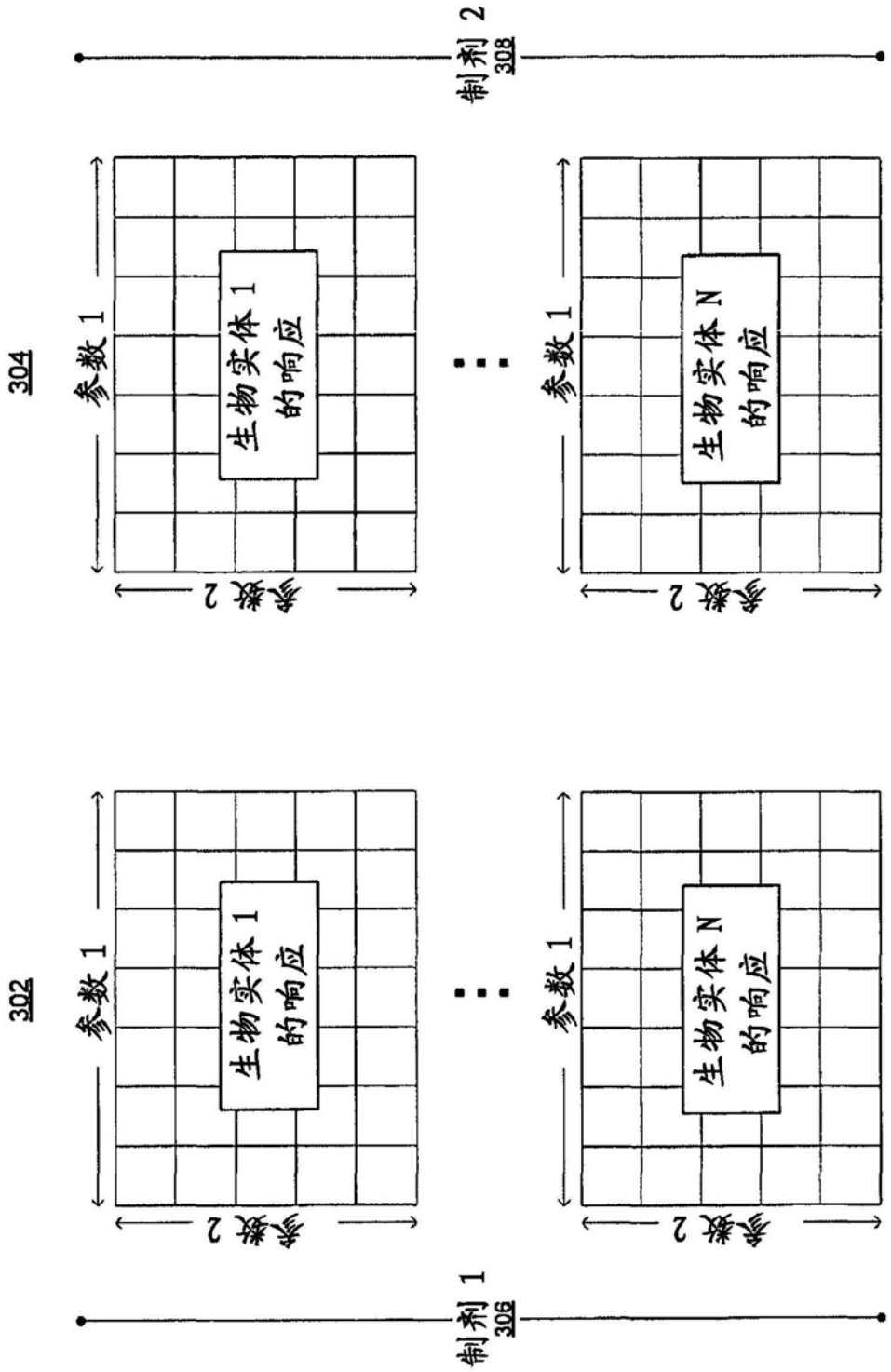


图3

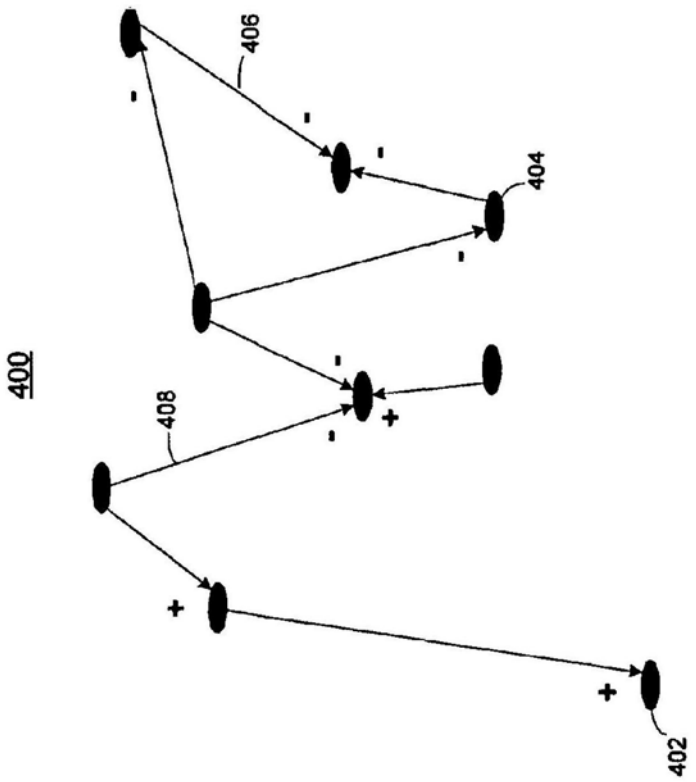


图4

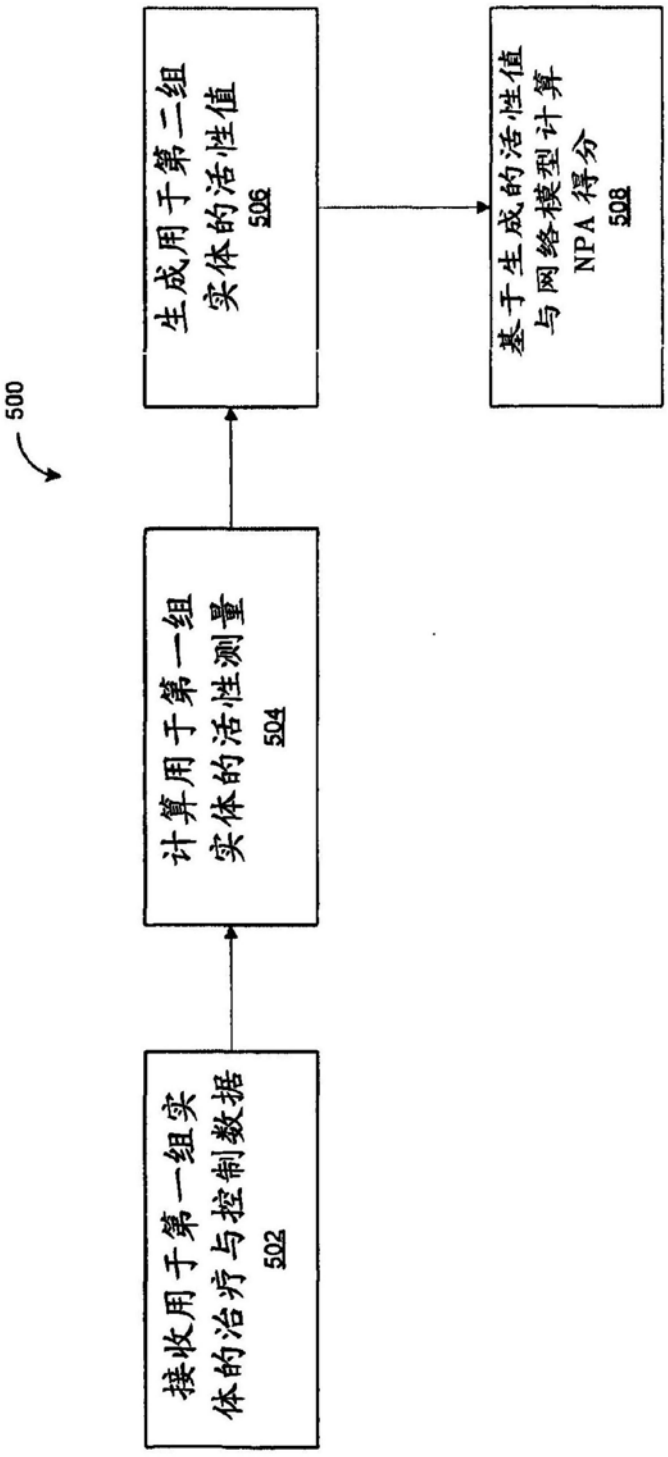

500
接收用于第一组实体的治疗与控制数据 502
计算用于第一组实体的活性测量 504
生成用于第二组实体的活性值 506
基于生成的活性值与网络-模型计算NPA得分 508

