



(12)发明专利



(10)授权公告号 CN 103765448 B

(45)授权公告日 2017.05.17

(21)申请号 201280028438.X

(22)申请日 2012.06.11

(65)同一申请的已公布的文献号

申请公布号 CN 103765448 A

(43)申请公布日 2014.04.30

(30)优先权数据

11195417.8 2011.12.22 EP

61/495,824 2011.06.10 US

61/525,700 2011.08.19 US

(85)PCT国际申请进入国家阶段日

2013.12.10

(86)PCT国际申请的申请数据

PCT/EP2012/061033 2012.06.11

(87)PCT国际申请的公布数据

W02012/168481 EN 2012.12.13

(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 J·霍恩格 F·马丁 M·派奇

A·塞沃尔

(74)专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 马景辉

(51)Int.Cl.

G06N 7/06(2006.01)

G06N 5/02(2006.01)

G06F 19/24(2006.01)

(56)对比文件

US 2007225956 A1,2007.09.27,

US 2009117589 A1,2009.05.07,

US 7599933 B2,2009.10.06,

US 2005240563 A1,2005.10.27,

审查员 刘雪

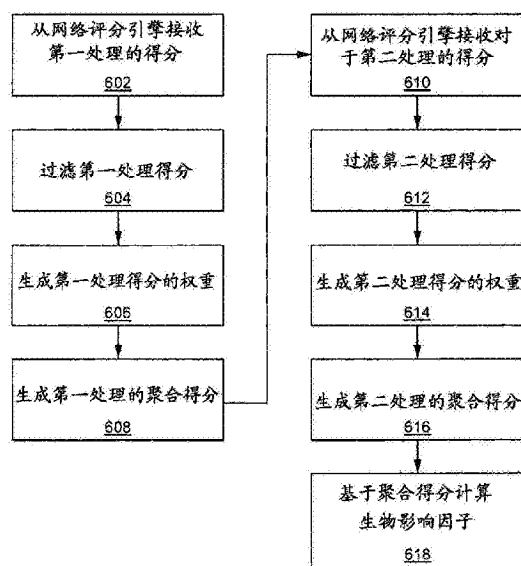
权利要求书3页 说明书40页 附图11页

(54)发明名称

用于量化生物扰动的影响的系统和方法

(57)摘要

提供了用于使用响应数据和基础生物机制的计算网络模型生成生物影响因子的系统和方法,该生物影响量化因子生物系统的扰动。生物影响因子可以表示扰动对于结果的长期影响。作为示例,该系统和方法可以基于暴露到诸如香烟烟雾的物剂来生成表示诸如肺癌的疾病的进展的生物影响因子。



1. 一种用于确定扰动对于生物系统的影响的计算机化的方法,包括:

在处理器处接收与生物系统对于第一处理的响应对应的第一数据集,其中,该生物系统包括多个生物实体,其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;

在处理器处接收与生物系统对于第二处理的响应对应的第二数据集,该第二处理与第一处理不同;

在处理器处提供表示生物系统的多个计算网络模型,每个模型包括表示多个生物实体的节点以及表示模型中的节点之间的关系;的边;

在处理器处基于第一数据集和所述多个模型来生成表示生物系统的扰动的第一得分集,以及基于第二数据集和所述多个计算模型来生成表示生物系统的扰动的第二得分集;以及

在处理器处生成具有单个标量值的数值生物影响因子,该生物影响因子基于第一得分集和第二得分集中的每一个得分集并且表示扰动对于生物系统的整体生物影响。

2. 根据权利要求1所述的方法,其中,数量多于两个的数据集被接收并且相应数量的得分集被生成。

3. 根据权利要求1所述的方法,其中,对于每个处理生成生物影响因子。

4. 根据权利要求1所述的方法,其中,至少一个数据集包括处理数据和相应的控制数据。

5. 根据权利要求1所述的方法,其中,多个网络中的至少一个是因果网络。

6. 根据权利要求1所述的方法,其中,每个的得分集内的得分由几何扰动指数评分技术、概率扰动指数评分技术或者预期扰动指数评分技术来独立地计算。

7. 根据权利要求1所述的方法,其中,第一得分集和第二得分集中的每个得分都包括得分向量,并且生成生物影响因子的步骤还包括在处理器处过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影。

8. 根据权利要求7所述的方法,其中,过滤还包括从经分解的第一得分和第二得分中的至少一个移除多个投影中的至少一个。

9. 根据权利要求7所述的方法,其中,基础向量集包括描述至少一个模型的矩阵的本征向量。

10. 根据权利要求1所述的方法,其中,生成第一得分集和第二得分集包括:

在处理器处基于相应的计算网络模型以及第一数据集和第二数据集中的至少一个来对第一得分集和第二得分集中的每一个得分分配权重;

聚合第一得分集中经加权的得分;

聚合第二得分集中经加权的得分;

其中,一个或多个生物影响因子是第一得分集和第二得分集的经聚合的得分的函数。

11. 根据权利要求1所述的方法,其中,生物影响因子是第一得分集和第二得分集的经聚合的得分的线性结合、线性变换或二次函数形式。

12. 根据权利要求10所述的方法,其中,对第一得分集和第二得分集中的每一个分配权重包括选择对于多个计算模型中每一个的权重,以将第一得分集内的得分和第二得分集内的得分之间的差异最大化。

13. 根据权利要求1所述的方法, 其中, 生成生物影响因子包括确定表示第一得分集经聚合的得分的第一向量与表示第二得分集经聚合的得分的第二向量之间的内积, 其中生物影响因子是与该内积相关联的角度。

14. 根据权利要求1所述的方法, 其中, 生成生物影响因子包括确定由表示第一得分集的第一向量限定的第一表面与由表示第二得分集的第二向量限定的第二表面之间的距离。

15. 根据前述权利要求中的任意一项所述的方法, 其中, 所述计算网络模型是从细胞增殖网络、炎性过程网络、细胞应激网络、以及DNA损伤、自吞噬、细胞死亡及衰老网络中选择的两个以上计算网络模型。

16. 根据权利要求1所述的方法, 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。

17. 根据权利要求1所述的方法, 其中, 第一处理包括以下项中的至少一个: 暴露到通过加热烟草生成的烟雾剂, 暴露到通过燃烧烟草生成的烟雾剂, 暴露到烟草烟雾, 暴露到香烟烟雾, 暴露到包括不存在于或不能从生物系统得到的分子或实体的混杂物质, 暴露到毒素、治疗性化合物、兴奋剂、松弛剂、天然产品、制造产品、食品物质中的至少一种, 以及暴露到镉、汞、铬、尼古丁、烟草特有亚硝胺及烟草特有亚硝胺的代谢物中的一个或多个。

18. 根据权利要求17所述的方法, 其中, 烟草特有亚硝胺的代谢物包括4-甲基亚硝胺基-1-(3-吡啶基)-1-丁酮4 (NNK)、N'-亚硝基去甲基尼古丁 (NNN)、N-亚硝基新烟草碱 (NAT)、N-亚硝基假木贼碱 (NAB) 和4-(甲基亚硝氨基)-1-(3-吡啶基)-1-丁醇 (NNAL)。

19. 根据权利要求1所述的方法, 还包括:

将生物影响因子与在不存在扰动或存在不同扰动的状态下已经获得的一个或多个附加生物影响因子相比较; 并且

其中该比较表示扰动对于生物系统的生物影响。

20. 根据权利要求1所述的方法, 其中, 生物影响因子表示或者被用来估计或确定由病原体、有害物质、生产的产品、用于安全评估或风险使用比较的生产的产品、治疗化合物或在环境或环境活性物质中的变化引起的期望的或不利的生物影响的幅度。

21. 根据权利要求19所述的方法, 其中, 两个以上的不同扰动被用来比较不同扰动对于生物系统的影响。

22. 根据权利要求19所述的方法, 其中, 一个或多个扰动表示至少两个不同处理条件。

23. 根据权利要求22所述的方法, 其中, 至少一个处理条件包括以下项中的至少一个: 暴露到通过加热烟草生成的烟雾剂, 暴露到通过燃烧烟草生成的烟雾剂, 暴露到烟草烟雾, 暴露到香烟烟雾, 暴露到包括不存在于或不能从生物系统得到的分子或实体的混杂物质, 以及暴露到毒素、治疗性化合物、兴奋剂、松弛剂、天然产品、制造产品、食品物质中的至少一种。

24. 根据权利要求1所述的方法, 其中, 所述扰动是由一个或多个物剂引起的。

25. 根据权利要求24所述的方法, 其中, 所述一个或多个物剂是从由以下各项构成的组中选择的: 通过加热烟草生成的烟雾剂、通过燃烧烟草生成的烟雾剂、烟草烟雾、香烟烟雾、以及任何其气体成分或粒子成分、镉、汞、铬、尼古丁、烟草特有亚硝胺及烟草特有亚硝胺的代谢物或者以上一个或多个的组合。

26. 根据权利要求25所述的方法, 其中, 烟草特有亚硝胺的代谢物包括4-甲基亚硝胺

基-1-(3-吡啶基)-1-丁酮4 (NNK)、N'-亚硝基去甲基尼古丁 (NNN)、N-亚硝基新烟草碱 (NAT)、N-亚硝基假木贼碱 (NAB) 和4-(甲基亚硝氨基)-1-(3-吡啶基)-1-丁醇 (NNAL)。

27. 根据权利要求1所述的方法, 其中, 将生物影响因子与已知的生物结果进行比较以校准生物影响因子的值。

28. 一种用于确定生物影响因子的计算机系统, 该计算机系统包括处理器, 该处理器被构造为:

接收与生物实体集对于第一处理的响应对应的第一数据, 其中, 该生物系统包括多个生物实体, 该多个生物实体包括生物实体集并且其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;

接收与生物实体集对于第二处理的响应对应的第二数据, 该第二处理与第一处理不同;

提供表示生物系统的多个计算因果网络模型, 每个计算模型包括表示多个生物实体的节点以及表示多个生物实体中的节点之间的关系的边;

基于第一数据和所述多个计算模型来生成表示生物系统的扰动的第一得分, 以及基于第二数据和所述多个计算模型来生成表示生物系统的扰动的第二得分; 以及

生成具有单个标量值的数值生物影响因子, 该生物影响因子基于第一得分和第二得分并且表示扰动对于生物系统的整体生物影响。

29. 根据权利要求28所述的计算机系统, 其中, 第一得分和第二得分中的每一个都包括得分向量, 并且其中处理器被进一步配置为:

过滤第一得分和第二得分, 以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影; 以及

从第一得分和第二得分中的至少一者移除多个投影中的至少一个。

30. 根据权利要求28所述的计算机系统, 其中, 基础向量集包括描述至少一个计算模型的矩阵的本征向量, 或者其中, 生成生物影响因子包括确定表示第一得分的第一向量与表示第二得分的第二向量之间的内积。

31. 根据权利要求28所述的计算机系统, 其中, 生成生物影响因子包括确定表示第一得分的第一表面与表示第二得分的第二表面之间的距离。

用于量化生物扰动的影响的系统和方法

背景技术

[0001] 人体经常受到暴露于在长期来看能够造成严重的健康风险的潜在有害物剂(agent)而扰动。暴露于这些物剂下能够危害人体内部的生物机制的正常机能。为了理解和量化这些扰动对人体的影响,研究人员研究生物系统对暴露于物剂作出响应的机制。有些团体已经在活体动物试验方法中广泛地使用。但是,动物试验方法并非总是足够的,因为关于它们的可靠性和关联性存在着疑问。在不同动物的生理方面存在着众多差异。因此,不同的物种对于暴露于物剂下可能会作出不同的响应。因此,关于从动物试验中获得的响应是否可以外推到人类生物学中是存在疑问的。其他方法包括通过对人类志愿者的临床研究来评估风险。但是这些风险评估是以后验方式执行的,并且因为疾病可能要花费数十年来表现出,所以这些评估可能不足以阐明将有害物质与疾病联系起来的机制。另外一些方法包括体外实验。尽管,基于体外细胞和组织的方法作为对它们的基于动物的相应方法的完全或部分替代方法已经获得普遍接受,但是这些方法价值有限。因为体外方法集中于细胞和组织机制的具体方面;它们并不总是会考虑在整个生物系统中发生的复杂的相互作用。

[0002] 在过去十年中,与传统的剂量相关的疗效和毒性测定结合的核酸、蛋白质及代谢物水平的高吞吐量测量结果已经作为用于阐明许多生物过程的作用机制的方法而出现。研究人员已经尝试将来自这些全异测量结果的信息与来自科学文献的关于生物路径的知识结合以构建(assembly)有意义的生物模型。为了这个目的,研究人员已经开始使用能够挖掘大量数据的数学及计算技术(例如,聚类和统计方法)来识别可能的生物作用机制。

[0003] 之前的工作同样探讨了揭示由对生物过程的一种或多种扰动引起的 基因表达变化的特性特征的重要性,以及关于该特征存在于作为对该过程的具体活动幅度的测量结果的附加数据集内的后续评分。在这方面的大部分工作已经涉及对与疾病表型关联的特征的识别和评分。这些表型衍生的特征提供显著的分类能力,但是在单个具体扰动与特征之间缺少机理或因果关系。因此,这些特征可以表示通过通常未知的机制导致相同的疾病表型的或由其引起的多种不同的未知扰动。

[0004] 一个挑战在于理解在生物系统中的各种个体生物实体的活动如何允许活化或抑制不同的生物机制。因为个体实体(例如,基因)会涉及多种生物过程(例如,发炎和细胞增殖),对基因活动的测量并不足以识别触发该活动的基础生物过程。

[0005] 当前技术都没有被应用来执行预测风险评价并且解决短期暴露到扰动与长期疾病结果之间的关系。通常,该问题是通过传统的纵向流行病学研究来解决的,但是这种研究可能呈现出道德挑战并且不能满足对于风险评价的当前急迫需求。实际上,对于新的物剂,传统的纵向流行病学研究不能被使用。因此,需要改善的系统和方法来研究扰动对于人体的影响。

发明内容

[0006] 这里描述了用于与从生物系统的实体的子集测量的活动数据来量化生物系统对于一个或多个扰动的响应的系统、方法和产品。

[0007] 在一个方面,提供了一种用于确定扰动对于生物系统的影响的计算机化的方法,包括:在处理器处接收与生物系统对于第一处理的响应对应的第一数据集,其中,该生物系统包括多个生物实体,其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;在处理器处接收与生物系统对于第二处理的响应对应的第二数据集,该第二处理与第一处理不同;在处理器处提供表示生物系统的多个计算网络模型,每个模型包括表示多个生物实体的节点以及表示模型中的节点之间的关系的边;在处理器处基于第一数据集和所述多个模型来生成表示生物系统的扰动的第一得分集,以及基于第二数据集和所述多个计算模型来生成表示生物系统的扰动的第二得分集;以及在处理器处基于第一得分集和第二得分集中的每一个来生成表示扰动对于生物系统的生物影响的一个或多个生物影响因子。

[0008] 在一个实施例中,两个以上的数据集被接收并且相应数量的得分集被生成。在某些实施例中,接收了三个以上、四个以上、五个以上、六个以上、七个以上、八个以上、九个以上或十个以上的数据集。在某些实施例中,接收了与扰动或处理至少相同数目的数据集。

[0009] 在一个实施例中,对于每个处理生成生物影响因子。

[0010] 在一个实施例中,至少一个数据集包括处理数据和相应的控制数据。

[0011] 在一个实施例中,多个网络中的至少一个是因果网络。

[0012] 在一个实施例中,每个的得分集内的得分由几何扰动指数评分技术、概率扰动指数评分技术或者预期扰动指数评分技术来独立地计算。

[0013] 在一个实施例中,第一得分集和第二得分集中的每个得分都包括得分向量,并且生成生物影响因子的步骤还包括在处理器处过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影。

[0014] 在一个实施例中,过滤还包括从经分解的第一得分和第二得分中的至少一个移除多个投影中的至少一个。

[0015] 在一个实施例中,基础向量集包括描述至少一个模型的矩阵的本征向量。

[0016] 在一个实施例中,生成第一得分集和第二得分集包括:在处理器处基于相应的计算网络模型和第一数据集和第二数据集中的至少一个来对第一得分集和第二得分集中的每一个分配权重;聚合第一得分集中经加权的得分;聚合第二得分集中经加权的得分;其中,一个或多个生物影响因子是第一得分集和第二得分集的经聚合的得分的函数。

[0017] 在一个实施例中,一个或多个生物影响因子是第一得分集和第二得分集的经聚合的得分的线性组合、线性变换或二次函数形式。

[0018] 在一个实施例中,对第一得分集和第二得分集中的每一个分配权重包括选择对于多个计算模型中每一个的权重,以将第一得分集内的得分和第二得分集内的得分之间的差异最大化。

[0019] 在一个实施例中,生成生物影响因子包括确定表示第一得分集的经聚合的得分的第一向量与表示第二得分集的经聚合的得分的第二向量之间的内积。

[0020] 在一个实施例中,生成生物影响因子包括确定由表示第一得分集的经聚合的得分的第一向量限定的第一表面与由表示第二得分集的经聚合的得分的第二向量限定的第二表面之间的距离。

[0021] 在一个实施例中,所述计算网络模型是从细胞增殖网络、炎性过程网络、细胞应激

网络、以及DNA损伤、自吞噬、细胞死亡及衰老网络中选择两个以上。

[0022] 在另一个方面,描述了一种计算机系统,该计算机系统包括处理器,该处理器被构造为:接收与生物实体集对于第一处理的响应对应的第一数据,其中,该生物系统包括多个生物实体,该多个生物实体包括生物实体集并且其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;接收与生物实体集对于第二处理的响应对应的第二数据,该第二处理与第一处理不同;提供表示生物系统的多个计算因果网络模型,每个计算模型包括表示多个生物实体的节点以及表示多个生物实体中的节点之间的关系的边;基于第一数据和所述多个计算模型来生成表示生物系统的扰动的第一得分,以及基于第二数据和所述多个计算模型来生成表示生物系统的扰动的第二得分;以及基于第一得分和第二得分来生成生物影响因子。

[0023] 在一个实施例中,第一得分和第二得分中的每一个都包括得分向量,并且其中处理器被进一步配置为:过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影;以及从第一得分和第二得分中的至少一者移除多个投影中的至少一个。

[0024] 在一个实施例中,基础向量集包括描述至少一个计算模型的矩阵的本征向量,或者其中,生成生物影响因子包括确定表示第一得分的第一向量与表示第二得分的第二向量之间的内积。

[0025] 在一个实施例中,生成生物影响因子包括确定表示第一得分的第一表面与表示第二得分的第二表面之间的距离。

[0026] 在一个实施例中,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。

[0027] 在一个实施例中,第一处理包括暴露到通过加热烟草生成的烟雾剂、暴露到通过燃烧烟草生成的烟雾剂、暴露到烟草烟雾、暴露到香烟烟雾、暴露到不存在于或不能从生物系统得到的分子或实体的混杂物质以及暴露到毒素、治疗性化合物、兴奋剂、松弛剂、天然产品、制造产品、食品物质中的至少一者、暴露到镉、汞、铬、尼古丁、烟草特有亚硝胺及其代谢物(4-甲基亚硝胺基-1-(3-吡啶基)-1-丁酮4(NNK)、N'-亚硝基去甲基尼古丁(NNN)、N-亚硝基新烟草碱(NAT)、N-亚硝基假木贼碱(NAB)和4-(甲基亚硝氨基)-1-(3-吡啶基)-1-丁醇(NNAL))中的一个或多个。

[0028] 在另一个方面,描述了一种计算机程序产品,其包括适合于执行这里公开的方法的计算机代码。

[0029] 在另一个方面,描述了一种计算机或计算机可读介质,其包括计算机程序产品。

[0030] 在另一个方面,提供了一种用于确定扰动对于生物系统的生物影响的方法,包括:生成表示扰动对于生物系统的生物影响的一个或多个生物影响因子,其中至少一个生物影响因子是根据这里描述的计算机化方法来确定的;将所述一个或多个生物影响因子与不存在扰动或存在不同扰动的状态下已经获得的一个或多个生物影响因子相比较;并且其中该比较表示扰动对于生物系统的生物影响。