图5

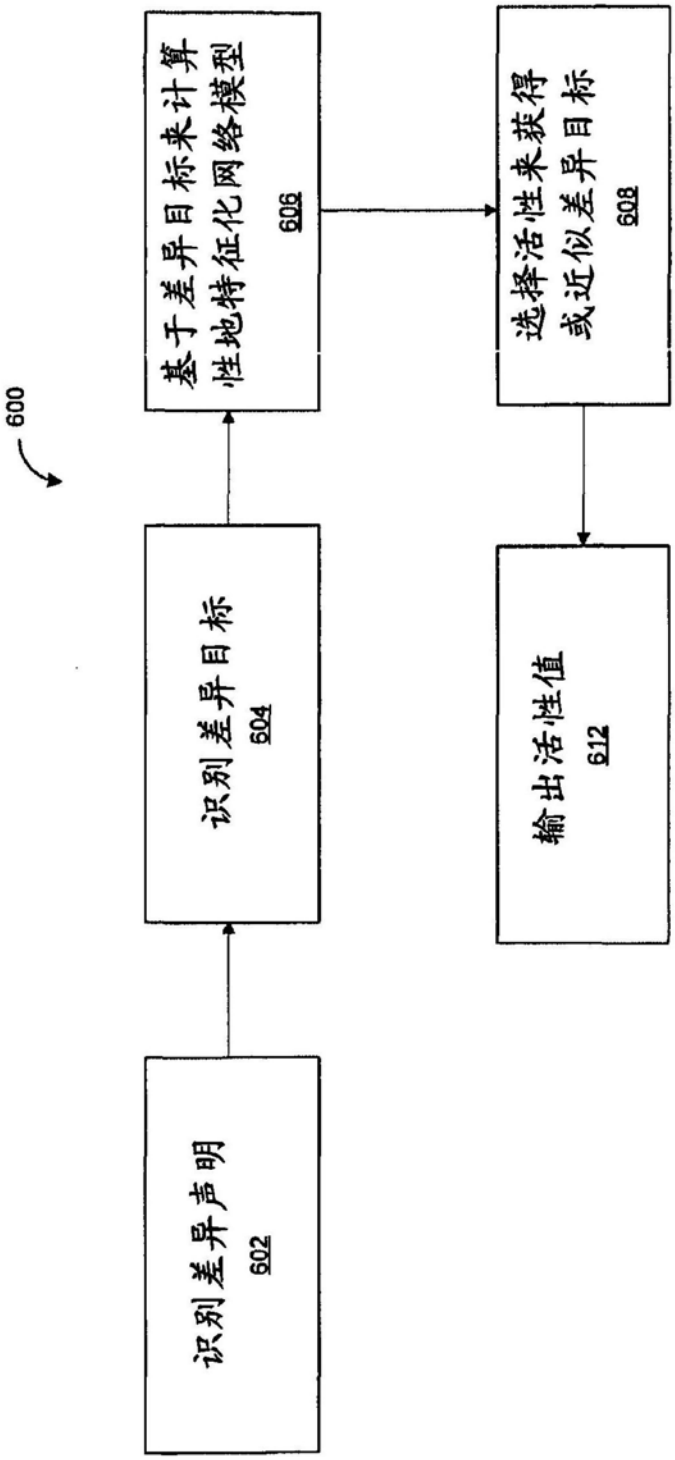

600
识别差异声明
602
识别差异目标
604
基于差异目标来计算性地特征化网络模型
606
选择活性来获得或近似差异目标
608
输出活性值
612

图6

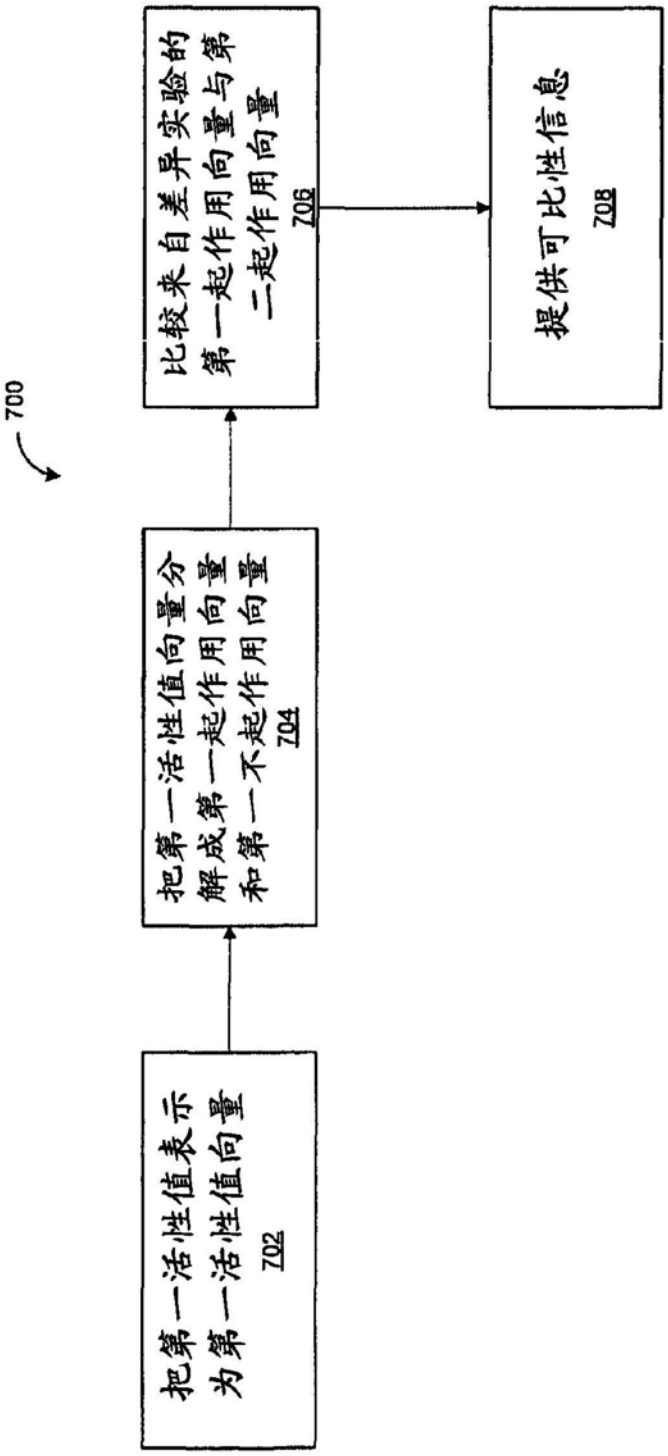

700
把第一活性值表示为第一活性值向量 702
把第一活性值向量分解成第一起作用向量和第一不起作用向量 704
比较来自差异实验的第一起作用向量与第二起作用向量 706
提供可比性信息 708