[0031] 在另一个方面,提供了一种用于确定扰动对于生物系统的生物影响的计算机化的方法,包括:生成表示扰动对于生物系统的生物影响的一个或多个生物影响因子,其中至少一个生物影响因子是根据权利要求1-15、21或22中任意一项描述的计算机化方法来确定

的;将所述一个或多个生物影响因子与在不存在扰动或存在不同扰动的状态下已经获得的一个或多个生物影响因子相比较;并且其中该比较表示扰动对于生物系统的生物影响。

[0032] 在另一个方面,提供了一种用于确定扰动对于生物系统的生物影响的方法,包括:生成表示扰动对于生物系统的生物影响的一个或多个生物影响因子,其中至少一个生物影响因子是根据这里描述的计算机化方法来确定的;将所述一个或多个生物影响因子与在不存在扰动或存在不同扰动的状态下已经获得的一个或多个生物影响因子相比较;并且其中该比较表示扰动对于生物系统的生物影响。

[0033] 在一个实施例中,一个或多个生物影响因子表示或者被用来估计或确定由病原体、有害物质、生产的产品、用于安全评估或风险使用比较生产的产品、治疗化合物或在环境或环境活性物质中的变化引起的期望的或不利的生物影响的幅度。

[0034] 在一个实施例中,两个以上的不同扰动被用来比较不同扰动对于生物系统的影响。

[0035] 在一个实施例中,一个或多个扰动表示至少两个不同处理条件。

[0036] 在一个实施例中,至少一个处理包括暴露到通过加热烟草生成的烟雾剂、暴露到通过燃烧烟草生成的烟雾剂、暴露到烟草烟雾、暴露到香烟烟雾、暴露到包括不存在于或不能从生物系统得到的分子或实体的混杂物质以及暴露到毒素、治疗性化合物、兴奋剂、松弛剂、天然产品、制造产品、食品物质中的至少一者。

[0037] 在一个实施例中,所述扰动是由一个或多个物剂引起的。

[0038] 在一个实施例中,物剂是从由以下各项构成的组中选择的:通过加热烟草生成的烟雾剂、通过燃烧烟草生成的烟雾剂、烟草烟雾、香烟烟雾、以及任何其气体成分或粒子成分、镉、汞、铬、尼古丁、烟草特有亚硝胺及其代谢物(4-甲基亚硝胺基-1-(3-吡啶基)-1-丁酮4 (NNK)、N'-亚硝基去甲基尼古丁 (NNN)、N-亚硝基新烟草碱 (NAT)、N-亚硝基假木贼碱 (NAB) 和4-(甲基亚硝氨基)-1-(3-吡啶基)-1-丁醇 (NNAL)) 或者以上一个或多个的组合。

[0039] 在一个实施例中,至少一个生物影响因子已经预先确定或预先计算。

[0040] 在另一个方面,提供了一种用于确定扰动对于生物系统的影响的计算机化的方法,包括:在处理器处接收与生物实体集对于第一处理的响应对应的第一数据,其中,该生物系统包括包含多个生物实体集的多个生物实体,其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;在处理器处接收与生物实体集对于第二处理的响应对应的第二数据,该第二处理与第一处理不同;在处理器处提供表示生物系统的多个计算因果网络模型,每个计算模型包括表示多个生物实体的节点以及表示多个生物实体中的实体之间的关系的边;在处理器处基于第一数据和所述多个模型来生成表示生物系统的扰动的第一得分,以及基于第二数据和所述多个计算模型来生成表示生物系统的扰动的第二得分;以及在处理器处基于第一得分和第二得分来生成表示扰动对于生物系统的生物影响的生物影响因子。

[0041] 在一个实施例中,第一得分和第二得分中的每个得分都包括得分向量,并且生成生物影响因子的步骤还包括在处理器处过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影。

[0042] 在一个实施例中,过滤还包括从经分解的第一得分和第二得分中的至少一个移除多个投影中的至少一个。

[0043] 在一个实施例中,基础向量集包括描述至少一个计算模型的矩阵的本征向量。

[0044] 在一个实施例中,生成第一得分和第二得分包括:在处理器处基于相应的计算模型和第一数据和第二数据中的至少一个来对多个计算模型中的每一个分配权重;在处理器处生成与多个计算模型相对应并且基于第一数据的多个第一得分;以及在处理器处生成与多个计算模型相对应并且基于第二数据的多个第二得分;根据所分配的权重将多个第一得分结合;根据所分配的权重将多个第二得分结合;其中,生物影响因子是结合的多个第一得分和结合的多个第二得分的函数。

[0045] 在一个实施例中,对多个计算模型中的每一个分配权重包括选择对于多个计算模型中每一个的权重,以将多个第一得分和多个第二得分之间的差异最大化。

[0046] 在一个实施例中,生成生物影响因子包括确定表示第一得分的第一向量与表示第二得分的第二向量之间的内积。

[0047] 在一个实施例中,生成生物影响因子包括确定表示第一得分的第一表面与表示第二得分的第二表面之间的距离。

[0048] 在一个实施例中,所述计算因果网络模型是从细胞增殖网络、炎性过程网络、细胞应激网络、以及DNA损伤、自吞噬、细胞死亡及衰老网络中选择两个以上。

[0049] 在另一方面,提供了一种用于确定生物影响因子的计算机系统,其包括适合于实施计算机化方法的设备。

[0050] 在一个实施例中,该计算机系统包括处理器,该处理器被构造为:接收与生物实体集对于第一处理的响应对应的第一数据,其中,该生物系统包括多个生物实体,该多个生物实体包括生物实体集并且其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;接收与生物实体集对于第二处理的响应对应的第二数据,该第二处理与第一处理不同;提供表示生物系统的多个计算因果网络模型,每个计算模型包括表示多个生物实体的节点以及表示多个生物实体中的节点之间的关系的边;基于第一数据和所述多个计算模型来生成表示生物系统的扰动的第一得分,以及基于第二数据和所述多个计算模型来生成表示生物系统的扰动的第二得分;以及基于第一得分和第二得分来生成生物影响因子。

[0051] 在一个实施例中,第一得分和第二得分中的每一个都包括得分向量,并且其中处理器被进一步配置为:过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影;以及从第一得分和第二得分中的至少一者移除多个投影中的至少一个。

[0052] 在一个实施例中,基础向量集包括描述至少一个计算模型的矩阵的本征向量。

[0053] 在一个实施例中,生成生物影响因子包括确定表示第一得分的第一向量与表示第二得分的第二向量之间的内积。

[0054] 在一个实施例中,生成生物影响因子包括确定表示第一得分的第一表面与表示第二得分的第二表面之间的距离。

[0055] 在一个实施例中,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。在一个实施例中,第一处理包括暴露到通过加热烟草生成的烟雾剂、暴露到通过燃烧烟草生成的烟雾剂、暴露到烟草烟雾、暴露到香烟烟雾、暴露到包括不存在于或不能从生物系统得到的分子或实体的混杂物质以及暴露到毒素、治疗性化合

物、兴奋剂、松弛剂、天然产品、制造产品、食品物质中的至少一者。

[0056] 在另一个方面,提供了一种计算机程序产品,其包括适合于执行本发明的计算机化方法的程序代码。

[0057] 在另一个方面,提供了一种计算机或计算机可读介质,其包括本发明的计算机程序产品。

[0058] 在一方面,这里描述的系统和方法涉及用于量化扰动对于生物系统的影响(例如,响应于诸如物剂暴露的处理条件或者响应与多个处理条件)的计算机化方法(例如,计算机实施的方法)和一个或多个计算机处理器。处理器接收与生物实体集对第一处理的响应相对应的第一数据。生物实体集是包括在生物系统中的多个生物实体的一部分。生物系统中的每个生物实体与生物系统中的至少另一个生物实体彼此影响。处理器也接收与生物实体集对第二处理的响应相对应的第二数据,该第二处理与第一处理不同。处理器也提供表示生物系统的多个计算因果网络模型。每个计算模型包括表示多个生物实体的节点以及表示多个生物实体中的实体之间的关系的边。

[0059] 处理器之后基于第一数据和多个计算模型生成生物系统的扰动的第一得分,以及基于第二数据和多个计算模型来生成表示生物系统的扰动的第二得分。处理器之后基于第一得分和第二得分生成“生物响应因子”或“BIF”。在各个实施方式中,计算机化方法将与多个处理(或物剂)相对应的多个模型得分结合,并且生成表示由处理(或物剂)引起的相对生物效果的BIF。在一些实施方式中,生成生物影响因子包括确定表示第一得分的第一向量与表示第二得分的第二向量之间的内积。在一些实施方式中,生成生物影响因子包括确定表示第一得分集的第一表面与表示第二得分的第二表面之间的距离。

[0060] 在一些实施方式中,第一和第二得分中的每个都包括得分向量,并且生成生物影响因子的步骤还包括在处理器处过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影。过滤还可以包括从经分解的第一得分和第二得分中的至少一者移除多个投影中的至少一个。基础向量集包括表示至少一个计算模型的矩阵(诸如拉普拉斯算子矩阵)的本征向量。

[0061] 在一些实施方式中,生成第一和第二得分包括基于响应计算模型以及第一和第二数据中的至少一个来对于多个计算模型中的每个分配权重。权重可以被分配例如以将第一得分和第二得分之间的差异最大化。处理器还可以生成与多个计算模型相对应并基于第一数据的多个第一得分以及与多个计算模型相对应并基于第二数据的多个第二得分。处理器之后可以根据所分配的权重将多个第一得分结合并且根据所分配的权重将多个第二得分结合。在一些这种实施方式中,生物影响因子是结合的多个第一得分和结合的多个第二得分的函数。

[0062] 在某些实施方式中,生物系统包括但不局限于细胞增殖机制、细胞应激机制、细胞发炎机制、DNA修复机制、DNA损伤机制、自吞噬机制、细胞死亡机制和衰老机制中的至少一个。处理可以包括但不局限于暴露到多种物质,包括存在于生物系统中或从其得出的分子或实体。处理可以包括但不局限于:暴露到毒素、治疗性化合物、兴奋剂、松弛剂、天然产品、制造产品和食品物质。处理可以包括但不局限于暴露到通过加热烟草生成的烟雾剂、通过燃烧烟草生成的烟雾剂、烟草烟雾和香烟烟雾。处理可以包括但不局限于暴露到镉、汞、铬、烟碱、烟草特有的亚硝胺及它们的代谢物(4-甲基亚硝胺基-1-(3-吡啶基)-1-丁酮4 (NNK)、

N'-亚硝基去甲基尼古丁 (NNN)、N-亚硝基新烟草碱 (NAT)、N-亚硝基假木贼碱 (NAB) 和4-(甲基亚硝氨基)-1-(3-吡啶基)-1-丁醇 (NNAL))。在某些实施方式中,物剂包括用于尼古丁替换疗法的产品。

[0063] 本文所描述的计算机化的方法可以在具有一个或多个计算器件的计算机化系统中实现,每个计算器件包括一个或多个处理器。一般地,本文所描述的计算机化系统可以包括一个或多个引擎,该一个或多个引擎包括被配置为具有硬件、固件和软件以执行本文所描述的一种或多种计算机化的方法的一个或多个处理器件,例如,计算机、微处理器、逻辑器件或者其他器件或处理器。在某些实现方式中,计算机化系统包括系统响应曲线引擎、网络建模引擎和网络评分引擎。引擎可以时不时地互连,并且还时不时地与一个或多个数据库连接,包括扰动数据库、可测量数据库、实验数据数据库和文献数据库。本文所描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。这样的实现方式可能适合于经由多种通信系统进行的分布式计算。

附图说明

[0064] 本公开内容的更多特征及其性质和各种优点在考虑了下面结合附图进行的详细描述的情况下将变得显而易见,在附图中同样的参考符号在所有附图中指代相同的部分,并且在附图中:

[0065] 图1是用于量化生物扰动的影响的示例性计算机化系统的框图。

[0066] 图2是用于生成生物影响因子(BIF)的示意性处理的流程图。

[0067] 图3是作为系统响应曲线的基础的数据的图形表示,系统响应曲线包括两个物剂、两个参数和N个生物实体的数据。。

[0068] 图4是具有若干生物实体以及它们的关系的生物网络的计算模型的图示。

[0069] 图5是用于生成BIF的示意性计算机化聚集引擎的框图。

[0070] 图6是用于从网络响应得分生成BIF的示例性过程的流程图。

[0071] 图7示出了网络响应得分向量的示例性分解。

[0072] 图8A和图8B示出了对于网络响应得分向量的是理性过滤运算。

[0073] 图9示出了在网络响应得分聚集期间网络加权的示例。

[0074] 图10描绘了可以由图1的计算机化系统分析的两个网络响应得分表面。

[0075] 图11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。

[0076] 图12是可以用来实现本文所描述的任意计算机化系统中的任意构件的示例性计算器件的框图。

[0077] 图13描绘了根据这里公开的系统和方法 and 示意性实施例生成的鼻上皮肿瘤发生BIF的实验结果。

[0078] 图14示出了用于生物影响因子聚集的实验设计的系统方法。一些良好挑选的生物系统被以时间和剂量相关的方式暴露到物质,以生成系统级数据,该系统级数据将会在与疾病袭击相关的每个生物网络的上下文中被解释。

[0079] 图15示出了使用在与疾病袭击相联系的生物网络的上下文中分析的系统级数据,对于给定生物活化物质得出生物影响因子的计算处理。

具体实施方式

[0080] 图1是用于量化一个或多个扰动的生物影响的计算机化系统100的框图。特别地,系统100包括系统响应曲线引擎110、网络建模引擎112、网络评分引擎114和聚合引擎116。引擎110、112、114和116时不时地互连,并且还时不时地与一个或多个数据库连接,包括扰动数据库102、可测量数据库104、实验数据数据库106和文献数据库108。如同本文所使用的,引擎包括一个或多个处理器件,例如,计算机、微处理器、逻辑器件或者参照图12所描述的一个或多个其他器件,该一个或多个处理器件以硬件、固件和软件来配置以执行这里描述的一个或多个计算技术。

[0081] 在操作期间,对于给定扰动,系统100生成生物影响因子(BIF),其为影响的量化测量结果,包括扰动对于包括人体的生物系统的长期影响。更具体地,系统100生成或提供一个或多个生物系统和机制相对于扰动的类型、感兴趣的理想生物机制或感兴趣的具体长期结果的计算机化模型(一同称作为“生物网络”)。例如,在细胞已经被暴露到香烟烟雾中时,系统100可以生成或提供用于细胞增殖的机制的计算模型。在这种示例中,系统100也可以生成或提供表示疾病的不同阶段的一个或多个计算模型,包括但不限于肺部疾病和心血管疾病。在特定方面中,系统100基于所施加的扰动(例如,暴露到物剂)、感兴趣的可测量量、被研究的结果(例如,细胞增殖、细胞应激、发炎、DNA修复)、通过科学文献获得的实验结果和知识中的至少一个来生成这些计算机化模型。系统100测量并量化处理的效果以生成BIF。预测/验证引擎122可以接收一个或多个BIF值并且可以使用这些BIF值来进行结果预测(例如,在有毒物质被从人的环境移除之后降低癌症发生率或可能性)。预测/验证引擎122也可以或者替换地将BIF值与已知的生物结果相比较,以校准BIF值或者炎症BIF值的预测。校准和验证的示例由以下的图13中示出的结果来表示。系统100的各个组件和引擎包括软件和硬件组件中的至少一个,并且将会参照图11和图12进一步描述。

[0082] 图2是根据一种实现方式的用于通过计算生物影响因子(BIF)来量化扰动对于生物网络的影响的过程200的流程图。过程200的步骤将被描述为通过图1的系统100的各种构件来执行,但是这些步骤中的任一个都可以通过任意适合的硬件或软件构件(本地或远程)来执行,并且可以按照任何适当的顺序排列或者并行地执行。在步骤210,系统响应曲线(SRP)引擎110接收来自各种不同来源的生物数据,并且数据本身可以是各种不同类型的。在步骤210的一些实施方式中,SRP引擎110接收与生物实体集对于第一处理的响应相对应的第一数据并且接收与该生物实体集对于第二处理的响应相对应的第二数据,第二处理与第一处理不同。例如,在步骤210接收的数据可以包括来自于其中生物系统通过暴露到物剂或环境条件而受到扰动的实验的数据,并且也可以包括控制数据。

[0083] 在本发明的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,有机体在此称为受试者。受试者一般为哺乳动物,包括人类。受试者能够是人类群体中的个体人。本文所使用的术语“哺乳动物”包括(但不限于)人类、非人灵长类、小鼠、老鼠、狗、猫、牛、羊、马和猪。除人之外的哺乳动物能够有利地被用作能够用来提供人类疾病的模型的受试者。非人受试者能够是未改性的、转基因动物(例如基因改造的动物或者携带有一个或多个基因突变或沉默基因的动物)。受试者能够是雄性或雌性。取决于操作的目的,受试者能够是已经暴露于所感兴趣的物剂的受试者。受试者能够是已经在很长一段时间(可选地包

括本研究之前的时间)内暴露于物剂的受试者。受试者能够是已经在一段时间内暴露于物剂的、但是不再与物剂接触的受试者。受试者能够是已经被诊断或被识别为具有疾病的受试者。受试者能够是已经经历过或者正在经历疾病或不利的健康状况的处理的受试者。受试者还可以是已经展示出针对具体的健康状况或疾病的一个或多个症状或风险因素的受试者。受试者能够是预感染的疾病,并且可以有疾病症状或没有疾病症状的受试者。在某些实现方式中,所讨论的疾病或健康状况与在很长的一段时间内暴露于物剂或使用物剂关联。根据某些实现方式,系统100(图1)含有或生成与所感兴趣的扰动类型或结果相关的一个或多个生物系统的计算机化模型及其功能机制(共同地,“生物网络”或“网络模型”)。

[0084] 取决于操作的环境,生物系统能够被定义于不同的水平,因为它涉及总体中的个体有机体的功能,有机体一般为器官、组织、细胞类型、细胞器、细胞组分或具体个体的细胞。每个生物系统包括一个或多个生物机制或路径,该一个或多个生物机制或路径的操作显现为系统的功能特征。用于再现所定义的人类健康状况的特征且适用于暴露于所感兴趣的物剂的动物系统是优选的生物系统。用于反映在疾病病因或病理中所涉及的细胞类型和组织的细胞和器官型系统同样是优选的生物系统。能够对尽可能地概括体内的人体生物学的原始细胞或器官培养赋予优先级。同样重要的是将体外人类细胞培养与从动物模型中得出的最相当的体内培养匹配。这允许使用匹配的体外系统作为参照系统来生成从体内的动物模型到人类生物学的转化连续。因此,为了用于本文所描述的系统和方法而构想出的生物系统能够通过功能特征(生物功能、生理功能或细胞功能)、细胞器、细胞类型、组织类型、器官、发展阶段或者前述项的组合(不作限定)来定义。生物系统的示例包括,但不限于,肺、外皮、骨骼、肌肉、神经(中枢和末梢)、内分泌、心血管、免疫、循环、呼吸、泌尿、肾脏、胃肠、结直肠、肝脏和生殖系统。其他的示例生物系统包括,但不限于,在上皮细胞、神经细胞、血液细胞、连接组织细胞、平滑肌细胞、骨骼肌细胞、脂肪细胞、卵子细胞、精子细胞、干细胞、肺细胞、脑细胞、心肌细胞、喉部细胞、咽部细胞、食管细胞、胃部细胞、肾脏细胞、肝细胞、乳腺细胞、前列腺细胞、胰腺细胞、胰岛细胞、精巢细胞、膀胱细胞、宫颈细胞、子宫细胞、结肠细胞和直肠细胞中的各种细胞功能。某些细胞可以是细胞系的细胞,在适当的培养条件下体外培养或体外无限保持。也可以被认为是生物系统的功能特征的细胞功能的示例包括但不限于,细胞增殖(例如,细胞分裂)、变性、再生、衰老、由核子对细胞活化的控制、细胞至细胞信令、细胞分化、细胞反分化、分泌、迁移、吞噬、修复、凋亡和发展规划。能够将其视为生物系统的细胞组分的示例包括,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、肽聚糖和抗体。