图7

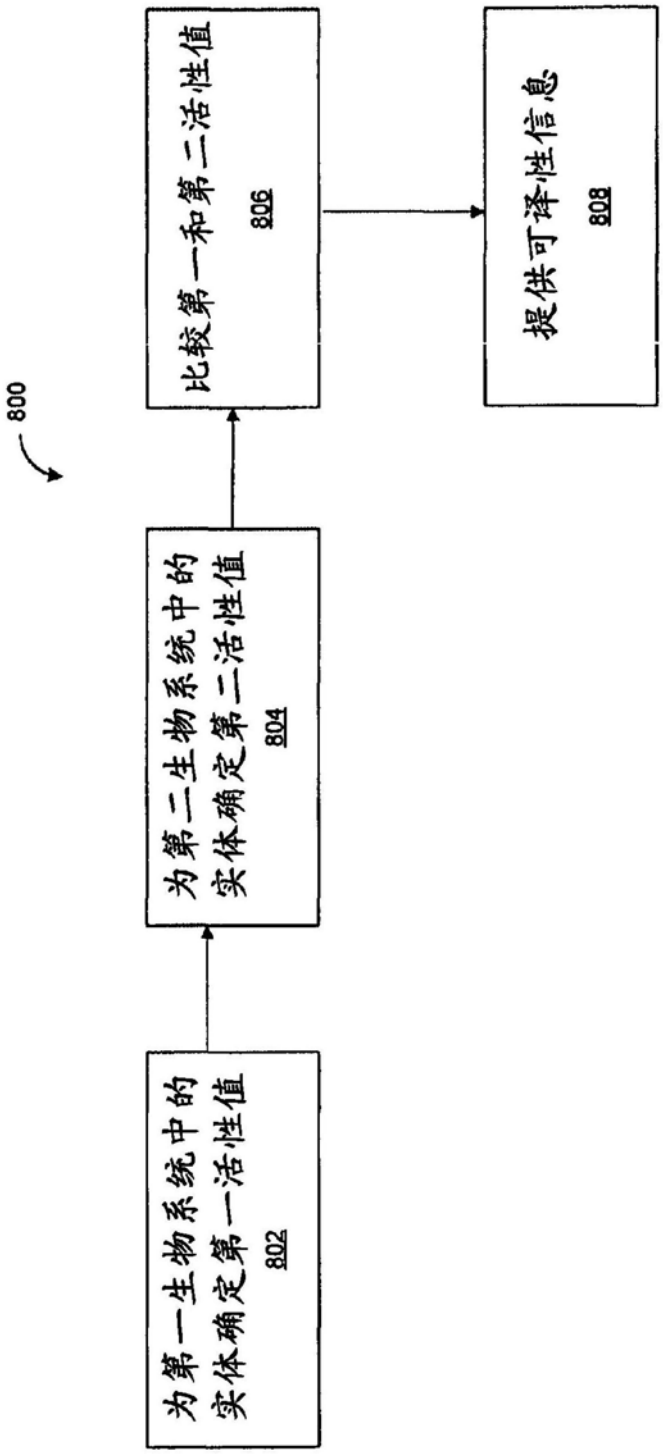

800
为第一生物系统中的实体确定第一活性值
802
为第二生物系统中的实体确定第二活性值
804
比较第一和第二活性值
806
提供可译性信息
808

图8

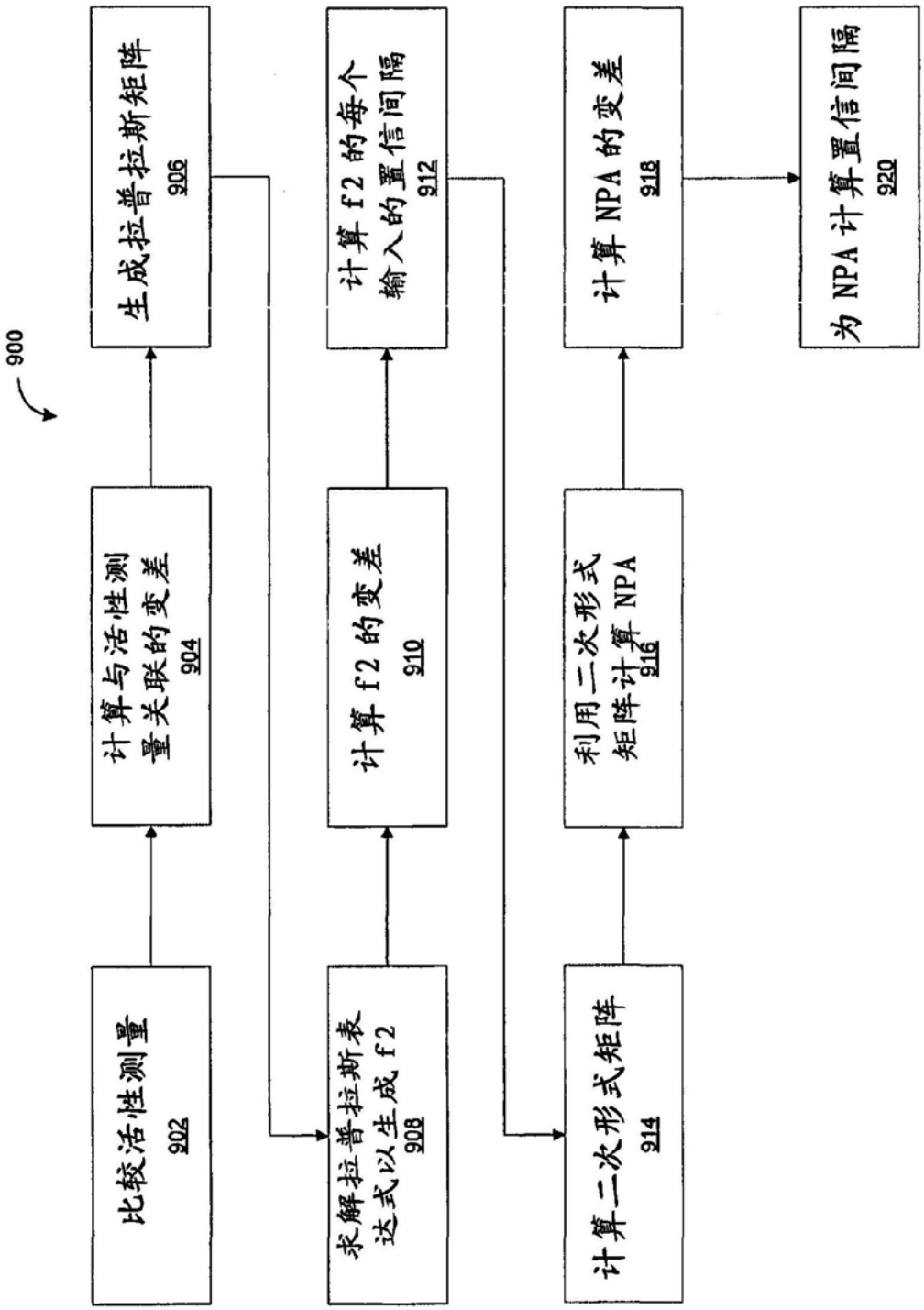

900
比较活性测量
902
计算与活性测量关联的变差
904
生成拉普拉斯矩阵
906
求解拉普拉斯表达式以生成 f2
908
计算 f2 的变差
910
计算 f2 的每个输入的置信间隔
912
计算二次形式矩阵
914
利用二次形式矩阵计算 NPA
916
计算 NPA 的变差
918
为 NPA 计算置信间隔
920