[0085] 在生物系统中的扰动能够通过暴露或者与生物系统的一个或多个部分接触而由一个或多个物剂在一段时间内导致。物剂能够是单一物质或物质的混合物,包括其中并非全部成分都被识别或被特征化的混合物。物剂或其成分的化学和物理性质可以是没有完全特征化的。物剂能够由其结构、其成分或者在某些限定条件下会生成物剂的来源来定义。物剂的示例是非均质物质,即,不存在于生物系统内的或者不是从其中得出的分子或实体,以及在与生物系统接触之后由其生成的任何中间物或代谢物。物剂能够是碳水化合物、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、神经递质、无机化合物、有机化合物、环境物剂、微生物、粒子、环境条件、环境力或体力。物剂的非限制性示例包括(但不限于)营养物、代谢废物、毒物、毒品、毒素、治疗化合物、刺激物、弛缓剂、天然产

品、制造产品、食物材料、病原菌(朊蛋白、病毒、细菌、真菌、原生动物)、其尺寸在微米范围内或以下的粒子或实体、上述项目的副产品以及上述项的混合物。物理物剂的非限制性示例包括放射物、电磁波(包括阳光)、温度的升高或降低、剪切力、流体压力、放电或其后果或外伤。

[0086] 有些物剂不会扰动生物系统,除非它达到阈值浓度或者它与生物系统接触达一段时间,或者两者的结合。导致扰动的物剂暴露或接触可以按照剂量来量化。因而,扰动能够由长期暴露于物剂引起。暴露的时长能够由时间的单位、由暴露的频率或者由在受试者的实际或估计的寿命内的时间百分比来表示。扰动还能够由从生物系统的一个或多个部分中抑制物剂(如同以上所描述的)或者限制对生物系统的一个或多个部分的物剂供给导致。例如,扰动能够由营养物、水、碳水化合物、蛋白质、脂质、生物碱、维生素、矿物质、氧、离子、酶、激素、神经递质、抗体、细胞因子、光线的减少供应或缺乏,或者通过限制有机体的某些部分的移动,或者通过抑制或要求锻炼导致。

[0087] 物剂可以导致不同的扰动,这取决于生物系统的哪个(哪些)部分被暴露以及暴露条件。物剂的非限制性示例可以包括通过加热烟草生成的烟雾剂、通过燃烧烟草生成的烟雾剂、烟草烟雾或香烟烟雾,以及它们的气态成分或微粒成分中的任一种。物剂的更多非限制性示例包括镉、汞、铬、烟碱、烟草专有的亚硝胺及它们的代谢物(诸如4-甲基亚硝胺基-1-(3-吡啶基)-1-丁酮4(NNK)、N'-亚硝基去甲基尼古丁(NNN)、N-亚硝基新烟草碱(NAT)、N-亚硝基假木贼碱(NAB)和4-(甲基亚硝氨基)-1-(3-吡啶基)-1-丁醇(NNAL)),以及用于烟碱替代疗法的任何产品。用于物剂的暴露方案或复合刺激应当反映在日常设定中暴露的范围和环境。标准的暴露方案的设置能够被设计为用于系统地应用于同样良好定义的实验系统。每个化验能够被设计为用于收集时间和剂量相关的数据,以采集早期的及后期的事件并且确保覆盖典型的剂量范围。但是,本领域技术人员应当理解,本文所描述的系统和方法可以被调整和修改,使其适用于正在处理的应用,并且本文所设计的系统和方法可以在其他合适的应用中使用,并且其它此类添加和修改不应脱离本发明的范围。

[0088] 在各种实现方式中,对基因表达、蛋白质表达或周转、微核糖核酸表达或周转、翻译后修饰(post-translational modifications)、蛋白质修饰、易位、抗体生成代谢物图谱或者上述项的两项或更多项的组合的高输出的系统级测量结果在各种条件(包括各自的控制)下生成。功能性结果测量在本文所描述的方法中是所希望的,因为它们一般地能够用作用于评估的锚定并且表示在疾病病因中的清晰步骤。

[0089] 如同本文所使用的,“样品”指的是独立于受试者或实验系统的任何生物样品(例如,细胞、组织、器官或整个动物)。样品能够包括(不作限定)单个细胞或多个细胞、细胞组分、组织活检、切除的组织、组织提出物、组织、组织培养提出物、组织培养介质、呼出气体、全血、血小板、血清、血浆、红血球、白细胞、淋巴球、中性粒细胞、巨噬细胞、B细胞或其子集、T细胞或其子集、造血细胞子集、内皮细胞、滑体、淋巴液、腹水液、间质液、骨髓、脑脊液、胸腔积液、肿瘤渗入物、唾液、黏液、痰、精液、汗液、尿液,或者任何别的体液。样品能够通过以下方法从受试者中获得,包括(但不限于):静脉穿刺、排泄、活检、针刺、灌洗、刮切、手术切除,或者本技术领域已知的其他方法。

[0090] 在操作期间,对于给定的生物机制、结果、扰动或者前述项的组合,系统100能够生成响应得分值,该网络幅度值是对在网络中的生物实体响应于处理条件而生成的状态变化

的量化测量结果。得分集中的得分的数目可以对应于网络的数目。

[0091] 系统100(图1)包括与感兴趣的健康状况、疾病或生物结果相关的一个或多个计算机化的网络模型。这些网络模型中的一个或多个基于现有的生物知识,并且能够由外部来源上传并在系统100内管理。还能够基于测量在系统100内重新生成模型。可测量元件因此通过使用现有知识被集成到生物网络模型之内。以下所描述的是表示能够用来生成或精炼网络模型的感兴趣的生物系统的变化的或者表示对于扰动的响应的数据类型。

[0092] 返回图2,在步骤210,系统响应曲线(SRP)引擎110接收生物数据。SRP引擎110可以接收从各种不同来源接收该数据,并且数据本身可以是各种不同类型的。由SRP引擎110使用的生物数据可以从文献数据库(包括来自医药产品或医疗装置的临床前的、临床的及临床后的试验的数据)、基因组数据库(基因组序列及表达数据,例如,国家生物技术信息中心的基因表达库(Gene Expression Omnibus)或者欧洲生物信息研究所的ArrayExpress(Parkinson等,2010,Nucl.Acids Res.,doi:10.1093/nar/gkq1040.Pubmed ID21071405))、可购买的数据库(例如,美国马里兰州盖瑟斯堡的Gene Logic)或实验工作取得。数据可以包括来自一个或多个不同来源的原始数据,例如,使用为了研究特定的处理条件或暴露于特定物剂的影响而特别设计的一个或多个物种的体外实验、间接体内或直接体内实验。体外实验系统可以包括表示人类疾病的关键方面的组织培养或器官型培养(三维培养)。在这样的实现方式中,用于这些实验的物剂剂量和暴露方案可以基本上反映可以针对人类在正常使用或活动条件期间或者在特殊使用或活动条件期间预料的暴露范围和环境。实验参数和试验条件可以根据期望来选择以反映物剂的性质和暴露条件、所讨论的生物系统的分子和路径、所涉及的细胞类型和组织、感兴趣的结果以及疾病病因的各方面。从特定的动物模型中得到的分子、细胞或组织可以与特定的人类分子、细胞或组织培养匹配,以提高基于动物的发现的可译性。

[0093] 除了别的以外,由SRP引擎110接收到的许多都由高吞吐量的实验技术生成的数据包括(但不限于)与核酸(例如,具体DNA或RNA种类的绝对量或相对量、DNA序列的变化、RNA序列、三级结构的变化,或者由序列化、在微阵列上的核酸特定的杂交、量化的聚合酶链式反应或者本技术领域已知的其他技术所确定的甲基化模式)、蛋白质/肽(例如,由本技术领域已知的方法确定的蛋白质的绝对量或相对量、蛋白质的具体片段、肽聚糖、二级或三级结构的变化或者翻译后修饰)以及在某些条件下的机能活动(例如,酶的活动、蛋白质水解活动、转译 调整活动、运输活动、与某些结合配对物的结合亲合力)相关的数据。包括蛋白质或肽的翻译后修饰在内的修饰能够包括(但不限于)甲基化、乙酰化、法呢酰化、生物素酰化、硬脂酰化、甲酰化、豆蔻酰化、蛋白质棕榈酰化、香叶酰香叶酰化、聚乙二醇化、磷酸化、硫酸化、糖基化、糖化改变(sugar modification)、脂质化、脂质改变、泛素化、sumolation、二氧化硫键合、半胱氨酰化、氧化、谷胱甘肽、羧化、醛糖酸化反应和脱酰胺化。另外,蛋白质还能够通过一系列的反应来进行翻译后修饰,例如,产生糖化蛋白质产物的Amadori反应、席夫碱反应和美拉德反应。

[0094] 数据还可以包括测得的机能结果,例如,但不限于,处于细胞层面的机能结果,包括细胞增殖、发展命运和细胞死亡,处于生理层面的机能结果,包括肺活量、血压、锻炼熟练度。数据还可以包括疾病活动或严重性的测量值,例如,但不限于,肿瘤转移、肿瘤好转、功能丧失以及在疾病的某一阶段的预期寿命。疾病活动能够通过其结果为能够在所定义的条件

件下根据来自一个或多个受试者的样品(或者样品的总体)的评价来获得的值或一组值的临床评估来测量。临床评估还能够基于由受试者提供的对访问或调查问卷的回答。

[0095] 这种数据可以为了明确地在确定系统响应曲线中使用而生成,或者可以在之前的实验中产生或者已发表于文献中。一般地,数据包括与分子、生物结构、生理条件、基因特征或表型相关的信息。在某些实现方式中,数据包括关于条件、位置、数量、活动,或者分子的结构、生物结构、生理条件、基因特征或表型的描述。如后面将描述的,在临床设定中,数据可以包括从对从人类受试者中获得的样品执行的化验或对暴露于物剂的人类受试者的观察中获得的原始数据或已处理的数据。

[0096] 在步骤212,系统响应曲线(SRP)引擎110基于在步骤212接收到的生物数据来生成系统响应曲线(SRP)。SRP是用于表示在生物系统内的一个或多个测量的实体(例如,分子、核酸、肽、蛋白质、细胞等)响应于施加于生物系统的扰动(例如,暴露于物剂)而单独改变的程度表示。该步骤可以包括背景校正、标准化、倍数变化计算、显著性确定以及差别响应(例如,表达不同的基因)的识别中的一项或多项。在一个示例中,为了生成SRP,SRP引擎110收集用于施加于给定的实验系统(“系统-处理”对)的一组给定参数(例如,处理或扰动条件)的一组测量值。图3示出了两个SRP:包括经受到具有变化的参数(例如,暴露于第一处理物剂的剂量和时间)的第一处理306的N个不同的生物实体的生物活动数据的SRP302,以及包括经受到第二处理308的N个不同的生物实体的生物活动数据的类似的SRP304。包括于SRP内的数据可以是原始实验数据、已处理的实验数据(例如,被过滤以去除异常值,以置信估计来标记,针对多次试验求均值)、由计算生物模型生成的数据,或者从科学文献中取得的数据。SRP可以以众多方式来表示数据,例如,绝对值、绝对变化、倍数变化、对数变化、函数和表格。SRP引擎110将SRP传递给网络建模引擎112。

[0097] 在步骤214,网络建模引擎112提供生物系统的多个计算模型,其包括在步骤210已经获得其数据的生物实体。每个计算模型包括表示生物实体的节点和表示生物系统中的生物实体之间的关系的边。网络建模引擎112可以从包括多个网络模型的一个或多个数据库得出这些计算模型,每个网络模型被选择为与物剂或感兴趣的特征相关。选择可以基于以系统的生物功能为基础的机制的在先知识。在特定实施方式中,网络建模引擎112可以使用系统响应曲线、数据库中的网络和在文献中预先描述的网络来提取系统内的实体之间的因果关系,由此产生、改良或扩展网络模型。

[0098] 在步骤214的一些实施方式中,网络建模引擎112将来自于SRP引擎110的系统响应曲线用于网络模型,该网络模型基于以系统的生物功能为基础的机制。虽然在前一步骤中得出的SRP表示将用于确定网络扰动的大小的实验数据,但其是作为用于计算和分析的基础的生物网络模型。该分析要求初始开发与生物系统的特征相关的机制和路径的详细网络模型。这样的架构提供除已经在更典型的基因表达分析中使用的基因列表的检查之外的机理层理解。生物系统的网络模型是表示动态生物系统的且通过组装有关生物系统的各种基本性质的量化信息来建立的数学构造。

[0099] 这样的网络架构可以是迭代过程。关于网络边界的描绘通过研究与感兴趣的特征(例如,在肺中的细胞增殖)相关的机制和路径的科学文献来指导。用于描述这些路径的因果关系从现有的知识中取得以使网络集结。基于文献的网络能够使用含有相关的表型端点的高吞吐量的数据集来验证。SRP引擎110能够用来分析数据集,该分析结果能够用来证实,

精炼或生成网络模型。在一些实施方式中,网络建模引擎112被用来识别已经基于SRP而生成的网络。网络建模引擎112可以包括用于接收模型的更新和改变的构件。网络建模引擎112可以通过并入新的数据并生成附加的或精炼的网络模型来重复网络生成的过程。网络建模引擎112还可以促进一个或多个数据集的合并或者一个或多个网络的合并。从数据库中取得的网络集可以手动补充另加的节点、边或者全新的网络(例如,通过挖掘用于描述由特定的生物实体直接调节的附加基因的文献的文字)。这些网络含有可以允许进行过程评分的特征。网络拓扑结构被保持;因果关系的网络能够从网络中的任意点追踪到可测量的实体。此外,模型是动态的,并且用来构建它们的假定能够被修改或被重申并允许适应于不同的组织环境和种类。这允许在有新的知识可用时进行反复的测试和改进。网络建模引擎112可以去除具有低置信度或者作为与科学文献中的实验结果冲突的受试者的节点或边。网络建模引擎112还可以包括可以使用有监管或无监管的学习方法(例如,度量学习、矩阵填充、模式识别)推断出的附加的节点或边。

[0100] 在某些方面,生物系统被建模为由顶点(或节点)和用于连接节点的边构成的数学图。例如,图4示出了具有9个节点(包括节点402和404)和边(406和408)的简单网络400。节点能够表示在生物系统中的生物实体,例如,但不限于,化合物、DNA、RNA、蛋白质、肽聚糖、抗体、细胞、组织和器官。边能够表示在节点之间的关系。图中的边能够表示节点之间的各种关系。例如,边可以表示“绑定至”的关系、“用以表达”的关系、“基于表达图谱来共同调控”的关系、“抑制”的关系、“共同发生于手稿中”的关系或者“共用结构元件”的关系。一般地,这些类型的关系描述一对节点间的关系。图中的节点还能够表示节点间的关系。因而,可以表示关系之间的关系或者在关系与表示于图中的另一类型的生物实体之间的关系。例如,在表示化学品的两个节点之间的关系可以表示反应。该反应可以是处于反应与用于抑制该反应的化学品之间的关系中的节点。

[0101] 图可以是无方向性的,意指在与每个边关联的两个顶点之间没有方向。作为选择,图形的边可以从一个顶点指向另一个顶点。例如,在生物的环境中,转录调控网络和代谢网络可以被建模为有向图。在转录调控网络的图形模型中,节点将表示基因,边表示节点之间的转录关系。作为另一个示例,蛋白质-蛋白质相互作用网络描述在有机体的蛋白质组中的蛋白质之间的直接的物理相互作用,并且在此类网络中通常没有与相互作用关联的方向。因而,这些网络可以被建模为无向图。某些网络可以具有有向边和无向边。构成图形的实体和关系(即,节点和边)可以作为相关节点的网络存储于系统100中的数据库内。

[0102] 在数据库中表示的知识可以是各种不同类型的,是从各种不同来源中取得的。例如,某些数据可以表示基因组数据库,包括关于基因的信息以及它们之间的关系。在这样的示例中,节点可以表示致癌基因,而与致癌基因节点连接的另一个节点可以表示用于抑制致癌基因的基因。数据可以表示蛋白质及它们之间的关系、疾病及它们的相互关系,以及各种疾病状态。存在着能够结合于图形表示中的许多不同类型的数据。计算模型可以表示在表示在例如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、流行病学数据、化学数据集、毒理学数据集、患者数据集和人口数据集中的知识的节点之间的关系的网络。如同本文所使用的,数据集是由在所定义的条件对样品(或一组样品)的评价得出的数字值的合集。数据集能够通过例如在实验上测量样品的可量化实体来获得;或者作为选择,从服务提供者(例如,实验室、临床研究组织)处获

得,或者从公共或专有数据库中获得。数据集可以含有数据以及由节点表示的生物实体,并且在每个数据集中的节点可以与同一数据集中的其他节点 或者其他数据集中的节点相关。而且,网络建模引擎112可以生成用于将在例如DNA、RNA、蛋白质或抗体的数据集中的遗传信息表示为在医疗数据集中的医疗信息,表示为在患者数据集中的关于个体患者的信息,以及表示为在流行病学数据集中的关于整个人群的信息的计算模型。除了以上所描述的各种数据集之外,还可以存在许多别的数据集,或者可以在生成计算模型时包括的生物信息的类型。例如,数据库能够还包括医疗记录数据、结构/活动关系数据、关于传染病理的信息、关于临床试验的信息、暴露模式数据、与产品的使用历史相关的数据,以及任何其他类型的生命科学相关的信息。

[0103] 网络建模引擎112可以生成表示例如在基因之间的调节的相互作用、在蛋白质之间的相互作用或者在细胞或组织内的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模型和动态模型。网络建模引擎112可以采用任何可应用的数据方案来表示系统,例如,超图和加权二分图,在所述数学方案中使用两种类型的节点来表示反应和化合物。网络建模引擎112还可以使用其他推理技术来生成网络模型,例如,基于在表达不同的基因中的功能相关的基因的过表达的分析、贝叶斯网络分析、图形化高斯模型技术或者基因关联网络技术,以基于一组实验数据(例如,基因表达、代谢物浓度、细胞应答等)来识别相关的生物网络。

[0104] 如同以上所描述的,网络模型基于作为生物系统的功能特征的基础的机制和路径。网络建模引擎112可以生成或含有表示有关与对物剂的长期健康风险或健康效益的研究相关的生物系统的特征的结果的模型。因此,网络建模引擎112可以生成或含有用于细胞功能的各种机制的网络模型,尤其是与在生物系统中的感兴趣的特征有关的或对其有贡献的那些机制,包括(但不限于)细胞增殖、细胞应激、细胞再生、凋亡、DNA破坏/修复或炎症反应。在其他实施例中,网络建模引擎112可以含有或生成与急性全身毒性、致癌性、经皮渗透、心血管疾病、肺部疾病、生态毒性、眼部灌注/腐蚀、遗传毒性、免疫毒性、神经毒性、药代动力学、药物代谢、器官毒性、生殖和发育毒性、皮肤刺激/腐蚀或 皮肤致敏相关的计算模型。一般地,网络建模引擎112可以含有或生成用于核酸(DNA、RNA、SNP、siRNA、miRNA、RNAi)、蛋白质、肽聚糖、抗体、细胞、组织、器官和任何其他生物实体的状态以及它们各自的相互作用的计算模型。在一个示例中,计算网络模型能够用来表示免疫系统的状态以及各种类型的白细胞在免疫响应或炎症反应期间的机能。在其他示例中,计算网络模型能够用来表示心血管系统的性能以及内皮细胞的机能和代谢。

[0105] 在本公开的某些实现方式中,网络是从因果生物知识的数据库中得出的。该数据库可以通过对不同的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)而生成,这些关系中的一些可以是因果关系,并且可以与市场上可购得的数据库结合,例如,由美国马萨诸塞州剑桥市的Selventa公司管理的Genstruct技术平台(Genstruct Technology Platform)或Selventa知识库(Selventa Knowledgebase)。使用因果生物知识的数据库,网络建模引擎112可以识别用于链接扰动102与可测物104的网络。在某些实现方式中,网络建模引擎112使用来自SRP引擎110的系统响应曲线以及之前在文献中生成的网络来提取生物实体之间的因果关系。除了其他处理步骤之外,数据库还可以被进一步处理以通过在不同的生物实体集之间应用同源推理来去除逻辑不一致性并生成新的生物知识。