图9

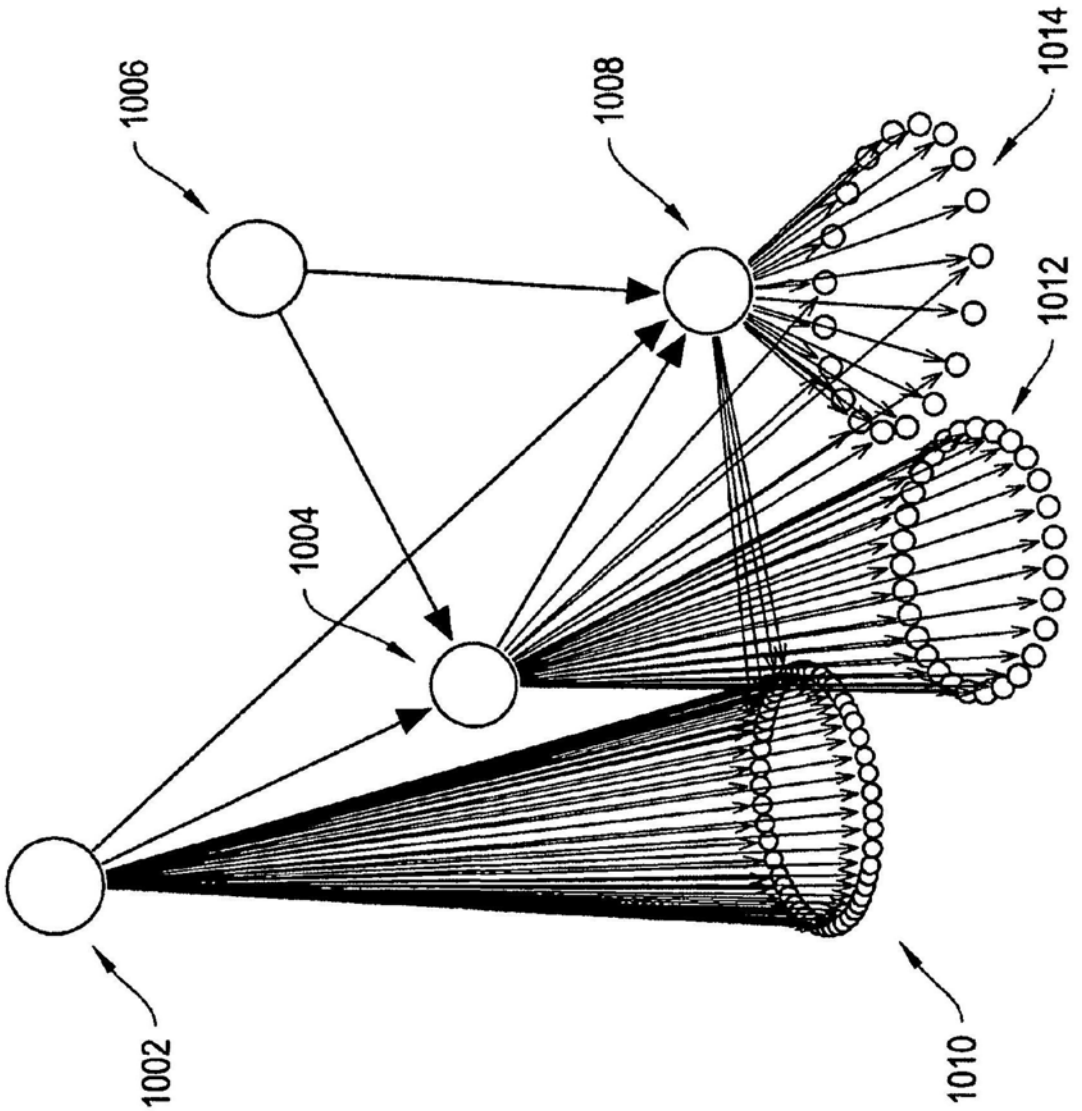


图10

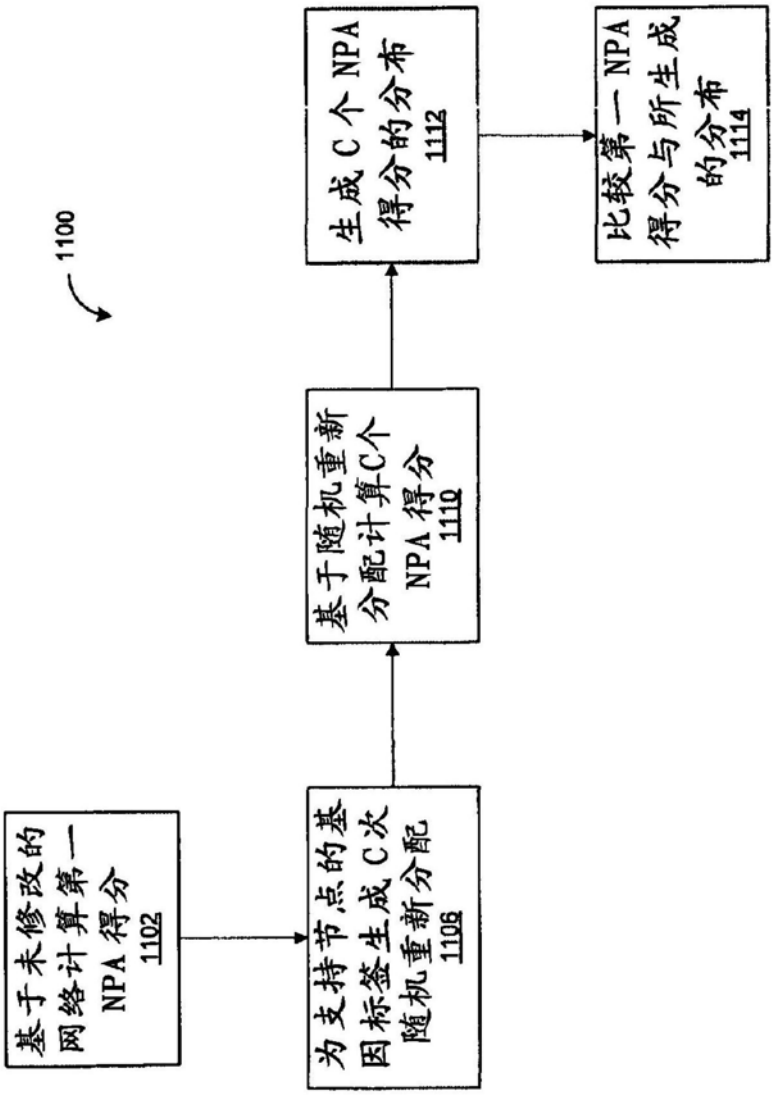

1100
基于未修改的网络计算第一NPA得分 1102
为支持节点的基因标签生成C次随机重新分配 1106
基于随机重新分配计算C个NPA得分 1110
生成C个NPA得分的分布 1112
比较第一NPA得分与所生成的分布 1114

图11

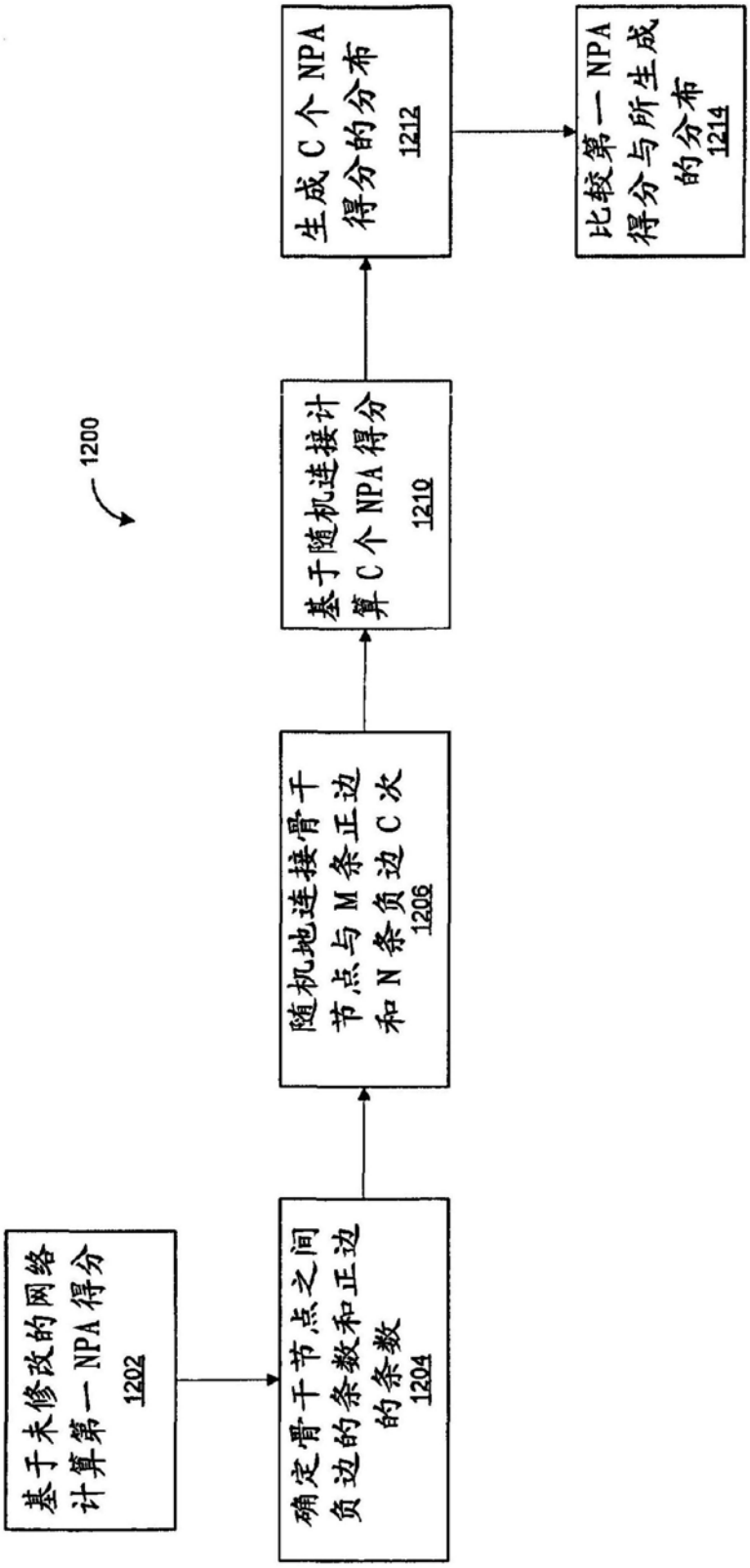

1200
基于未修改的网络计算第一 NPA 得分
1202
确定骨干节点之间负边的条数和正边的条数
1204
随机地连接骨干节点与 M 条正边和 N 条负边 C 次
1206
基于随机连接计算 C 个 NPA 得分
1210
生成 C 个 NPA 得分的分布
1212
比较第一 NPA 得分与所生成的分布
1214