[0106] 在某些实现方式中,用从数据库中提取出的信息构建的网络模型基于反向因果推理(RCR)、用于处理因果关系的网络以制定机制假设并然后针对差别测量的数据集来评价那些机制假设的自动推理技术。每个机制假设将生物实体链接至它能够影响的可测量的量。例如,除了别的以外,可测量的量还能够包括浓度的增大或减小、生物实体的数量或相对多度、生物实体的激活或抑制或者生物实体的结构、功能或逻辑的改变。RCR使用作为计算的基础的在生物实体之间的在实验上观察到的因果相互作用的有向网络。有向网络可以用Biological Expression Language™(BEL™)语言(用于记录生物实体之间的相互关系的语法)来表达。RCR计算为网络模型生成指定某些约束,例如,但不限于,路径长度(连接上游节点与下游节点的边的最大数量)以及用于将上游节点连接至下游节点的可能的因果路径。RCR的输出是根据用于评价关联性和精确性的统计数据来排名的一组机制假设,这组假设表示实验测量值的差异的上游控制者。机制假设输出能够被组合到因果链和较大的网络中,以在较高的互连机制和路径层面解释数据集。

[0107] 一种类型的机制假设包括存在于表示可能原因的节点(上游节点或控制者)与表示测得量的节点(下游节点)之间的因果关系集。这种类型的机制假设能够用来进行预测,例如,如果由上游节点表示的实体的多度增大,则由因果增大关系链接的下游节点将被推断为是增大的,而由因果减小关系链接的下游节点将被推断为减小的。

[0108] 机制假设表示在一组测得的数据(例如,基因表达数据)与作为那些基因的已知控制者的生物实体之间的关系。另外,这些关系包括在上游实体与下游实体(例如,下游基因)的差别表达之间的影响的符号(正号或负号)。下游基因的机制假设可以从管理文献的因果生物知识的数据库中取出。在特定实施方式中,形式为可计算的因果网络模型的用于将上游实体链接至下游实体的机制假设的因果关系是通过网络响应评分方法来计算网络变化的基础。

[0109] 在特定实施方式中,生物实体的可评分的复杂因果网络模型能够通过搜集表示模型中的生物系统的各种特征的个体机制假设并将所有下游实体(例如,下游基因)与单个上游实体或过程的连接重新分组而转换成单个因果网络模型,由此表示整个复杂因果网络模型;这实际上是基础图结构的扁平化。在网络模型中表示的生物系统的特征和实体的改变因此可以通过将各个机制假设结合来得到。

[0110] 在某些实现方式中,系统100可以在细胞已经暴露于香烟烟雾时含有或生成用于细胞增殖机制的计算机化模型。在这样的示例中,系统100还可以含有或生成表示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,包括(但不限于),癌症、肺部疾病和心血管疾病。在某些方面,这些网络模型基于所施加的扰动(例如,暴露于物剂)、在各种条件下的响应、感兴趣的可测量的量、正在研究的结果(例如,细胞增殖、细胞应激、发炎、DNA修复)、实验数据、临床数据、流行病学数据和文献中的至少一项。

[0111] 作为说明性的示例,网络建模引擎112可以被配置用于生成细胞应激的网络模型。网络建模引擎112可以接收用于描述从文献数据库中获知的应激响应所涉及的相关机制的网络。网络建模引擎112可以基于已知的生物机制来选择一个或多个网络,以响应于在肺和心血管的环境下的应激而操作。在某些实现方式中,网络建模引擎112识别出在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来结合较小的网络而构建较大的网络模型。特别地,对于细胞应激模型,网络建模引擎112可以考虑与对氧化应激、基因毒性应激、

低氧应激、渗透、外源性应激和剪切应激的响应相关的功能单元。因此,用于细胞应激模型的网络构件可以包括外源性代谢响应、基因毒性应激、内皮剪切应激、低氧响应、渗透应激和氧化应激。网络建模引擎112还可以接收来自对在特定的细胞分组中执行的应激关联实验的公开可用的转录数据的计算分析的内容。

[0112] 当生成生物机制的网络模型时,网络建模引擎112可以包括一个或多个规则。这样的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中选出一个或多个数据集,包括体外和体内实验结果的结合。网络建模引擎112可以使用实验数据来验证在文献中所识别的节点和边。在对细胞应激进行建模的示例中,网络建模引擎112可以基于实验表示在无疾病的肺或心血管组织内的生理上相关的应激的良好程度来选择实验的数据集。数据集的选择可以基于例如表型应激端点数据的可用性、基因表达图谱实验的统计严密性,以及实验环境与正常肺或心血管无疾病的生物的关联性。

[0113] 在识别了相关网络的合集之后,网络建模引擎112还可以处理并精炼那些网络。例如,在某些实现方式中,多种生物实体及它们的连接可以被分组并且由一个或多个新节点来表示(例如,使用聚类或其他技术)。

[0114] 网络建模引擎112还可以包括与所识别的网络中的节点和边有关的描述性信息。如上所述,节点可以通过其关联的生物实体、关于关联的生物实体是否是可测量的量的指示,或者生物实体的任何其他描述符来描述,而边可以通过例如它表示的关系的类型(例如,因果关系(例如,上调或下调)、相关性、条件相关或独立)、该关系的强度或者在该关系中的统计置信度来描述。在某些实现方式中,对于每个处理,表示可测量的实体的每个节点与响应于该处理的活动变化的预期方向(即,增大或减小)关联。例如,当支气管上皮细胞暴露于诸如肿瘤坏死因子(TNF)之类的物剂时,特定基因的活动可以增加。这种增加可以由于可从文献中获知的直接调节关系(并且以由网络建模引擎112识别出的一个网络来表示)而出现,或者通过经由网络建模引擎112所识别的一个或多个网络的边来跟踪众多的调节关系(例如,自分泌信令)而出现。在某些情形中,网络建模引擎112可以识别每个可测量实体响应于特定的扰动而改变的预期方向。当在网络中的不同路径指示特定实体的改变的相反的预期方向时,这两个路径可以被更详细地检查,以确定净改变方向,或者该特定实体的测量值可以被丢弃。计算网络模型可以由系统100产生、导入到系统100中或者在系统100内识别(例如,来自于生物知识的数据库)。

[0115] 返回图2,在步骤216,网络评分引擎114使用由网络建模引擎112在步骤214识别的网络以及由SRP引擎110在步骤212以SRP的形式生成并接收的数据来生成来为每个扰动生成网络响应得分。网络响应得分在生物实体之间的基础关系(由所识别的网络表示)的背景下对处理的生物响应(由SRP表示)进行量化。这些网络响应得分可以数字地或图形地呈现例如通过暴露到潜在有害的物剂来扰动生物系统的效果。通过提供对于处理的网络响应的测量结果,这些网络响应得分可以允许分子事件(由实验数据所测量的)与表征细胞、组织或器官水平的网络的表现型相关联。网络评分引擎114可以包括用于生成包含于网络建模引擎112内的或者由其识别的每个网络的网络响应得分的硬件和软件构件。

[0116] 网络评分引擎114可以被配置用于实现生成表示网络对于处理的响应的整体强度的标量值得分的上述技术,诸如强度评分技术。强度得分是对于在SRP中表示的不同实体

的活动观察值的均值。在一些实施方式中,网络响应的强度按照下式来计算:

$$[0117] \quad \text{强度} = \frac{\sum_i d_i \beta_i}{\text{NumMeasNodes}}$$

[0118] 其中 d_i 表示与节点 i 关联的实体的活动改变的预期方向, β_i 表示在处理与控制条件之间的活动的倍数变化的对数(即,用于描述从初始值变为最终值的数量变化程度的数字),并且NumMeasNodes是具有关联的已测量的生物实体的节点的数量。正强度得分指示SRP与从所识别的网络中得出的预期活动变化匹配,而负强度得分指示SRP与预期活动变化不匹配。

[0119] 除了或者代替上述标量值网络得分,网络评分引擎114可以生成向量值得分。用于计算表示网络响应的值的方法的示例在2011年8月19日递交的美国专利临时申请no.61/525,700中描述了,例如,集合扰动指数(GPI)、概率扰动指数(PPI)和预期扰动指数(EPI),通过引用将该文献全部结合在这里。一个向量值得分是每个被测量节点的活性的倍数变化或绝对变化的向量。如上所述,倍数变化是用于描述在不同的条件下(例如,在扰动与控制条件之间)可测量变化从初始值变为最终值的程度的数字。该倍数变化向量具有 N 个分量,对应于在具有关联的测量的生物实体的网络中的节点的数量。在某些实现方式中,几何扰动指数(GPI)值被用在本公开的方法中。在计算GPI的一些实施方式中,倍数变化向量的第 i 分量 β_i 表示第 i 测量生物实体的活动在扰动与控制条件之间的倍数变化的对数(例如,底为2)(即,实体的活动在两个条件之间变化的因子的对数)。在其他实施方式中,其他线性或非线性函数可以替换算法函数。权重向量 r 也被包括在计算GPI中。权重向量 r 也具有 N 个成分,其中每个加权向量 r 的每个分量 r_i 表示待赋予观察到的第 i 倍数变化 β_i 的权重。在某些实现方式中,权重表示第 i 测量实体关于感兴趣的特征或结果(例如,在癌症研究中的已知致癌物质)的已知生物学意义。可以有利地用来加权的一个值是本地错误未发现率 $fndr_i$ (即,倍数变化值 β_i 表示违反关于零倍数变化的基础原假设的概率,在某些情形中,在所观察到的 p 值的条件下),如同由Strimmer等人在“A general modular framework for gene set enrichment analysis”(BMC Bioinformatics10:47,2009)中以及由Strimmer在“A unified approach to false discovery rate estimation”(BMC Bioinformatics9:303,2008)中所描述的,这两篇文章均通过引用的方式全文并入本文。用于计算GPI的另一个加权因子是每个成分在倍数变化向量 β 上的期望方向。网络评分引擎可以通过询问网络建模引擎以检索因果生物网络模型的变化预期方向来生成这个值。网络评分引擎然后能够将这些预期的变化方向结合到 N 个分量的向量 d 中,其中向量 d 的第 i 分量 d_i 表示第 i 测量生物实体的预期的变化方向(例如,+1用于增大的活动,而-1用于减小的活动)。在某些实现方式中,该组合是算术组合,其中所缩放的倍数变化 $r_i \beta_i$ 每个都乘以其相应的预期变化方向 d_i ,并且结果对于所有 N 个生物实体进行叠加。在算术上,这种实现方式能够由下式表示:

$$[0120] \quad \sum_i d_i r_i \beta_i$$

[0121] 在其他实现方式中,向量 d 、 r 和 β 可以按照任何线性的或非线性的方式结合。这种结合是通过乘以预定的缩放因子来标准化。一个这样的缩放因子是生物实体数 N 的平方根。

在该实现方式中,GPI得分能够由下式表示:

$$[0122] \quad GPI = \frac{\sum_i d_i r_i \beta_i}{\sqrt{N}}$$

[0123] 在某些实施方式中,概率扰动指数(PPI)值被用在本公开的方法中。PPI是通过将正活化度量PPI⁺和负活化度量PPI⁻结合来计算的,例如通过以下公示:

$$[0124] \quad PPI = \frac{1}{2} (PPI^+ + PPI^-)$$

[0125] 活化度量是基于在给定所观察的系统响应曲线的状态下,由感兴趣的网络表示的生物机制被活化的可能性的量化。对于计算GPI,倍数变化向量 β 被收集并且倍数变化强度被生成为具有表示在处理条件下倍数变化值能够于生物系统中采用的值的集合的近似的范围,并且可以由范围 $[-W, W]$ 来近似,其中 W 是 $\log_2$ 倍数变化的理论的预期最大绝对值。正活化度量表示SRP指示所观察到的生物实体的活化/抑制与由 d_i 表示的预期变化方向一致的程度。网络与SRP一致的行为在本文中称为“正活化(Positively Activated)”,并且可以使用的一个正活化度量是一个或多个网络被正向活化的概率,即PPI⁺,其可以按照以下表达式来计算:

$$[0126] \quad PPI^+ = \Pr(\text{Positively Activated}) = \frac{1}{W} \int_0^W \Pr(\text{Positively Activated} | \varphi) d\varphi$$

[0127] 其中:

$$[0128] \quad \Pr(\text{Positively Activated} | \varphi) = \frac{1}{N} \sum_{0 < d_i \beta_i < \varphi} fndr_i$$

[0129] 其中 $fndr_i$ 是以上讨论的错误为发现率。按照下式来计算正活化度量PPI⁺的近似值:

$$[0130] \quad PPI^+ \approx \frac{1}{WN} \sum_{0 < d_i \beta_i} fndr_i d_i \beta_i$$

[0131] 不一致的行为在这里被称作为负活化(Negatively Activated)。可以使用的一个负活化度量是一个或多个网络被负活化的概率。这样的概率(称为PPI⁻)可以按照以下表达式来计算:

$$[0132] \quad PPI^- = \Pr(\text{Negatively Activated}) = \frac{1}{W} \int_{-W}^0 \Pr(\text{Negatively Activated} | \varphi) d\varphi$$

[0133] 其中

$$[0134] \quad \Pr(\text{Negatively Activated} | \varphi) = \frac{1}{N} \sum_{\varphi < d_i \beta_i < 0} f n d r_i$$

[0135] 可以根据以下公示计算负活化量度PPI的近似:

$$[0136] \quad PPI^- \approx \frac{1}{WN} \sum_{d_i \beta_i < 0} f n d r_i d_i \beta_i$$

[0137] 另一种计算表示网络响应的值的方法是预期扰动指数(EPI)评分技术。对于每个SRP表示测量生物实体在处理条件下的活动(或活动变化),那么每个SRP与所测得的活动的数量关联,每个测量的生物实体一个。EPI是由SRP表示的在所有生物实体上的平均活动变化的量化。一般地,以SRP表示的测得活动可以从测得活动的分布中随机取得,EPI表示该分布的期望值。如果每个倍数变化 β_i 都是从分布 $p(\cdot)$ 中取得的,则该分布的期望值是

$$[0138] \quad EPI = \int \varphi \cdot p(\varphi) \cdot d\varphi$$

[0139] 由于真实的理论分布 $p(\cdot)$ 不容易知道,因而EPI值可以通过使用所观察的活化来尽速并且生成倍数变化密度。如果从分布 $p(\cdot)$ 中取得的每个倍数变化 β_i ,则分布 $p(\cdot)$ 能够是由下式近似表示:

$$[0140] \quad \hat{p}(\varphi) \propto \begin{cases} \frac{1}{N} \sum_{\varphi < d_i \beta_i < \varphi} \frac{|\beta_i|}{W}, & \varphi > \varepsilon \\ \frac{1}{N} \sum_{\varphi < d_i \beta_i < \varphi} \frac{|\beta_i|}{W}, & \varphi < \varepsilon \end{cases}$$

[0141] 在某些实现方式中,网络评分引擎114应用计算内插技术(例如,线性或非线性的内插技术)来根据以上公式的分布生成近似的连续分布,然后计算该分布的期望值。在其他实现方式中,网络评分引擎被配置用于将离散分布用作对连续分布的矩形近似,并且按照下式来计算EPI:

$$[0142] \quad EPI \approx \frac{1}{WN} \left[\sum_{d_i \beta_i > 0} (d\beta)_{(i)} \left(\sum_{j=1}^{n^+} (d\beta)_{(j)} \right) ((d\beta)_{(i)} - (d\beta)_{(i-1)}) - \sum_{d_i \beta_i < 0} -(d\beta)_{(i)} \left(\sum_{j=1}^{n^-} -(d\beta)_{(j)} \right) (-(d\beta)_{(i)} - (-(d\beta)_{(i-1)})) \right]$$

[0143] 在该公式中,下标 $(\cdot)$ 表示按照从最小倍数变化到最大倍数变化的顺序取得的值, n^+ 是预期其活动将响应于处理($d_i \beta_i > 0$)而增大的实体的数量,而 n^- 是预期其活动将响应于处理($d_i \beta_i < 0$)而减小的实体的数量。在EPI得分中,值较高的倍数变化会比值较低的倍数变化受到更多的考虑,从而提供具有高特异性的活动的测量结果。

[0144] 在特定实施方式中,对于每个扰动(例如,暴露到已知或未知物剂),网络评分单元114可以生成多个网络响应得分,其构成用于相应的扰动或相应处理的得分集。例如,网络评分引擎114可以生成对于具体网络、具体物剂剂量和具体暴露时间的网络响应得分。全部这些网络响应得分的集合被发送到聚合引擎116。

[0145] 在步骤218,聚合引擎116基于在步骤216由聚合引擎116生成的多个网络响应得分来生成生物响应因子(BIF)。聚合引擎116可以仅使用可以从一个或多个网络得出的其他补充信息来生成BIF。在特定实施方式中,聚合引擎116可以从对应于不同生物网络的SRP来直接生成BIF。在特定实施方式中,BIF值可以被用来比较暴露到不同处理的预测生物结果,其中不同结果可以由各个处理条件导致的不同机制来引起。在特定实施方式中,BIF可以被当做扰动对多个可能受到疾病袭击或生物结果影响的生物网络的影响的聚集测量结果。用于生成BIF的大量图形理论计算技术已经被研究出来了,其中的任何技术都可以由聚合引擎116来执行;这些技术的示例在下文中讨论。在特定实施方式中,得分是向量值得分。在特定实施方式中,得分不是标量值得分。在特定实施方式中,一个或多个生物影响因子由第一和第二得分集的聚集得分的线性组合、线性变换或二次方程形式来确定。将由网络建模引擎112提供的M个计算模型表示为Net-1、Net-2、...、Net-M,其中M大于 等于1。为了生成BIF,聚合引擎116可以使用图形统计技术,该技术利用一些或全部网络模型的统计或特征,诸如完全网络结构、节点数目、边数目、节点或边的权重(如果加权的话)、节点或边的任何其他特性(例如,与生物实体的测量结果以及分别由节点和边表示的关系相关的统计置信度)、在不同网络模型中重复的任何节点或边、网络模型的结构中的置信度(例如,网络结构如何在文献中一致地复制的测量结果)、或者由网络建模引擎112提供的网络模型表示的任何其他数据。一些这些数据可以从由SRP引擎110执行的计算来获得(例如,对于测量结果的统计置信度估计),并且可以被经由网络建模引擎112传送到聚合引擎116,或者从SRP引擎110直接传送到聚合引擎116。

[0146] 对于每个处理和每个网络模型Net-i,聚合引擎116也从网络评分引擎114接收一个或多个网络响应得分的向量 S_i 。如上所述, S_i 可以包括表示Net-i对于物剂扰动的响应的整体强度的一个或多个标量值得分; S_i 也可以包括表示Net-i对于物剂扰动的响应的拓扑分布的一个或多个向量值得分。与不同网络模型Net-i和Net-j分别相关联的网络响应得分向量 S_i 和 S_j 不需要具有相同维度,也不需要基于任何相同的网络响应得分生成技术。

[0147] 在特定实施方式中,聚合引擎116使用来自于网络建模引擎112的数据和来自于网络评分引擎114的网络响应得分来生成生物影响因子(BIF)。图5示出了可以包括在聚合引擎116中的四个模块:过滤模块510、网络加权模块512、聚合模块514和相对评分模块516。这些模块510、512、514、516中的一个或多个可以实施在硬件和软件至少一者上,如参照图11和图12所描述的。

[0148] 聚合引擎116可以被构造为生成与图6中描绘的示意性图形理论处理600。处理600的步骤现在将会被描述为由聚合引擎116的模块510-516(图5)来执行,但是将会明白这些步骤可以由任何合适的顺序执行并在一个或多个处理组件之间分割。

[0149] 在步骤602处,聚合引擎116从网络建模引擎112接收关于计算网络模块的信息以及从网络评分引擎114接收网络响应得分。在步骤604 处,过滤模块510过滤得分向量 S_1 、 S_2 、...、 S_M 。在一些实施方式中,在步骤604处执行的过滤操作包括将一个或多个得分向量的