图12

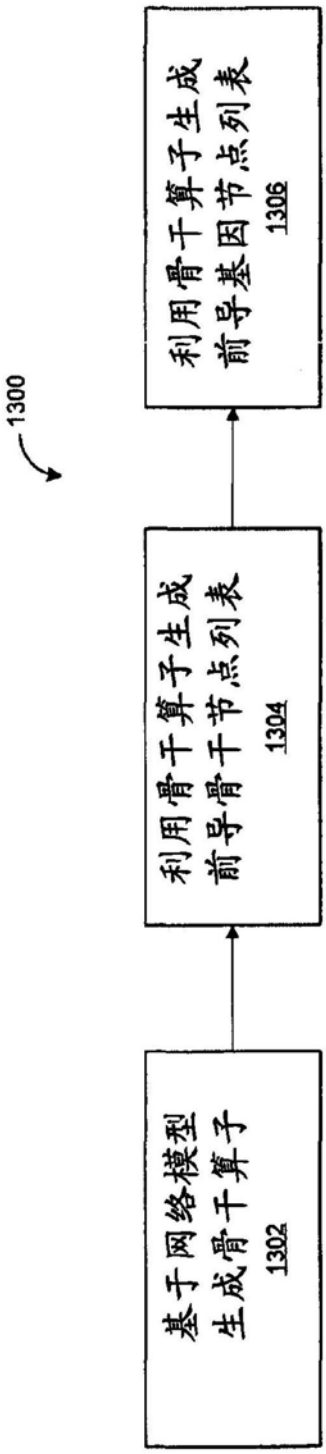

1300
基于网络模型生成骨干算子
1302
利用骨干算子生成前导骨干节点列表
1304
利用骨干算子生成前导基因节点列表
1306

图13

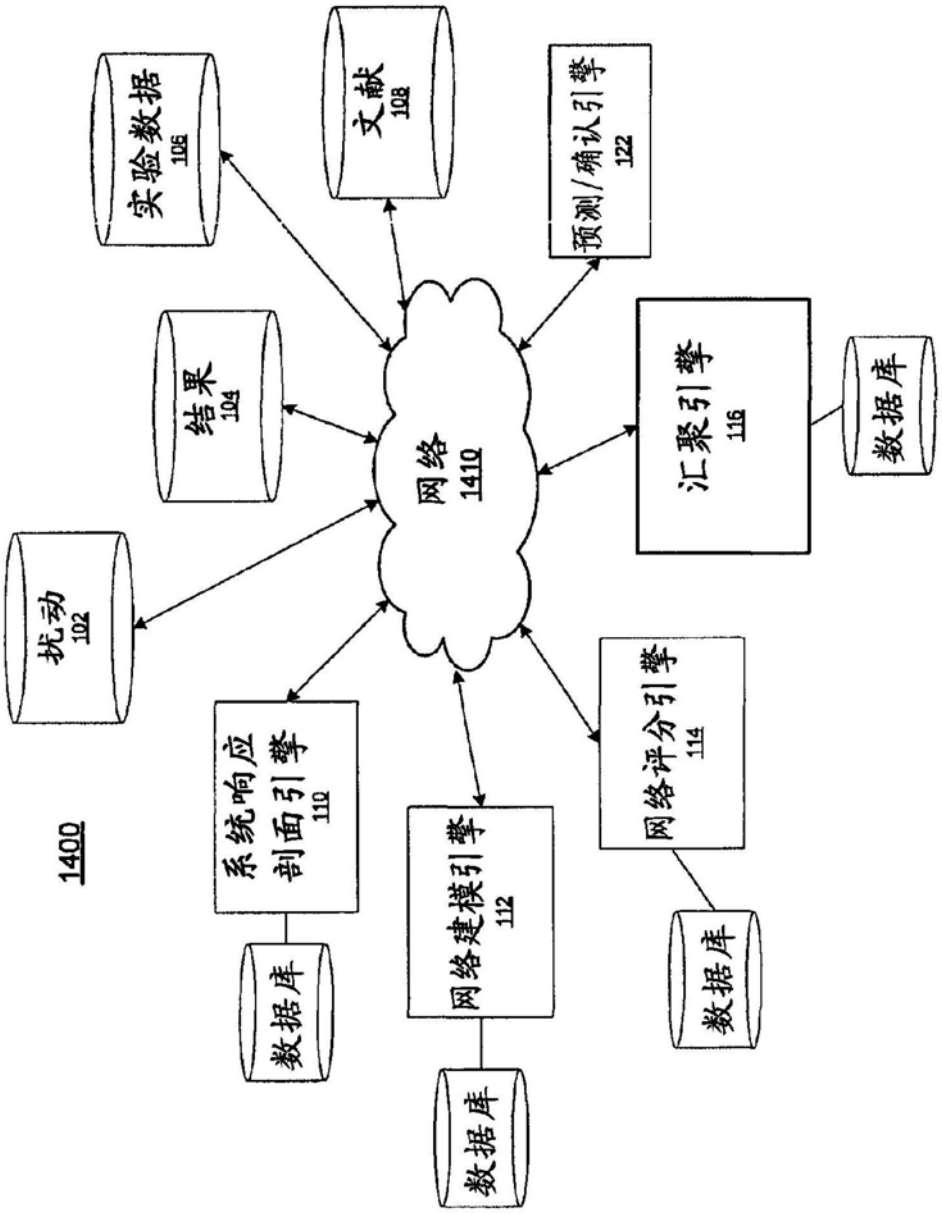

实验数据
106
文献
108
预测/确认引擎
122
结果
104
网络
1410
汇聚引擎
116
数据库
扰动
102
系统响应剖面引擎
110
网络评分引擎
114
网络建模引擎
112
数据库
数据库
数据库
1400