一个或多个成分标准化。例如,如果每个得分向量的第一成分是表示相关网络模型的响应的整体强度的标量值得分,那么这些第一成分可以由适当的值标准化,使得得分全部落入期望范围内。合适的标准化值的一个选择是全部得分向量中第一成分的最大值;如果全部第一成分值是非负的,则将每个第一成分除以该最大值将会将第一成分先知道范围[0,1]内。在一些实施方式中,在步骤602执行的过滤操作包括移除异常值。当得分向量的成分值距指定值(例如,平均值、中值、模值)超出指定量(例如,一定量的标准差)时,该得分向量的成分可以被认为是异常值。指定量和值可以是已知先验的,或者可以基于网络响应得分向量 $S_1, S_2, \dots, S_m$ 的组合来计算。

[0150] 在一些实施方式中,在步骤604处执行的过滤操作包括几何图形技术。一个这种分解技术将得分向量 $S_1, S_2, \dots, S_m$ 的一个或多个部分分解。为了示意性目的,整体向量 S_i 的分解在以下描述中讨论,但是分解可以仅在得分向量的特定成分上执行。在分解中,向量 S_i 被写入为两个以上的向量的组合。图7示出了将向量702分解为两个成分708和710。如本领域中已知的,如果 S_i 具有维度 p ,则 S_i 可以被写作 p 个不同基础向量的线性组合,这些向量可以跨越其中嵌入 S_i 的 p 维向量空间,数学表达式为:

$$[0151] \quad S_i = a_1 v_1 + \dots + a_p v_p$$

[0152] 其中 $\{v_1, \dots, v_p\}$ 是向量的跨度集合,并且 $\{a_1, \dots, a_p\}$ 是相应的标量系数。向量 $a_1 v_1$ 被称作为 S_i 在 v_1 上的投影。在图7中,向量704和706是基础向量,并且向量702在每个基础向量上的投影分别是向量708和710。在不损失一般性的情况下, $\{v_1, \dots, v_p\}$ 被假设为正交基础。标量系数的值可以通过计算 S_i 与相应向量之间的内积来计算。

[0153] 聚合引擎116可以被配置为选择大量基础向量 $\{v_1, \dots, v_p\}$ 中的任何一者(或者由其预编程)。在一些实施方式中,使用网络模型Net- i 的结构 例如使用谱图计算技术基础向量来确定。通常,谱技术使用从表示网络模型的矩阵的本征分析来得到的技术。在一个具体谱技术中,基础向量 $\{v_1, \dots, v_p\}$ 可以是与网络模型Net- i 相关联的组合拉普拉斯算子矩阵的本征向量。如果Net- i 表示具有 n_i 节点的无向网络,那么组合拉普拉斯算子被计算为:

$$[0154] \quad L = D - A$$

[0155] 其中 D 是 n_i 乘 n_i 的对角矩阵,在对角线上是每个节点Net- i 的度数,并且 A 是Net- i 的节点-节点的 n_i 乘 n_i 的邻接矩阵。其本征向量可以在步骤604提供用于分解的合适基础的其他矩阵包括节点-节点邻接矩阵、节点-边邻接矩阵、标准化的拉普拉斯算子矩阵、格拉姆矩阵或者表示Net- i 的结构的其他任何矩阵。

[0156] 因此,在一个实施方式中,第一得分集和第二得分集内的每个得分都包括得分向量,并且生成生物影响因子的步骤还包括在处理器处过滤第一得分和第二得分以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影。过滤还可以包括从经分解的第一和第二得分中的至少一者移除多个投影中的至少一个。基础向量集可以包括表示至少一个模型的矩阵的本征向量。在一些实施方式中, S_i 的分解可以被用来使用例如几何过滤技术或几何图形计算技术来调整 S_i 的值。在特定方面,使用几何图形过滤技术包括调整一个或多个网络模型的图形表示,诸如向量、网眼(mesh)或更高纬度表示。两个这种示例分别在图8A和图8B中示出。在第一示例中, S_i 向一些基础向量的投影可以被从 S_i 减去(“减小 S_i 维度”)。这在图8A中示出:向量702被分解为向量708和710,并且过滤模块510从向量702移除向量708,保留向量806。被移除的投影可以是幅度(例如,长度)最小的那些。当基础向量生

成为具体矩阵的本征向量时,被移除的投影可以是与本征值具有最小幅度本征向量相关联的那些投影。固定数目的投影可以被移除或保持。代替或者附加于减小 S_i 的维度, S_i 的每个投影可以被分别缩放,之后经缩放的投影被加到一起,以形成新的得分向量 S_i 。这在图8B中示出:向量702被分解为向量708和710,并且过滤模块510缩放向量708以形成新的向量812,并且缩放向量710以形成新的向量814。每个投影的缩放向量可以以多种方式选择,包括基于每个投影的相对重要性(significance)的经验观察或数学建模。在一些实施方式中,使用图谱信息。例如,当基础向量生成为具体矩阵的本征向量时,每个投影的缩放因子可以基于与本征向量相关的本征值。例如, S_i 投影到向量 v_j 上的缩放值可以由以下式子给出:

$$[0157] \quad e^{-\lambda_j t}$$

[0158] 其中, λ_j 是与本征向量 v_j 相关的本征值。参数 t 是可调谐的,使得更大的值导致更小的缩放投影。

[0159] 返回图6,在步骤604,网络加权模块512可以对于每个 M 计算模型相关联的网络响应得分向量 $S_1, S_2, \dots, S_m$ 中的每个进行加权。生成第一得分集和第二得分集可以包括:在处理器处基于相应的计算网络模型和第一数据集和第二数据集中的至少一个来对第一得分集和第二得分集中的每一个分配权重;聚合第一得分集中经加权的得分;聚合第二得分集中经加权的得分;其中,一个或多个生物影响因子是第一得分集和第二得分集的经聚合的得分的函数。对第一得分集和第二得分集中的每一个分配权重包括选择对于多个计算模型中每一个的权重,以将第一得分集内的得分和第二得分集内的得分之间的差异最大化。这种加权可以基于在步骤210接收的数据(图2)并且基于相应的网络模型。在一些实施方式中,步骤604包括图形优化计算技术。在一个这种实施方式中,与每个得分向量相关的权重被选择,以基于表示对于生物系统相对“弱”的扰动的处理条件以及基于表示对于生物系统相对“强”的扰动的处理条件,将得分向量之间的差异最大化。图9示出了示例,但是不局限本发明应用的范围。图9描绘了其中生物系统被暴露到有毒物剂持续三种不同暴露时间的处理条件图900:短902、中904和长906。对于每个暴露时间,SRP引擎110收集表示生物实体集的所测量的活动的的数据。网络建模引擎112相对于有毒物剂和生物系统(包括所测量的生物实体)识别三个不同网络Net-1908、Net-2910、Net-3912,并且网络评分引擎114计算三个网络和三个暴露时间中每一个的标量值网络响应得分。网络加权模块512之后分别对于三个网Net-1908、Net-2910、Net-3912选择权重集 C_1, C_2 和 C_3 ,使得在使用相同权重的情况下,短暴露网络响应得分的加权总和与长暴露网络响应得分的加权总和的差异最大。权重 C_1, C_2 和 C_3 可以以一些方式来约束(例如, C_1, C_2 和 C_3 必须是非负的并且总和为1)。换言之,网络加权模块512执行以下优化方法(使用已知的计算优化方法):

$$[0160] \quad \max \left[c_1 \left\| S_{\text{长}}^1 - S_{\text{短}}^1 \right\|^2 + c_2 \left\| S_{\text{长}}^2 - S_{\text{短}}^2 \right\|^2 + c_3 \left\| S_{\text{长}}^3 - S_{\text{短}}^3 \right\|^2 \right]$$

$$[0161] \quad \text{s. t. } c_1, c_2, c_3 \geq 0$$

$$[0162] \quad c_1 + c_2 + c_3 = 1$$

[0163] 在步骤604对网络响应得分向量进行加权之后,聚合模块514在步骤606对分别用

于每个处理条件的网络响应得分向量相加。这些向量可以已经受到过滤模块510的过滤、受到网络加权模块512的加权、受到这两者或者受到任意一者。在一些实施方式中,步骤606包括将用于具体处理条件的全部网络响应得分向量联系成为单一向量。令ASV-i表示对于处理i的经聚合的得分向量。

[0164] 之后,对于第二处理条件重复步骤602-606(在图6中表示为步骤608-610)。这些步骤可以对于感兴趣的许多附加处理条件重复,但是如这里所公开的,在一些实施方式中,仅研究了两个处理条件。这些处理条件之一可以包括暴露到其长期生物影响被合理理解的物剂(诸如来自于标准烟草香烟的烟),而第二处理条件可以包括暴露到其长期生物影响还没有被良好了解的物剂(诸如来自于烟草相关制品的烟雾剂或蒸气)。无论研究多少处理条件,在步骤606的结论中,对于每个处理条件i生成经聚集的得分向量ASV-i。

[0165] 在步骤608处,相对评分模块516根据经聚合的得分向量生成BIF。在一些实施方式中,相对评分模块516将这些经聚合的彼此比较,以生成一个或多个BIF。如上文所讨论的,BIF可以表示不同的扰动之间哪些生物途径被类似地活化,这可以允许基于其他扰动的长期影响来进行关于一个扰动的长期影响的预测。这里讨论了BIF的大量优点和用途。相对评分模块516可以以许多方式从ASV的集合生成BIF。在一些实施方式中,步骤608包括几何图形技术。例如,BIF可以由计算两个ASV之间的内积并使用与内积相关的角度作为BIF测量结果来生成。在这种实施方式中,更小的BIF数目表示由两个处理条件活化的生物机制之间更大的一致性,从而提示根据这些机制的长期结果的类似性。任何数目的内核可以被用于内积计算,包括对角实体中具有各种锁放因子的单位矩阵或对角矩阵。一些这种实施方式包括图谱信息。例如,相对评分模块516可以对于内积计算使用块对角矩阵内核,其中第i个块根据以下公式计算:

$$[0166] \quad \sum_j e^{-2t\lambda_j} v_j v_j^T$$

[0167] 其中, v_j 是与Net-i的拉普拉斯算子矩阵相关联的第j个本征向量,并且 λ_j 是相关联的第j个本征值。使用该内核来计算原始得分向量 S_1 、 S_2 、...、 S_M 之间的内积是聚合引擎116如上所述参照以上公式实施本征向量分解和指数缩放技术的一种替换方式。

[0168] 在一些实施方式中,每个ASV被用来限定(可能具有多个维度)的表面,并且BIF由比较这些表面来生成。因此,生成生物影响因子可以包括确定由至少一个第一向量限定的至少一个第一表面与由至少一个第二向量限定的至少一个地儿表面之间的距离,第一向量表示至少一个第一得分集的经聚合的得分,第二向量表示至少一个第二得分集的经聚合的得分。其中,这种实施方式可以包括几何和优化技术。这种方法以图10中的简单示例示出,其为描绘了与第一处理条件对应的表面1002和与第二处理条件对应的表面1004的图示1000。这些表面被限定在剂量-暴露时间空间内(剂量轴1008和时间轴1010),并且每个表面在具体剂量和暴露时间下的高度等于标量网络响应得分1006的值(或者向量值得分或多个不同得分的标量值聚合)。BIF可以通过以许多方式进行的表面比较框架来生成。在一些实施方式中,相对评分模块516识别两个表面彼此最接近的剂量和时间。网络响应得分在该点的差异(即,表面的高度的差异)表示在相同剂量-时间条件下由一个扰动活化的生物机制与由第二扰动活化的生物机制最接近的条件。在一个示例中,当第一扰动是暴露到已知

的有毒物质并且第二扰动暴露到未知物质,该最小距离比较表示对于未知物质的生物响应可能与已知毒素的生物响应类似的“最差情况”。这种最差情况可能对于研究和公共健康目的是重要的。在一些实施方式中,相对评分模块516识别两个表面彼此最远离的剂量和时间。当研究药品或原理的有益特性时,这种实施方式可以是有用的,因为最大差异点可以示出与已知的有效药物相比新药的功效的“最差情况”。在一些实施方式中,相对评分模块识别第一表面与第二表面的任何值最接近的值,而不考虑这些点是否对应于相同的剂量-暴露时间条件。识别这些最接近的点可以使得能够做出两个扰动之间的有益比较;例如,由吸传统香烟持续特定时间段引起的扰动的效果类似于由吸入来自于关于烟草的产品的烟雾剂或蒸气持续不同时间段所产生的扰动的效果。

[0169] 相对评分模块516可以以许多不同方式表示相对得分。在一些实施方式中,相对评分模块可以输出总结前述实验和分析的标量值BIF。例如,如果相对评分模块将图10的两个表面相比较,找出两个表面的值最相似的点,并且识别用于第一处理的响应剂量和暴露时间(分别为剂量1和时间1)以及用于第二处理的响应剂量和暴露时间(分别为剂量2和时间2),那么标量值BIF可以根据以下公式计算:

$$[0170] \quad BIF = \frac{\text{剂量1}}{\text{剂量2}} + \frac{\text{时间1}}{\text{时间2}}$$

[0171] 在前述示例中,BIF是被描述为与生物机制的扰动相关。该BIF值在某些方面被具体描述为量化所选择的扰动对各个生物机制的长期结果的数值。然而,系统100不局限于识别具体扰动的BIF,并且作为代替还可以被用来生成数个不同扰动的BIF值以及用来预测一个或多个这些扰动的数个不同的长期结果。

[0172] 对于扰动和结果附加地和可选地,系统100可以被用来对于包括疾病结果、疾病进展、生物机制和环境条件的一个或多个其他参数的一个或多个BIF值。例如,多个BIF值可以被生成,每个值表示肺癌进展的不同水平——早期、中期和晚期。系统100可以包括用于生成和存储用于这些不同参数的多个BIF值的硬件和软件组件。例如,系统100可以包括用于存储与肺癌进展相关的不同BIF值的数据库和存储装置。该数据库中的每个实体可以包括表示疾病(即,肺癌)进展的不同阶段的不同BIF值。该数据库中的实体可以包括与BIF相关联的附加信息,诸如相关生物机制和生物实体的列表。数据库可以被用于不同目的,例如,临床诊断和预诊断。

[0173] 在用于临床分析的一个示例中,系统100可以被用来研究病人的肺癌的进展。系统100可以包括表示特定疾病(例如但不限于肺癌)的进展的不同阶段的BIF值的数据库。在这种示例中,病人可能已经暴露到具有未知来源或未知身份的物质。病人可以告知临床医生他们曾暴露到可能是他们怀疑可能潜在影响他们健康(特别是肺部健康)的颗粒和气体物质的混合物的物质。临床医生可以选择一个或多个化验来从病人获得生物样本并且产生病人的可测量数据。在特定实施方式中,系统100可以辅助选择化验。例如,在临床医生对于告知肺癌的进展的化验发出请求时,系统100可以向临床医生显示一个或多个推荐化验的列表。由一个或多个化验获得的病人的数据可以被输入到系统100中用来计算。基于该数据,系统100可以查询数据库来获得具有类似实验结果的实体。例如,对于基因表达化验,系统100可以查询数据库来识别基因或基因表达水平与从病人的数据获得的那些匹配的实

体。在特定实施方式中，系统100可以基于可能不应用到病人的其他属性来过滤数据库中的一个或多个实体。系统100可以确定与所选择的数据库实体相对应的一个或多个BIF值并且将这些一个或多个BIF值归属于该病人。可选择地，系统100可以使用病人的数据来计算该病人独有并且可以被用来与数据库中表示特定生物结果的BIF值相比较的BIF。例如，数据库可以包括范围从0到100的BIF值，每个值表示肺癌进展的水平。在这种示例中，更接近零的数字可以表示更早的肺癌阶段，而更接近100的数字可以表示晚期阶段。系统100可以确定病人的数据生成在范围10-20内的BIF值，并且输出该结果用于显示。临床医生或系统100可以解释该结果并且通知病人它们暴露到可能有害的物质并且可能呈现特定的肺癌阶段。系统100可以包括合适的硬件和软件组件来接收数据并且生成并输出BIF值。

[0174] 图11是用于量化生物扰动的影响分布式计算机化系统1100的框图。系统1100的构件与图1的系统100中的那些构件是相同的，但是系统100的布局是这样的：使得每个构件通过网络接口1110来通信。这样的实现方式可能适合于经由包括无线通信系统的多种通信系统进行的分布式计算，该无线通信系统可以共用对共同网络资源的访问，例如，“云计算”范式。

[0175] 图12是计算装置的框图，例如，图1的系统100或者用于执行参照图1-图10所描述的过程的图11的系统1100的任意构件。包括SRP引擎110、网络建模引擎112、网络评分引擎114、聚合引擎116以及一个或多个数据库（包括结果数据库、扰动数据库和文献数据库）的系统100的每个构件都可以在一个或多个计算装置1200上实现。在某些方面，多个上述构件和数据库可以包括于一个计算装置1200内。在某些实现方式中，构件和数据库可以跨若干个计算装置1200而实现。

[0176] 计算装置1200包括至少一个通信接口单元、输入/输出控制器1210、系统存储器和一个或多个数据存储装置。系统存储器包括至少一个随机存取存储器（RAM1202）和至少一个只读存储器（ROM1204）。所有这些元件都与中央处理单元（CPU1206）通信以促进计算装置1200的操作。计算装置1200可以按照许多不同的方式来配置。例如，计算装置1200可以是常规的独立式计算机，或者作为选择，计算装置1200的功能可以跨多个计算机系统和体系架构分布。计算装置1200可以被配置用于执行建模、评分和聚合操作中的一些或全部。在图10中，计算装置1200经由网络或本地网络链接至其他服务器或系统。

[0177] 计算装置1200可以按照分布式体系架构来配置，其中数据库和处理器被安放于分离的单元或位置。某些此类单元执行初级处理功能并且最低程度地含有通用控制器或处理器和系统存储器。在这方面，这些单元每个都经由通信接口单元1208连结至用作与其他服务器、客户端或用户计算机及其他相关装置间的初级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身可以具有最小限度的处理能力，主要用作通信路由器。多种通信协议可以作为系统的一部分，包括，但不限于：以太网（Ethernet）、SAP、SASTM、ATP、BLUETOOTHTM、GSM和TCP/IP。

[0178] CPU1206包括处理器，例如，一个或多个常规的微处理器和一个或多个辅助协处理器，例如，用于转移CPU1206的工作负载的数学协处理器。CPU1206与通信接口单元1208和输入/输出控制器1210通信，由此CPU1206与诸如其他服务器、用户终端或装置之类的其他装置通信。通信接口单元1208和输入/输出控制器1210可以包括用于与例如其他处理器、服务器或客户端终端同步通信的多种通信通道。与每个其他通信的装置不需要持续地相互发送

信号。相反地,这样的装置只需要在必要时彼此发送信号,可以实际上大部分时间实际都避免交换数据,并且可以需要执行几个步骤来建立装置之间的通信链路。

[0179] CPU1206同样与数据存储装置通信。数据存储装置可以包括磁存储器、光存储器或半导体存储器的适当组合,并且可以包括例如RAM1202、ROM1204、闪存驱动器、光盘(例如,压缩盘)或者硬盘或硬盘驱动器。例如,CPU1206和数据存储装置每个都可以完全位于单个计算机或其他计算装置之内;或者通过通信介质(例如,USB端口、串口线、同轴线、以太网型网线、电话线、射频收发器或者其他类似的无线或有线介质,或者上述限的组合)彼此连接。例如,CPU1206可以经由通信接口单元1208与数据存储装置连接。CPU1206可以被配置用于执行一个或多个特定的处理功能。

[0180] 数据存储装置可以存储例如(i)用于计算装置1200的操作系统1212;(ii)适用于根据本文所描述的系统和方法并且尤其是根据针对CPU1206详细描述的过程来引导CPU1206的一个或多个应用1214(例如,计算机程序代码或计算机程序产品);或者(iii)适用于存储可以用来存储程序所需的信息的信息的数据库1216。在某些方面,数据库包括用于存储实验数据的以及发表的文献模型的数据库。