图14

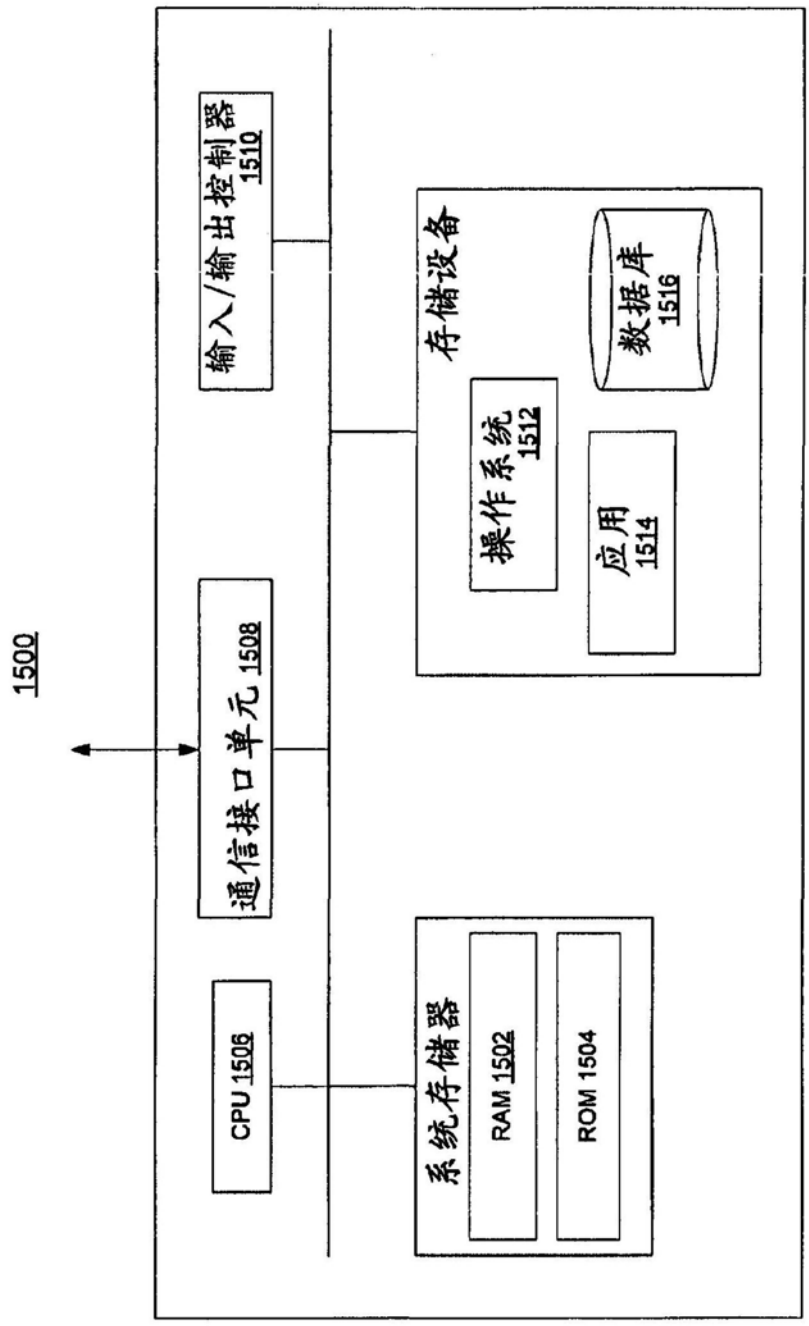

1500
输入/输出控制器 1510
通信接口单元 1508
CPU 1506
系统存储器
RAM 1502
ROM 1504
存储设备
操作系统 1512
应用 1514
数据库 1516

图15

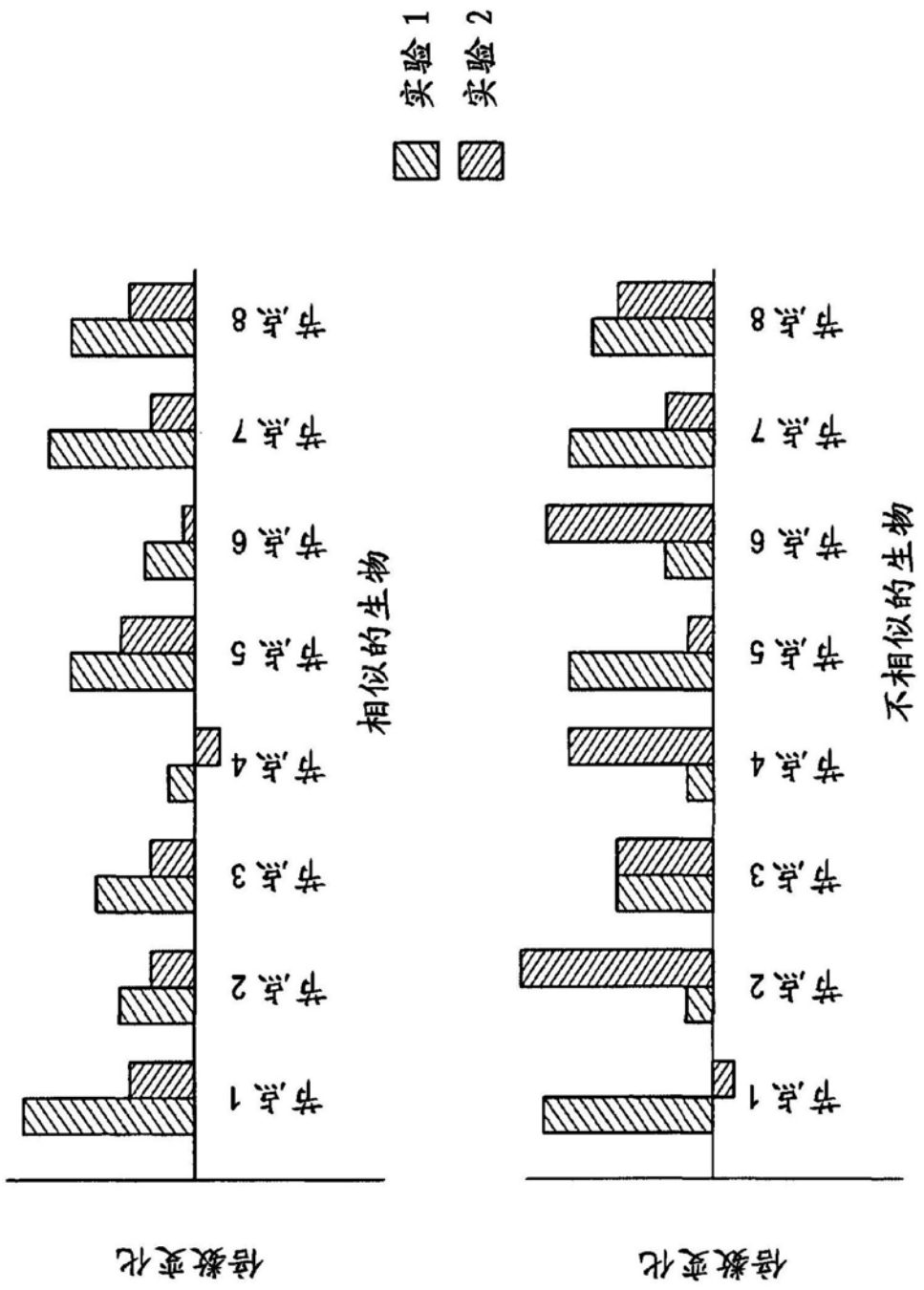


图16

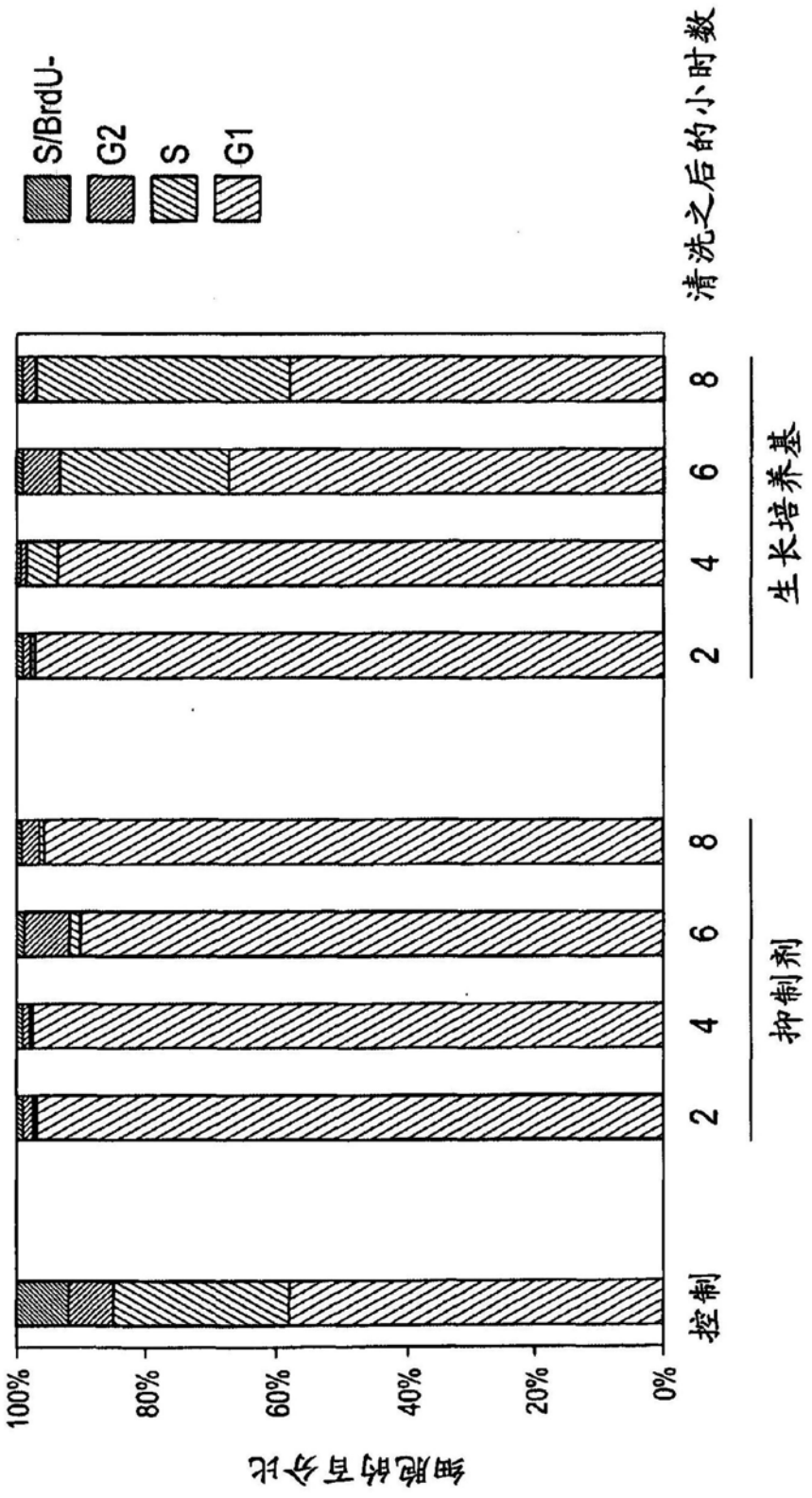

S/BrdU-
G2
S
G1
清洗之后的小时数
生长培养基
抑制剂
控制
2
4
6
8
2
4
6
8
100%
80%
60%
40%
20%
0%
细胞的百分比