[0181] 操作系统1212和应用1214可以按照例如压缩的、不压缩的和加密的格式来存储,并且可以包括计算机程序代码。程序的指令可以从数据存储装置以外的计算机可读介质(例如,ROM1204或RAM1202)读入处理器的主存储器内。虽然在程序中的指令序列的执行促使CPU1206执行本文所描述的过程步骤,但是也可以使用硬连线电路来代替用于实现本发明的过程的软件指令或者与其结合。因而,所描述的系统和方法并不限于硬件和软件的任何具体结合。

[0182] 合适的计算机程序代码可以被提供用于执行与本文所描述的建模、评分和聚合相关的一个或多个功能。程序同样可以包括程序单元,例如,操作系统1212、数据库管理系统以及允许处理器经由输入/输出控制器1210与计算机外围装置(例如,视频显示器、键盘、计算机鼠标等)接口连接的“装置驱动器”。

[0183] 本文所使用的术语“计算机可读介质”指的是用于给计算装置1200的处理器(或者本文所描述的装置的任何其他处理器)提供或参与提供用于执行的指令的任何非临时性介质。这样的介质可能要采取许多形式,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盘、磁盘或光磁盘,或者集成电路的存储器,例如,闪存存储器。易失性介质包括通常构成主存储器的动态随机存取存储器(DRAM)。计算机可读介质的共同形式包括,例如,软盘、柔性盘、硬盘、磁带、任何别的磁介质、CD-ROM、DVD、任何别的光介质、打孔卡、纸带、具有孔图形的任何别的物理介质、RAM、PROM、EPROM或EEPROM(电可擦可编程只读存储器)、FLASH-EEPROM、任何别的存储器芯片或盒,或者计算机能够读取的任何别的非临时性介质。

[0184] 各种形式的计算机可读介质可以涉及将一个或多个指令的一个或多个序列传递给CPU1206(或者本文所描述的装置的任何其他处理器),以用于执行。例如,指令最初可以位于远程计算机(未示出)的磁盘上。远程计算机能够将指令装载到其动态存储器内,以及经由以太网连接、电线或甚至是使用调制解调器的电话线来发送指令。位于计算装置1200(例如,服务器)本地的通信装置能够接收在各自的通信线路上的数据,并且将数据放置于处理器的系统总线上。系统总线将数据输送到主存储器,处理器从该主存储器中检索出指

令并执行。由主存储器接收到的指令可以在由处理器执行之前或之后任选地存储于存储器内。另外,指令可以经由通信端口作为电信号、电磁信号或光信号来接收,这些信号是用于载送各种类型的信息的无线通信或数据流的示例性形式。

[0185] 如上文所讨论的,系统100可以被用来构建生物机制的网络来进一步在系统水平评价扰动的生物影响。以下的段落描述了数个示例网络,其中每一者可以被用来计算与基础机制相关的不同结果的BIF得分。

[0186] 作为第一示例,系统100可以被用来构建用于细胞增殖的关注肺的网络。该关注肺的细胞增殖网络是使用生物表达语言(BEL)构建的,由Selventa(剑桥,马塞诸塞,美国)研究的用于生物途径表示的可计算框架,使其能够应用到基于由高吞吐量装置获得的数据的细胞增殖的评估。细胞增殖网络包括854个节点,1598个边(1017个因果边和581个非因果边),并且是使用来自于429个PubMed摘要的文献源的信息来构建的。数个示意性网络节点类型包括根蛋白质节点(例如,CCNE1)、经调整的蛋白质节点(例如,在特定丝氨酸残余处磷酸化的RB1,以及活性节点(例如,CDK2的激酶活性(kaof(CDK2)) and RBI的转录活性(例如,taof(RBI)))。因果边是生物实体之间的因果影响关系,例如,增加的CDK2的激酶活性因果关系地增加在丝氨酸373处磷酸化的RBI。非因果边将不同形式的生物实体,诸如mRNA或蛋白质络合物,连接到其基础蛋白质(例如,在酪氨酸(Y)641处磷酸化的STAT6与其根蛋白质节点STAT6具有非因果关系),而不具有暗含的因果关系。细胞增殖网络以模块方式构建,其中核心细胞周期子模型(core cell cycle submodel)被连接到对于肺中的细胞增殖有贡献的附加生物途径。简要地说,使用了五个符号,它们包括:细胞周期(包括哺乳动物细胞周期的进出的核心机制的标准元素(canonical element),包括但不限于细胞周期蛋白、CDK和E2F族成员);生长因子(包括调节肺细胞增殖所涉及的公共细胞外生长因子,即EGF、TGF-β、VEGF和FGF组成员);细胞内和细胞外信令(包括调节肺细胞增殖涉及的公共细胞内和细胞外途径,包括Hedgehog、Wnt和Notch信令途径,以及涉及肺细胞增殖的钙信令、MAPK、Hox、JAK/STAT、mTOR、前列腺素E2(PGE2)、时钟和核接收信令);细胞相互作用(包括来自于公共细胞粘附分子的相互作用的、引起细胞增殖的信号转导途径(包括与ITGA1-3链络合的ITGB1)以及分子外矩阵成分(具体是胶原蛋白、纤连蛋白和层粘连蛋白));以及实验胚胎学(包括肺细胞增殖的主要已知实验胚胎学调节器,包括组蛋白乙酰基转移酶(HDAC)族和DNA转甲基酶(DMT)组成员DNMT1)。

[0187] 为了检验网络的内容,系统100被用来使用反向因果推理(RCR)来分析转录数据集,该反向因果推理识别可以解释给定转录数据集中的mRNA状态的显著改变的上游控制器(“假设”)。转录数据集被用来使用诸如GEO(Gene Expression Omnibus)和ArrayExpress的公共数据仓库来检验和扩展模型。所使用的数据集包括EIF4G1数据集(GSE11011)、RHOA数据集(GSE5913)、CTNNB1数据集(PMID15186480)和NR3C1数据集(E-MEXP-861)。系统100被用来对这些四个细胞增殖转录数据集中的每一个进行RCR分析并且评价所产生的假设。对于核细胞周期块中的许多节点的预测(包括增加的E2F1、2、和3活动的预测)与他们在调节肺相关细胞类型中的细胞增殖的已公开角色一致。此外,对于RhoA和CTNNB1数据集中的增加的MYC的预测与所报告的MYC在正调节肺和肺相关细胞类型中的细胞增殖所具有的角色一致。除了在细胞增殖被实验性地引起增加的数据集中预测正细胞增殖调节器的增加的活动之外,RCR也预测增殖的负调节器的减少的活动。例如,RBI和E2F4的转录活动的减小(它们

都被已知为细胞周期进展的负调节器)被多个数据集中预测。类似地,CDK1A或CDK2A(其为具有有效反增殖效果的细胞周期检查点蛋白质)的含量的减小也被在全部三个数据集(在其中观察到增殖的增加)中预测。许多这些假设是多效性的信令分子,它们涉及除了增殖之外的其他处理,并且可以由所研究的数据集中的生物不可增殖区域的扰动来引起。除了检验细胞增殖文献模型之外,四个细胞增殖数据集上的RCR被用来识别影响肺的细胞增殖的其他机制。例如,Zbtb17(MIZ-1)的转录活动被预测为在CT B1数据集中增加,但是Zbtb17在调节普通肺细胞增殖中仍不具有直接文献描述的角色。因此,在特定实施例中,药剂对于哺乳动物受试者的生物影响可以通过分析至少肺细胞增殖的网络模块中的数据来取得。适当地,肺细胞增殖网络模型至少包括以下两个以上子模型中的至少一者或组合:细胞周期、生长因子、细胞内和细胞外信号传导、细胞间的相互作用以及表观遗传学。

[0188] 作为第二示例,系统100被用来通过将相关的出版文献的调查与多个转录数据集的计算分析结合来构建主肺部炎症过程的网络模型(炎症过程的网络或IPN)。为了捕获多种细胞类型对于肺部炎症的贡献,系统100被配置为使用模块化架构来构建IPN模型,更大的网络模型包括构成子模型。至少23刻痕的IPN子模型集中于已知涉及在香烟烟雾诱导的肺部炎症的主要细胞类型;具体而言,肺上皮细胞、巨噬细胞、嗜中性粒细胞、T-细胞亚群(Th1、Th2、Th17、Treg和Tc)、NK细胞、树突状细胞、巨核细胞和肥大细胞。在每个子模型内,输入-输出设计被使用;子模型输入对配体/触发器发送信号,该配体/触发器引起或抑制细胞内的信号级联反应,而子模型输出是这些信号传导途径(主要是分泌的细胞活素或生物过程)的细胞/生理产物。系统100被用来根据上述的过程(包括调查科学文献、从Selventa知识库提取的因果关系、从文献接收手工辅助陈述、并把从转录组分析实验(其评估特异性炎症相关的处理)的反因果推理(RCR)分析派生的节点)来构建IPN模型分析。RCR增大(augmentation)是基于从Gene Expression Omnibus(GEO)获得的数据集,表示小鼠全肺暴露于体内LPS(GSE18341)、响应于体外IL15/Th1分化/Th2分化(GSE22886)和肺中性粒细胞暴露于体内的LPS的树突状细胞活化/单核细胞-巨相分化/NK细胞的活化(GSE2322)。因此,在某些实施方案中,试剂对哺乳动物受试者(例如人类)的生物学的影响可以通过在至少一个网络的肺部炎症模型分析数据进行评估。适当地,肺部炎症网络模型包括分别包括一个或更多示例性的节点(在括号中)的两个或多个下面的子模型的一个或多个的组合:(1)粘液分泌过多(响应于细胞因子如IL13、CCL2、TNF和EGF的肺上皮细胞中的高分泌和MUCSAC表达);(2)上皮细胞屏障防御(响应于诸如EGF、TNF、ADAM17和ROS的信号,上皮的屏障功能和紧密连接的渗透性的变化);(3)上皮细胞促炎信令(响应于在上皮细胞活化过程中诸如TNF、TLR4、ELA2和IL-1beta的上行信号的炎症蛋白的表达);(4)中性粒细胞响应(响应于上行信号,如TNF、CSF3和FPR1);(5)巨噬细胞介导的嗜中性粒细胞(响应于诸如TNF的上行信号的,IL-8、丝氨酸1和白三烯B4的导致中性粒细胞的趋化性和补充的分泌);(6)中性粒细胞趋化作用(响应于诸如CSF3、F2、IL1A CXCL12、S100A8和S100A9的上行信号的,趋化的调节);(7)组织损伤(在导致TLR和NFkB信令的组织损伤之后,作为炎症触发器的DAMP和PAMP的释放);(8)巨噬细胞活化(响应于诸如Toll样受体结扎的上行信号的,促炎分子的依赖于NFkB的产生);(9)巨噬细胞的分化(响应于诸如IL-6、IGF-1和干扰素gamma的上行信号的分化);(10)Th1分化(响应于诸如CCL5和DLL1的上行信号的,Th1分化和IFNG表达);(11)Th1应答(上行信号,如IFNG、IL2、LTA和LTB);(12)Th2细胞分化(响应于上行信号,如IL4、IL25和

VIP); (13) Th17分化(响应于上行信号,如TGFB1和DLL4); (14) Th17响应(响应于上行信号,如IL21、IL22和IL26); (15) Treg应答(并且响应于诸如TGFB1和IL7的上行信号的,调节性T细胞分化和IL10表达); (16) Tc响应(响应于上行TCR结扎和IL15的,作为细胞毒性T细胞响应的FASLG的诱发); (17) NK细胞活化(响应与诸如IL-2、IL-4、IL-7、IL-12、IL-15、TGFBeta1、IFNalpha1和ITGB2的上行信号的,NK细胞的靶细胞细胞溶解的诱发); (18) 肥大细胞活化(响应于上行信号,如IL4、KITLG和FcIgE受体); (19) 树突状细胞的活化(响应于诸如LPS和 HMGB1的上游TLR配体的,细胞活素和其它炎症相关蛋白质的产生); (20) 树突状细胞向组织的迁移(响应于诸如补体、CCL3和CCL5的上行信号的,向感染部位的迁移的调节); (21) 树突状细胞向淋巴结的迁移(响应于诸如CXCL9、CXCL10、CXCL11、CCL19和CCL21的上行信号的,向淋巴结的迁移的调节); (22) Th2应答(响应于诸如IL-4和IL-13的上行信号的免疫响应); 以及(23) 巨核细胞分化(响应于诸如IL11和CXCL12的上行信号的巨核细胞分化)。因此,本发明用于测定生物学影响的计算机化的方法可以包括使用肺的炎性过程的网络模型,它包括23个子模型的一个或多个。

[0189] 作为第三个例子中,系统100被用于构建综合网络模型,该模型捕获响应于在非患病哺乳动物的肺和心血管细胞中的内源性和外源性压力的、基于生理细胞生物。系统100被用来根据上述的过程构建细胞应激网络(CSN)的模型,包括调查科学文献、从Selventa知识库提取因果关系、以及从文献中手工接收辅助陈述。CSN的模型包括六个子模型:(1) 异物代谢反应(包括AHR、Cytochrome P450酶、以及这种反应的各种环境的诱导剂); (2) 内质网(ER) 应激(包括解折叠蛋白应答和包括Perk (Eik2ak3)、ATF6和Irelalpha (Ern1)的三个压力传感器的途径下游,同时排除了响应的促凋亡臂); (3) 内皮剪应激(包括层流的(防动脉粥样硬化)和湍流(动脉粥样硬化)剪切应激对单核细胞的粘附的影响,包括NF- κ B和一氧化氮途径); (4) 低氧反应(包括Hif1a活化和靶、转录控制、蛋白质的合成、以及与氧化应激、内质网应激和渗透压应激反应途径的串扰); (5) 渗透应激(包括Nfat5、水通道蛋白、和高渗反应的Cftr下游途径); 以及(6) 氧化应激(包括细胞内自由基的管理途径,内源/外源性氧化剂(包括由暴露到高氧条件下而诱发的)、抗氧化剂、谷胱甘肽代谢、P38、Erk、JNK和NF- κ B途径,以及NRF2及其上游调节和下游抗氧化反应元素(ARE)介导的基因表达)。因此,在某些实施方案中,试剂对哺乳动物受试者(例如人类)的生物学的影晌可以通过在至少一个网络小区应激的模型分析数据进行评估。合适地,所述细胞应激网络模型包括至少一个或两个以下的子模型的一个或多个的组合:异物代谢反应、内质网(ER) 应激、内皮剪应激、低氧反应、渗透应激和氧化应激。

[0190] 系统100被用来针对表示对于香烟烟雾(CS)的转录响应的数据序列来评价CSN模型,该香烟烟雾作为小鼠肺中多效性的细胞应激的原型诱导(GSE18344)。该数据系列包括从暴露到环境空气(假暴露)的CS野生型和NRF2基因敲除动物的数据,有选择的测试CSN模型第1天的CS处理数据。对于三个比较确定显著mRNA的状态变化(SC):野生型1天CS v. 假暴露,NRF2基因敲除1天CS v. 假暴露,以及NRF2基因敲除1天CS v. 野生型1天CS暴露。实验结果与NRF2在到CS的肺细胞反应中的中心作用是一致的。特别是,在野生型小鼠中由1天CS暴露诱导的SC的35%可以通过活化NRF2基因来解释。当1天CS暴露的NRF2基因敲除小鼠与野生型小鼠相比时,减少NRF2转录活性被预测,与这些小鼠中不存在NRF2是一致的。因此,本发明用于确定生物学影响的计算机化的方法可以包括使用细胞应激的网络模型,它包括5

个子模型中的一个或多个。

[0191] 作为进一步的例子中,系统100被用于通过相关的出版文献的调查结合多种转录数据集的计算分析,构建DNA损伤应答、细胞凋亡、程序性坏死、自噬和衰老的网络模型。该网络被称为DNA损伤、吞噬、细胞死亡和衰老的DACS网络。这些DACS网络使用的是高度模块化的设计来构建的,其中较大的网络划分为子模型。影响细胞命运的离散机制(例如,抗细胞凋亡基因的NF- κ B-介导的转录上调的促存活作用)在5个DACS网络区域中由35个子模型描述。总体而言,DACS网络包含1052个独特的节点和1538个独特的边(959个因果边和579个非因果边),这是由1231个PubMed参考的文献引用的支持。在DACS网络节点是生物实体,如蛋白质丰度、mRNA的表达和蛋白质的活动。此外,节点还可以代表生物过程(例如细胞凋亡)。边是节点之间的系,以及被分类为因果的或非因果的。DACS网络被构建和填充来自两个主要来源的内容:从科学文献中描述的现有知识得出的节点和边,以及通过反向因果推理(RCR)从转录组分析数据的计算分析得到的节点。

[0192] 适当地,DACS网络模型包括两个以上的以下子模型的至少一个或组合:对于细胞凋亡——(1)胱天蛋白酶级联反应、(2)ER应激诱导的细胞凋亡、(3)MAPK信令、(4)NF- κ B信令、(5)PKC信令、(6)凋亡线粒体信令、(7)促存活线粒体信号、(8)TNFR/FAS信令、(9)TP53转录签名;对于自噬——(10)自噬作用的ATG诱导、(11)自噬的诱导、(12)mTOR信令、(13)养分转运体的合成和(14)蛋白质合成;对于DNA损伤——(15)影响TP53的活性成分、(16)影响TP53活性的成分、(17)影响TP53活动的成分、(18)DNA损伤到G1/S检测点、(19)DNA损伤到G2/M检查点、(20)双链断裂反应、(21)DNA修复的抑制、(22)NER/XP途径、(23)单链断裂反应、(24)TP53基因转录签名;对于坏死性凋亡——(25)Fas活化、(26)基因标记、(27)炎性介质、(28)RIPK/ROS介导的执行、(29)TNFR1活化;对于衰老——(30)原癌基因诱导的衰老、(31)复制衰老、(32)应激诱导早衰、(33)调节p16INK表达、(34)肿瘤抑制基因的调节以及(35)SASP的转录调控。

[0193] 使用简称为“建筑物(building)”数据集的四个转录数据集(2个DNA损伤和2对衰老)进行的DACS网络的基于RCR的增强。理想情况下,处理所有5个DACS区域的转录数据集被使用,以最大化网络覆盖。但是,由于三个DACS网络区域(细胞凋亡、细胞自噬和性坏死)还没有被经典描述为由转录组变化驱动的,工作的重点是来自于描述DNA损伤反应的实验和衰老的诱导的转录组数据。四个建筑物的数据集都从人或小鼠的成纤维细胞进行体外实验导出,并且代表对于通过紫外线照射或化学DNA交联剂引起的DNA损伤、通过连续传代诱导的复制衰老,以及由博莱霉素(GSE13330)诱发的应激诱导的早产衰老(SIPS)的响应。因此,在某些实施方案中,物剂对哺乳动物受试者(例如人类)的生物学的影 响可以通过在至少一个DACS网络模型分析数据进行评估。适当地,DACS网络模型包括至少一个或两个上面描述的子模型的一个或多个的组合。

[0194] 在处理器处提供多个计算因果网络模型,这些模型表示生物系统,每个计算模型包括代表多个生物实体的节点和代表较多个生物实体的实体之间的关系的边。在一个实施例中,计算因果网络模型从细胞增殖网络、炎性过程网络、蜂窝网络压力和DNA损伤选定、自噬、细胞死亡和衰老的网络中的两个或多个选择。每个网络模型可以包含子模型组成。

[0195] 在一个实施方案中,细胞增殖网络是以肺为重点的细胞增殖网络。适当地,子模型被选自由以下项目构成的组:细胞周期(包括哺乳动物细胞周期的进出的核心机制的标准

元素(canonical element),包括但不限于细胞周期蛋白、CDK和E2F族成员);生长因子(包括调节肺细胞增殖所涉及的公共细胞外生长因子,即EGF、TGF-β、VEGF和FGF组成);细胞内和细胞外信令(包括调节肺细胞增殖涉及的公共细胞内和细胞外途径,包括Hedgehog、Wnt和Notch信令途径,以及涉及肺细胞增殖的钙信令、MAPK、Hox、JAK/STAT、mTOR、前列腺素E2(PGE2)、时钟和核接收信令);细胞相互作用(包括来自于公共细胞粘附分子的相互作用的、引起细胞增殖的信号转导途径(包括与ITGA1-3链络合的ITGB1)以及分子外矩阵成分(具体是胶原蛋白、纤连蛋白和层粘连蛋白));以及实验胚胎学(包括肺细胞增殖的主要已知实验胚胎学调节器,包括组蛋白乙酰基转移酶(HDAC)族和DNA转甲基酶(DMT)组成员DNMT1),或两种或更多种的组合。