图17

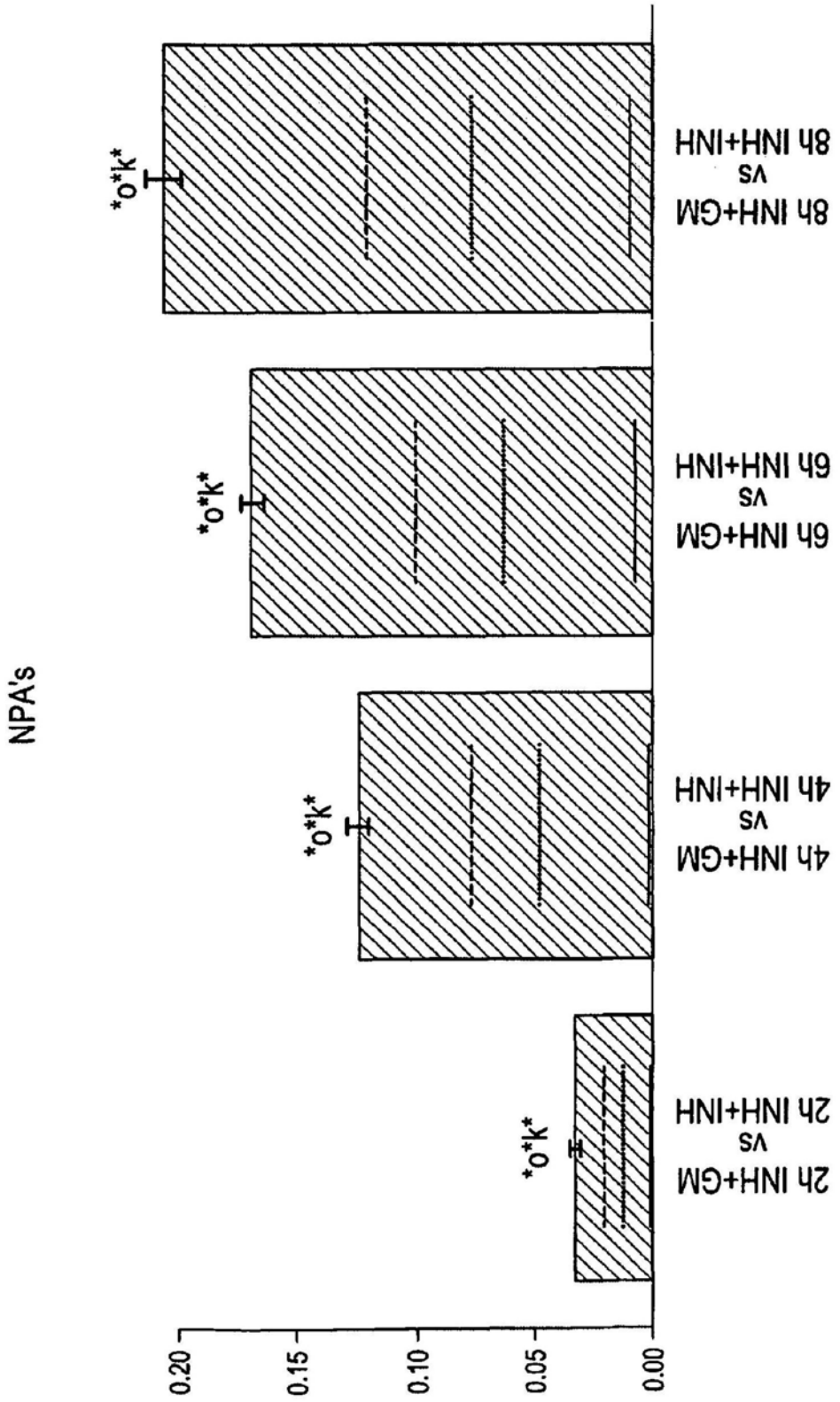

NPA's
0.20
0.15
0.10
0.05
0.00
2h INH+GM
vs
2h INH+INH
4h INH+GM
vs
4h INH+INH
6h INH+GM
vs
6h INH+INH
8h INH+GM
vs
8h INH+INH
*o*k*

图18