[0196] 在一个实施方案中,炎性过程网络是肺部炎性过程的网络。适当地,子模型集中于已知涉及在香烟烟雾诱导的肺部炎症的主要细胞类型。在一个实施例中,子模型选自肺上皮细胞、巨噬细胞、嗜中性粒细胞、T-细胞亚群(Th1、Th2、Th17、Treg和Tc)、NK细胞、树突状细胞、巨核细胞和肥大细胞,或两种或更多种的组合构成的组。

[0197] 在一个实施例中,细胞网络应激的子模型被选自由以下项目构成的组:异物代谢反应(包括AHR、Cytochrome P450酶、以及这种反应的各种环境的诱导剂);(2)内质网(ER)应激(包括解折叠蛋白应答和包括Perk(Erk2ak3)、ATF6和Irelalpha(Ern1)的三个压力传感器的途径下游,同时排除了响应的促凋亡臂);(3)内皮剪应激(包括层流的(防动脉粥样硬化)和湍流(动脉粥样硬化)剪切应激对单核细胞的粘附的影响,包括NF-κB和一氧化氮途径);(4)低氧反应(包括Hif1α活化和靶、转录控制、蛋白质的合成、以及与氧化应激、内质网应激和渗透压应激反应途径的串扰);(5)渗透应激(包括Nfat5、水通道蛋白、和高渗反应的Cftr下游途径);以及(6)氧化应激(包括细胞内自由基的管理途径,内源/外源性氧化剂(包括由暴露到高氧条件下而诱发的)、抗氧化剂、谷胱甘肽代谢、P38、Erk、JNK和NF-κB途径,以及NRF2及其上游调节和下游抗氧化反应元素(ARE)介导的基因表达),或者两种或更多种的组合。

[0198] 在DACS网络模型的实施例中,子模型是从由以下项目构成的组中选择的:对于细胞凋亡——(1)胱天蛋白酶级联反应、(2)ER应激诱导的细胞凋亡、(3)MAPK信令、(4)NFκB信令、(5)PKC信令、(6)凋亡线粒体信令、(7)促存活线粒体信号、(8)TNFR/FAS信令、(9)TP53转录签名;对于自噬——(10)自噬作用的ATG诱导、(11)自噬的诱导、(12)mTOR信令、(13)养分转运体的合成和(14)蛋白质合成;对于DNA损伤——(15)影响TP53的活性成分、(16)影响TP63活性的成分、(17)影响TP73活动的成分、(18)DNA损伤到G1/S检测点、(19)DNA损伤到G2/M检查点、(20)双链断裂反应、(21)DNA修复的抑制、(22)NER/XP途径、(23)单链断裂反应、(24)TP53基因转录签名;对于坏死性凋亡——(25)Fas活化、(26)基因标记、(27)炎性介质、(28)RIPK/ROS介导的执行、(29)TNFR1活化;对于衰老——(30)原癌基因诱导的衰老、(31)复制衰老、(32)应激诱导早衰; (33)调节p16INK表达、(34)肿瘤抑制基因的调节以及(35)SASP的转录调控,或者两个以上的组合。

[0199] 根据本文所描述的系统和方法中,计算模型可以用于表示生物系统及其部件的运行和结构的任何和所有方面。特别是,本文所描述的系统和方法被配置为量化物剂对生物系统及其部件的运行和结构的任何和所有方面的长期影响。因此,虽然大多数本说明书中谈论了生理水平的生化数据,但是计算模型可以被用来表示在离子和原子(例如,钙吞吐

量、神经传递)、核酸、蛋白质和代谢物生物化学的水平的相互作用、细胞器、亚细胞器、细胞、组织室、组织、器官、器官系统、个体、种群、饮食、疾病状态、临床试验、流行病学、捕食猎物的相互作用、以及寄生虫-宿主相互作用。

[0200] 生物系统在人类上下文中的示例包括,但不限于:肺、体壁、骨骼、肌肉、神经、内分泌、心血管、免疫、循环系统、呼吸系统、消化系统、泌尿系统和生殖系统。在一个特定示例中,计算模型可以被用来表示在肌肉系统中的骨骼肌纤维的功能和结构。在另一示例中,计算模型可以被用来表示在骨骼系统的肌纤维的收缩神经控制的运作。在进一步实施例中,计算模型可以用于表示在神经系统中用于内脏运动输出的途径的功能和结构或神经组织中的突触通信的功能。在其他示例中,计算模型可以用来表示在心血管系统中的心动周期和心脏速率的控制的功能和结构。在另外的其它实施例中,计算模型可以被用来表示在淋巴系统中淋巴细胞和免疫应答的功能和结构。在其他示例中,计算模型可以用来表示有症状或不利健康影响的显现和疾病的发作。在某些实施方案中,本发明的计算模型代表的疾病,如心血管疾病、癌症(尤其是肺癌)、慢性阻塞性肺病、哮喘和与吸烟香烟和其他含尼古丁的组合物的消费相关的不良健康状况。这样的计算模型可以用于在本发明的方法预测的抽烟和使用含尼古丁的组合物的生物学影响。

[0201] 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、食道癌细胞、胃细胞、肾细胞、肝细胞、乳腺细胞、前列腺细胞、胰腺细胞、睾丸细胞、膀胱细胞、子宫细胞、结肠细胞和直肠细胞。细胞功能的示例包括,但不限于:细胞分裂、细胞调节、由核细胞活性的控制、以及细胞-细胞信令,计算模型可以被用来表示蜂窝部件的功能和结构。细胞成分的例子包括,但不限于:细胞质、细胞骨架、核糖体、线粒体、细胞核、内质网(ER)、高尔基体或溶酶体。

[0202] 在某些方面,计算机模型可被用来代表蛋白质的结构,功能和合成。此外,计算模型可以被用来表示蛋白质的组分,包括,但不限于,氨基酸序列、二级和三级结构、翻译后修饰(如磷酸化)、构象的数据。此外,计算模型可以被用来表示与蛋白质,包括相关的分子,但不限于,酶。

[0203] 在某些方面中,计算模型被用来代表核酸的结构,功能和合成。核酸是不限于任何特定类型的核酸,包括,但不限于:总基因组DNA、cDNA RNA、mRNA、tRNA和rRNA。在某些方面,来自于生命科学信息的计算模型被用来表示DNA复制、DNA修复和DNA重组的结构和功能。在本文所描述的系统和方法的另一个方面,计算模型确定例如单核苷酸多态性(SNP)、剪接变体、小分子RNA、双链RNA(dsRNA)、小干扰RNA(也称为短干扰RNA或siRNA)、RNA干扰(RNAi)、染色体、染色体的修改或沉默基因。

[0204] 在某些方面中,计算模型被用来表示癌症的途径,包括但不限于,致癌基因和肿瘤抑制基因的功能。例如,一个或更多的计算模型可以用于表示人p53肿瘤抑制基因的基因表达。在另一个方面,计算机模型可被用来代表不同类型的癌症的途径,包括但不限于:血液(例如,白血病)、口腔、唇、鼻腔和鼻窦、喉、咽、食道、胃、肺、肝、胰腺、前列腺、肾脏、睾丸、膀胱、子宫、宫颈、结肠和直肠。

[0205] 在某些方面中,计算模型被用来表示各种疾病的途径,包括但不限于,相关疾病的分子机制的运作。疾病的示例包括但不限于:心血管疾病、冠心病、肺、呼吸系统、血液、神

经、精神、神经、肌肉、骨骼、眼科、胃肠道、泌尿生殖系统、内分泌、皮肤病、炎症、代谢、病原性和感染性疾病。

[0206] 在某些方面中,计算模型识别产物的关系。这样的关系的例子包括,但不限于以下各项:物剂X抑制分子Y的特定功能;物剂X充当药物;物剂X是在公开的专利中;物剂X被用于治疗疾病Y;物剂X抑制实体Y的活性;以及物剂X活化实体Y的ABC活动。

[0207] 在某些方面,计算机模型可被用来代表传染因子的功能和结构。这样的感染性试剂的示例包括但不限于病毒、细菌、酵母真菌或其它微生物如寄生虫。在另一个方面中,计算模型识别诸如病毒、细菌、真菌或朊病毒的病原体与表示在特定的疾病和其他特性的影响的关系连接器。在本公开的其它方面中,计算模型识别特定的可测量的实体是疾病状态的生物标志物、药物的功效或患者分级,识别模型有机体、组织或疾病的其它生物的模式之间的关系,以及相关的疾病或一种流行病及其特点。

[0208] 提供下列实施例作为说明而不是作为一种限制。除非另有指明,本发明采用了本领域已知的常规技术和方法。

[0209] 示例

[0210] 本文所描述的是新的计算方法,其使用定义因果生物(例如分子)网络模型作为数据分析的基板从基础系统范围的数据得出量化生物冲击——定义为生物影响因子(BIF)。这种方法使活性物质的生物影响评估在药理水平是先验的,并且可以通过因果生物网络模型的应用识别作用机制。由于生物活性物质中的一个或混合物引起的特定的生物网络扰动的影响可以是对网络中的每个所述分子实体确定,从而识别由每种物质或混合物诱发的因果机理的影响。由于我们的方法是基于系统范围的实验数据,所以这种量化的方法考虑到了整个生物系统并且因此考虑了由活性物质扰动的许多生物网络。这使得在所描述的生物网络中的每个分子实体(或节点)的量化和客观的评估单独使用或作为签名的一部分作为分子标志物,其密切表达系统中的每个生物网络的扰动的整体状态(相比于控制的活化或抑制)及其与诸如疾病发作或进展的事件的相关性。此外,我们的方法能够进行跨越机制水平的个体和物种的生物影响的量化比较,而基因水平的比较是通过基因组/基因变异而扰乱。这种能力提供了在体内和体外模型系统生物学和人类生物学之间进行转换的手段。

[0211] 这种方法提供了潜在的预测能力和所有假设通过确定的评分算法明确列出。这种方法可以使网络药理学和系统生物学的应用超越毒理学评估,并且可以在诸如药物开发、消费产品测试和环境影响分析的领域被应用。采用了五个步骤的方法的本发明的实施例在图2中描述。

[0212] 示例1——用于数据生产的设计实验

[0213] 为了使研究转化为人类系统,从临床研究中收集的数据是最适用的。然而,由于获得大量人的数据集的挑战,考虑体内非人模型以及基于代表人类疾病的主要方面的体外细胞和有机典型(organotypical, 3D)培养物的模型是有用的。从这些系统中获得的数据允许至少一些洞察由要得到的物质引起的生物网络扰动,以识别用于人类研究的机制特异性生物标志物以及将这些机制链接到疾病的发作影响评估。

[0214] 虽然实验系统在体外和体内都已知有许多不足之处,但是系统化的方法来使用它们将最大限度地减少这些问题(图14)。这样系统的方法可以包括考虑若干限制:

[0215] 暴露。接触途径的物质或复杂的刺激反映暴露在日常生活设定的范围和情况。一

套标准暴露方案被定义为系统地应用到同样定义的实验系统。此外,每个测定可被设计为收集时间和剂量依赖的数据,以捕获早期和晚期事件,并确保有代表性的剂量范围被覆盖。

[0216] 实验系统。如果可能的话,实验系统可以覆盖两个互补的目的:1)该重现人类疾病的定义特征和足够的暴露的动物模型,2)选择细胞和有机典型系统以反映疾病病因所涉及的细胞类型和组织,并优先考虑尽可能概括体内人体生物学的原代细胞或器官培养物。将每个人的体外培养与体内动物模型中得出的最等值的培养物最等价也是重要的。这使得使用体外匹配系统作为“枢纽”来建立从动物模型到体内人类生物学的“平移连续”。

[0217] 测量。用于基因表达、蛋白质表达、翻译后修饰的访问高吞吐量系统系统范围测量结果(如磷酸化和代谢物曲线)被生成并与系统暴露的功能结果相关。因为功能结果测量结果作为评估的锚并且表示疾病病因中明确步骤,所以他们对于策略是有用的。虽然动物模型和细胞系统并不总是完全转化为人类疾病,但是某些步骤可以被再现和他们代表在理解生物网络扰动如何导致疾病中的重要素材。

[0218] 示例2——计算机系统响应曲线

[0219] 在第一步骤中生成的质量控制测量结果在给定的实验系统中构成对于每个给定暴露的系统响应曲线(SRP)。因此,SRP表达每个分子实体由于系统的暴露而改变的程度,并且可以是严格的质量控制和统计分析的结果。以这种方式,不同的测量结果和数据类型可以被结合并共同分析,以提供生物学的更精确的量化表示。

[0220] 接下来,可测量元素(如mRNA表达)被因果关系地通过使用先验知识结合进生物网络模型。再结合发展中的计算方法,这使得能够进行机制评估和由活性物质引起的生物网络扰动的理解。

[0221] 示例3——构建生物网络模型

[0222] 虽然在上一步骤中得出的SRP表示从其确定生物影响的实验数据,但是因果生物网络模型才是用于SRP分析的基板。这种策略的应用需要与风险评估相关的机制生物过程的详细因果网络模型的发展。这样的框架提供了超越已经用在更经典的毒理基因组学中的基因列表的研究的一层机制理解。使用BEL(生物表达式语言,用于生物网络表示的Selventa可计算框架)发展了构建这样的模型的策略,使得其能够基于高吞吐量数据应用到感兴趣的生物过程的评估。

[0223] 这种网络的结构是一个迭代过程。选择网络的生物边界是由与感兴趣的过程(例如,肺中的细胞增殖)相关的信令途径的文献调查引导的。描述这些途径的因果关系是从Selventa的知识库摘录的,以使得网络以从相关细胞类型得出的那些关系为核心。基于文献的网络可以使用具有可用表型提供端点的高吞吐量数据集来验证。

[0224] 一个例子是由与增殖测定结合的关键细胞周期调节器CDK1的抑制剂扰动的人支气管上皮细胞的微阵列分析。这些数据集是使用反向因果推理(RCR)分析的,其为用于识别的生物实体(网络中的节点)的活化状态的预测的方法,该生物实体是统计学显著的并与对于给定的高吞吐量数据集进行的测量结果一致。

[0225] 与细胞增殖中用于生成高吞吐量的数据的观察一致的文献网络节点的RCR预测验证网络有能力捕捉调节所表示的生物过程的机制。此外,通过RCR预测的、尚未在文献中网络所表示的网络相关的节点是集成的。这种方法生成具有从文献的处的节点和边(节点之间的定向连接)以及从相关的高吞吐量数据集生成的节点的综合生物网络。

[0226] 这些网络包含的功能可以使进程评分。拓扑被维持；因果关系(信令途径)的网络可以从网络中的任何一点追溯到可测量的实体。此外,该模型是动态的,并且用于构建它们的假设可以被修改或重列,使适应不同的组织上下文和物种。这样就可以反复试验和改进新知识变为可用。

[0227] 示例4——由SRP计算生物网络的NPA得分

[0228] 为了使能够进行生物网络的扰动的量化比较,一个计算方法是开发出将SRP转换为网络响应得分。网络响应得分被应用到生物网络中的因果模型的上下文内的实验数据。具体来说,作为扰动的下游效果被因果关系地映射到模型中的各个元素的测量结果被经由这里描述的技术聚合到生物网络特定的得分中。通过提供的生物网络扰动的测量结果,网络响应得分允许分子时间与在细胞、组织或器官水平表征表型的分子事件的相关性。

[0229] 示例5——生物计算系统的生物学影响因子

[0230] 单个数字得分可以被计算,其表示混合物的给定物质的系统范围和泛生物机制的影响。在估计扰动剂的生物影响的另一个步骤是聚合网络响应得分——其表示对每个生物网络的影响——成为表示对于整个生物系统的整体影响的一个整体值。每个贡献网络的网络响应得分被聚合,以生成在既需要标准化网络之间的得分、又需要加权每个网络的分布的处理中生成生物影响的估计(图15)。因此,聚合算法的设计可解决限定每个生物网络对于系统的整体状态的相对贡献的问题。最后,当BIF被用作中长期疾病结果的预测器时,它可以使用的实验和(如果可用)的流行病学数据的组合进行校准。

[0231] 示例6——量化吸入化学制品对大鼠鼻上皮肿瘤发生的影响

[0232] 作为如本文所公开的图形理论BIF技术的应用示例,系统100被配置为使用细胞增殖和炎性网络生成生物影响因子(BIF),以量化吸入化学产品对大鼠的鼻上皮肿瘤的影响。从大鼠鼻腔组织随时间的基因表达微阵列分析和剂量下吸入甲醛的数据可以公开地根据取得编号GSE23179(Gene Expression Omnibus)获得。要获得此数据集,八周龄雄性F344/Cr1BR大鼠通过全身吸入暴露于甲醛。全身暴露剂量以0、0.7、2、6、10和15ppm(每天6小时,每周5天)的剂量进行。在暴露开始之后的第1、4和13周后处死吸入动物。在牺牲之后,从鼻子的二级区域获得的组织被解剖并以消化蛋白酶混合物除去上皮细胞。从鼻子的这部分获得的上皮细胞主要由移行上皮和一些呼吸道上皮构成。在上皮细胞中进行基因表达的微阵列分析。该系统响应曲线引擎110从暴露到各种剂量的甲醛13周的大鼠接收转录组数据,并将该数据组成系统响应曲线(SRP)。网络建模引擎112识别与肿瘤相关的两个网络:扩散网络和炎性网络。对于每个剂量,网络评分引擎114针对SRP评价增殖和炎性网络(和特别地,通过这些网络预测的转录行为),并为两个网络中的每个计算网络响应得分。接着,聚合引擎116通过平均两个网络响应得分(反映一个假设,即两个网络基础的机制对于所关心的结果(即,肿瘤发生)的贡献相同)生成每个剂量的BIF。该预测/验证引擎122然后针对从生物文献截取的剂量特异性肿瘤发生率比较每个剂量的BIF值。这种比较是在图13中示出。在图13所示的结果表示由BIF预测的肿瘤发生变得先出超过0.4的BIF阈值。在一些实施方案中,BIF被针对一个已知的或以其它方式预测的生物结果校准(如在图13中表示)。在其它实施例中,BIF未校准,但BIF值被彼此比较,以进行排名和比较生物结果。最初,这些得分是使用强度算法计算的,并且之后使用几何扰动指数评分技术来确认。

[0233] 虽然本公开已经描述参考具体的实施例,但应通过本领域技术人员 可以在形式

和细节上的各种变化可以做出而不脱离本发明的精神和范围内本领域中所理解。按照本发明,计算模型可以用于表示任何生命科学信息。本公开的其它方面阐述在下面的段落:

[0234] 1.一种用于确定物剂对于生物系统的生物影响的计算机化方法,包括:在网络建模引擎处接收与生物系统对于物剂的响应对应的数据,其中,该生物系统包括多个生物实体,每个生物实体都与生物系统中的至少一个其他生物实体相互作用;在网络建模引擎处接收与生物系统对应的数据;在网络建模引擎处生成生物系统的一部分的多个计算模型;其中,每个计算模型包括表示生物实体的节点以及表示生物实体之间的关系的边;在网络评分引擎处生成表示物剂对多个计算模型的影响的至少一个第一得分,以及表示没有暴露到物剂的生物系统的计算模型的至少一个第二得分;以及在聚合引擎处生成表示生物系统的经聚合的得分。

[0235] 2.根据段落1所述的计算机化方法,其中,与物剂对应的数据包括数据表示,其表示生物网络内的一个或多个生物实体由于生物系统暴露到物剂而改变的程度。

[0236] 3.根据段落1或2所述的计算机化方法,其中,至少部分基于与第一物剂和第二物剂中的至少一者相对应的数据,网络建模引擎识别在生物系统内表现响应于物剂的统计显著活动的生物实体。

[0237] 4.根据段落3所述的计算机化方法,其中,网络建模引擎构建一个或多个计算模型,其具有与所识别的生物实体相对应的节点以及与一个或多个所识别的生物实体之间的因果连接相对应的边。

[0238] 5.根据段落1-4中任意一个所述的计算机化方法,其中,多个计算模型中的一个或多个包括一个或多个可直接测量的节点,并且与第一物剂和第二物剂中的至少一个对应的数据包括由一个或多个可直接测量的节点表示的一个或多个生物实体的测量结果。

[0239] 6.根据段落5所述的计算机化方法,其中,基于与各个节点相对应的生物实体的测量结果的数值值,网络评分引擎对具有一个或多个可直接测量的节点的一个或多个计算模型分配得分。

[0240] 7.根据段落1-6中任意一个所述的计算机化方法,其中,多个计算模型中的一个或多个包括一个或多个可直接测量的节点,并且与第一物剂和第二物剂中的至少一个对应的数据包括因果关系地连接到一个或多个可直接测量的节点的一个或多个生物实体的测量结果。

[0241] 8.根据段落7所述的计算机化方法,其中,基于与各个节点因果关系地链接的生物实体的测量结果的数值值,网络评分引擎对具有一个或多个可直接测量的节点的一个或多个计算模型分配得分。

[0242] 9.根据段落7所述的计算机化方法,其中,基于与各个节点因果关系地链接的生物实体的测量结果的结合值,网络评分引擎对具有一个或多个可直接测量的节点的一个或多个计算模型分配得分。

[0243] 10.根据段落1-9中任意一个所述的计算机化方法,还包括:基于物剂对响应计算模型的影响,在聚合引擎处对于多个计算模型中的每个分配权重;以及在聚合引擎处通过分别根据所分配的权重结合第一得分集和第二得分集来生成第一经聚合的得分和第二经聚合的得分;其中,相对经聚合的得分是第二经聚合的得分的函数。

[0244] 11.根据段落10所述的计算机化方法,其中,生成第一经聚合的得分和第二经聚合

的得分包括应用几何图形计算技术。

[0245] 12. 根据段落10-11中任意一个所述的计算机化方法, 其中, 生成第一经聚合的得分和第二经聚合的得分包括应用谱图计算技术。

[0246] 13. 根据段落10-12中任意一个所述的计算机化方法, 其中, 生成第一经聚合的得分和第二经聚合的得分包括应用图形优化计算技术。

[0247] 14. 根据段落1-13中任意一个所述的计算机化方法, 其中, 至少一个第一得分和至少一个第二得分包括向量, 并且聚合的步骤还包括在聚合引擎处过滤至少一个第一得分和至少一个第二得分, 以将第一得分和第二得分中的每一个分解为具有相应标量系数的多个基础向量。

[0248] 15. 根据段落1-14中任意一个所述的计算机化方法, 其中, 过滤还包括移除具有相应标量系数的多个基础向量中的至少一个。

[0249] 16. 根据段落1-15中任意一个所述的计算机化方法, 还包括在聚合引擎处过滤至少一个第一得分和至少一个第二得分, 以移除统计异常值。

[0250] 17. 根据段落1-16中任意一个所述的计算机化方法, 还包括在聚合引擎处标准化至少一个第一得分和至少一个第二得分。

[0251] 18. 根据段落1-17中任意一个所述的计算机化方法, 还包括基于将至少一个第一得分和至少一个第二得分之间的差异最大化, 在聚合引擎处对于多个计算模型中的每个分配权重, 并且在聚合引擎处基于所分配的权重生成相对经聚合的得分。

[0252] 19. 一种用于确定物剂对于生物系统的生物影响的计算机系统, 包括: 网络建模引擎, 其用于接收与生物系统对于物剂的响应对应的数据以及与生物系统没有暴露到物剂相对应的数据; 其中, 生物系统包括多个生物实体, 每个生物实体都与至少一个其他生物实体相互作用; 生成由第一物剂和第二物剂扰动的生物系统的一部分的多个计算模型; 其中, 每个计算模型包括表示一个或多个生物实体的节点和表示生物实体之间的关系的关系的边; 网络评分引擎, 其用于生成表示物剂对多个计算模型的影响的至少一个第一得分, 以及表示没有暴露到物剂的生物系统的计算模型的至少一个第二得分; 以及聚合引擎, 其生成表示物剂对生物系统的生物影响的经聚合的得分。

[0253] 20. 根据段落19所述的计算机系统, 其中, 聚合引擎还包括: 过滤模块, 其用于过滤至少一个第一得分和至少一个第二得分, 以生成至少一个第一经过滤得分和至少一个第二经过滤得分; 网络加权模块, 其用于对于多个计算模型中的每个分配权重; 以及相对评分模块, 其用于基于至少一个第一经过滤得分和至少一个第二经过滤得分生成相对经聚合的得分。

[0254] 21. 一种用于确定表示物剂对生物系统的影响的得分的计算机化系统, 包括: 在网络建模引擎处接收与生物系统对于第一物剂的响应对应的数据, 其中, 该生物系统包括多个生物实体, 每个生物实体都与至少一个其他生物实体相互作用; 在网络建模引擎处生成生物系统的一部分的多个计算模型; 其中, 每个计算模型包括表示生物实体的节点以及表示生物实体之间的关系的关系的边; 在网络评分引擎处生成对于多个计算模型的每个节点的期望响应; 其中, 期望响应是基于暴露到物剂以及计算模型的节点和边中的至少一者; 在网络评分引擎处接收数据; 以及在网络评分引擎处结合期望响应和数据以生成表示计算模型针对数据的性能的性能的得分。

[0255] 22.一种用于确定第二物剂的生物影响相对于第一物剂的生物影响的计算机化方法,包括:在网络建模引擎处接收表示生物系统对于第一物剂的响应的数据,其中生物系统包括多个生物实体,每个生物实体都与至少一个其他生物实体相互作用;在网络建模引擎处接收与在网络建模引擎处生成多个生物系统的一部分的多个计算模型的响应相对应的数据;其中,每个计算模型包括表示生物实体的节点以及表示生物实体之间的关系的边;在网络评分引擎处生成表示第一物剂对多个计算模型的影响的至少一个第一得分,以及表示第二物剂对多个计算模型的影响的至少一个第二得分;以及在聚合引擎处基于至少一个第一得分和至少一个第二得分生成表示第二物剂的生物影响相对于第一物剂的生物影响的相对经聚合评分。

[0256] 23.一种用于确定第二物剂的生物影响相对于第一物剂的生物影响的计算机系统,包括:网络建模引擎,其用于接收表示生物系统对于第一物剂的响应的数据,以及生物系统对于第二物剂的响应的数据;其中生物系统包括多个生物实体,每个生物实体都与至少一个其他生物实体相互作用;生成由第一物剂和第二物剂扰动的生物系统的一部分的多个计算模型;其中,每个计算模型包括表示生物实体的节点以及表示生物实体之间的关系的边;网络评分引擎,其用于生成表示第一物剂对多个计算模型的影响的至少一个第一得分,以及表示第二物剂对多个计算模型的影响的至少一个第二得分;以及聚合引擎,其用于基于至少一个第一得分和至少一个第二得分生成表示第二物剂的生物影响相对于第一物剂的生物影响的相对经聚合评分。

[0257] 1a.一种用于确定扰动对于生物系统的影响的计算机化的方法,包括:

[0258] 在处理器处接收与生物实体集对于第一处理的响应对应的的第一数据,其中,该生物系统包括包含多个生物实体集的多个生物实体,其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;

[0259] 在处理器处接收与生物实体集对于第二处理的响应对应的的第二数据,该第二处理与第一处理不同;

[0260] 在处理器处提供表示生物系统的多个计算因果网络模型,每个计算模型包括表示多个生物实体的节点以及表示多个生物实体中的实体之间的关系的边;

[0261] 在处理器处基于第一数据和所述多个计算模型来生成表示生物系统的扰动的的第一得分,以及基于第二数据和所述多个计算模型来生成表示生物系统的扰动的的第二得分;以及

[0262] 在处理器处基于第一得分和第二得分来生成表示扰动对于生物系统的生物影响的生物影响因子。

[0263] 2a.根据段落1a的计算机化方法,其中,第一得分和第二得分中的每个得分都包括得分向量,并且生成生物影响因子的步骤还包括在处理器处过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影,适合地,其中,过滤还包括从经分解的第一得分和第二得分中的至少一个移除多个投影中的至少一个。

[0264] 3a.根据段落2a的计算机化方法,其中,基础向量集包括描述至少一个计算模型的矩阵的本征向量。

[0265] 4a.根据段落1a到3a中任何一者的计算机化方法,其中,生成第一得分和第二得分包括:在处理器处基于相应的计算模型和第一数据和第二数据中的至少一个来对多个计算

模型中的每一个分配权重;在处理器处生成与多个计算模型相对应并且基于第一数据的多个第一得分;以及在处理器处生成与多个计算模型相对应并且基于第二数据的多个第二得分;根据所分配的权重将多个第一得分结合;根据所分配的权重将多个第二得分结合;其中,生物影响因子是结合的多个第一得分和结合的多个第二得分的函数。

[0266] 5a.根据段落4a的计算机化方法,其中,对多个计算模型中的每一个确定权重包括选择对于多个计算模型中每一个的权重,以将多个第一得分和多个第二得分之间的差异最大化。

[0267] 6a.根据段落1a到5a中任意一个的计算机化方法,其中,生成生物影响因子包括确定表示第一得分的第一向量与表示第二得分的第二向量之间的内积,或者其中,生成生物影响因子包括确定表示第一得分的第一表面与表示第二得分的第二表面之间的距离。

[0268] 7a.根据段落1a到6a中任意一个的计算机化方法,其中,所述计算因果网络模型是从细胞增殖网络、炎性过程网络、细胞应激网络、以及DNA损伤、自吞噬、细胞死亡及衰老网络中选择两个以上。

[0269] 8a.一种用于确定生物影响因子的计算机系统,该计算机系统包括处理器,该处理器被构造为:接收与生物实体集对于第一处理的响应对应的第一数据,其中,该生物系统包括多个生物实体,该多个生物实体包括生物实体集并且其中生物系统中的每个生物实体都与生物系统中的至少一个其他生物实体相互作用;接收与生物实体集对于第二处理的响应对应的第二数据,该第二处理与第一处理不同;提供表示生物系统的多个计算因果网络模型,每个计算模型包括表示多个生物实体的节点以及表示多个生物实体中的节点之间的关系;基于第一数据和所述多个计算模型来生成表示生物系统的扰动的第一得分,以及基于第二数据和所述多个计算模型来生成表示生物系统的扰动的第二得分;以及基于第一得分和第二得分来生成生物影响因子。

[0270] 9a.根据段落8a的计算机系统,其中,第一得分和第二得分中的每一个都包括得分向量,并且其中处理器被进一步配置为:过滤第一得分和第二得分,以将第一得分和第二得分中的每一个分解为在基础向量集上的多个投影;以及从第一得分和第二得分中的至少一者移除多个投影中的至少一个。

[0271] 10a.根据段落8a或9a的计算机系统,其中,基础向量集包括描述至少一个计算模型的矩阵的本征向量。

[0272] 11a.根据段落8a到10a中任意一项的计算机系统,其中,生物影响因子包括确定表示第一得分的第一向量与表示第二得分的第二向量之间的内积。

[0273] 12a.根据段落1a到6a中任意一项的计算机化方法或者根据段落8a到11a中任意一项的计算机系统,其中,生成生物影响因子包括确定表示第一得分的第一表面与表示第二得分的第二表面之间的距离。

[0274] 13a.根据段落1a到6a或12a中任意一项的计算机化方法或者根据段落8a到12a中任意一项的计算机系统,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。在一个实施例中,第一处理包括暴露到通过加热烟草生成的烟雾剂、暴露到通过燃烧烟草生成的烟雾剂、暴露到烟草烟雾、暴露到香烟烟雾、暴露到包括不存在于或不能从生物系统得到的分子或实体的混杂物以及暴露到毒素、治疗性化合物、兴奋剂、松弛剂、天然产品、制造产品、食品物质中的至少一者。

[0275] 14. 一种计算机程序产品,其包括适合于执行段落1a到6a或12a到13a中任意一项的方法的程序代码。

[0276] 15. 一种计算机或计算机可读介质,其包括段落14a的计算机程序产品。

[0277] 虽然本发明的实现方式已经参照具体的示例进行了具体地展示和描述,但是本领域技术人员应当理解,在不脱离由所附权利要求所定义的本发明的精神和范围的情况下能够于其中进行形式和细节的各种改变。本发明的范围因而由所附的权利要求所指示,并且因此意指包含落入权利要求的等价物的意思及范围之内内的所有变化。以上说明书中提到的全部公开文献全部通过引用结合到这里。

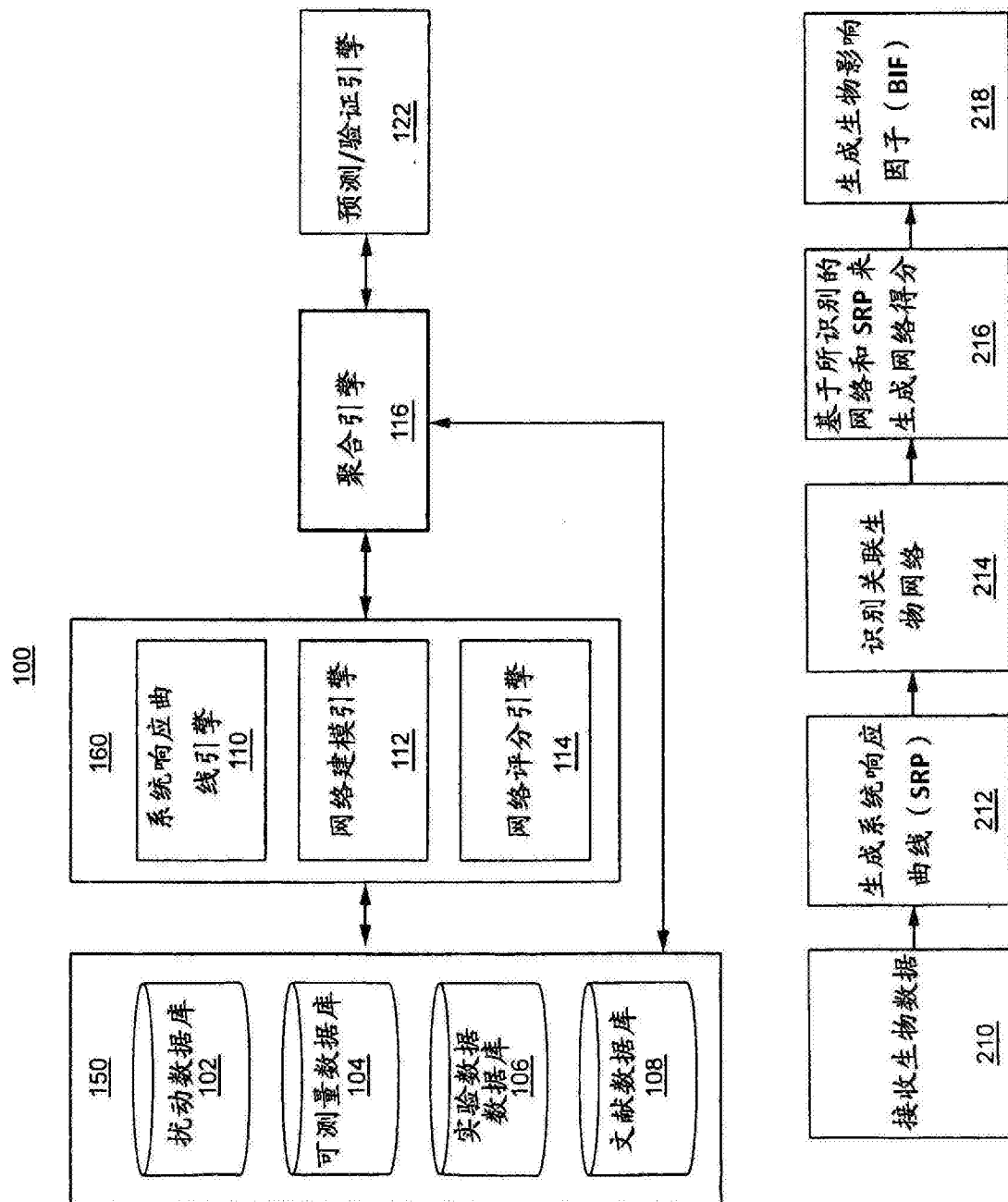


图2

图1

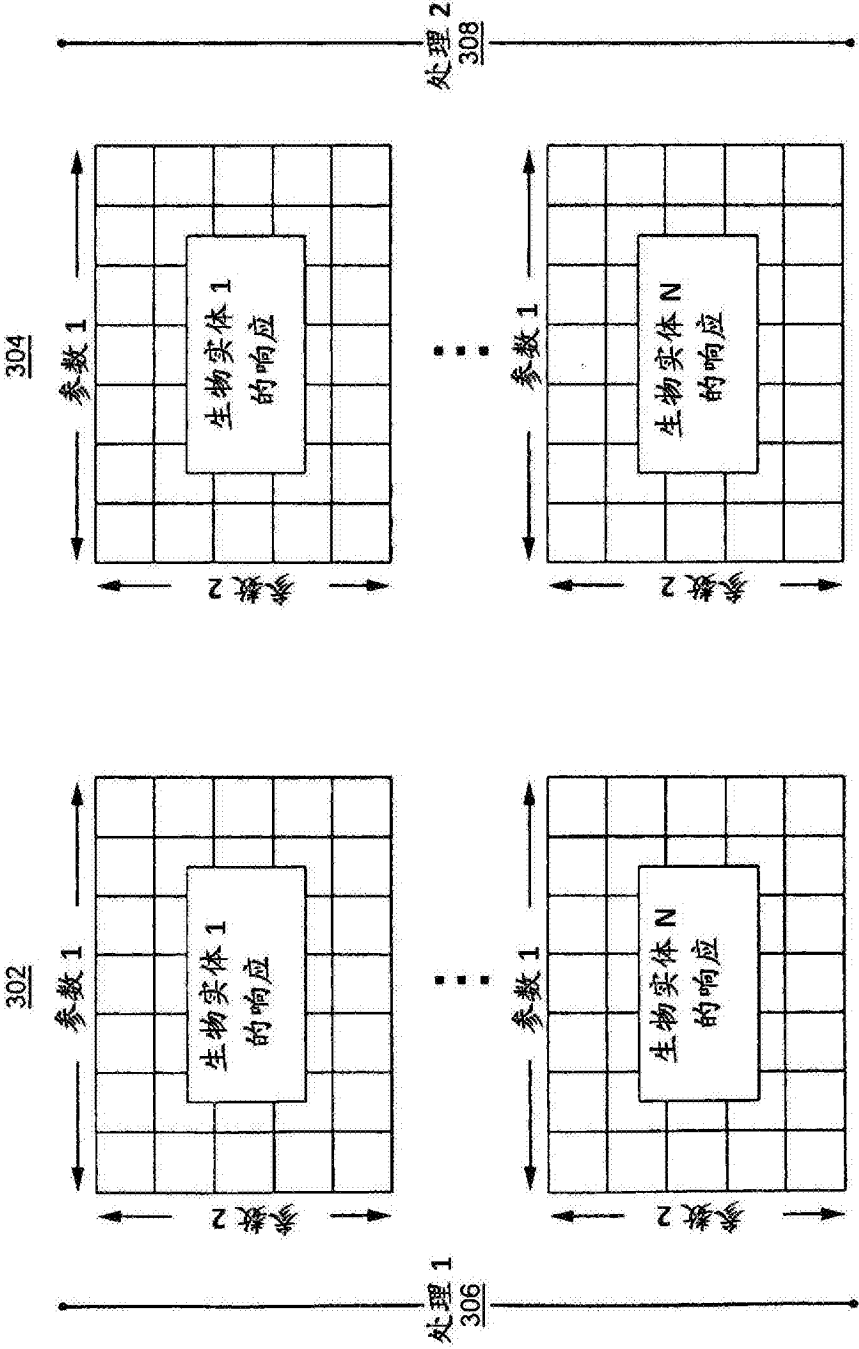


图3

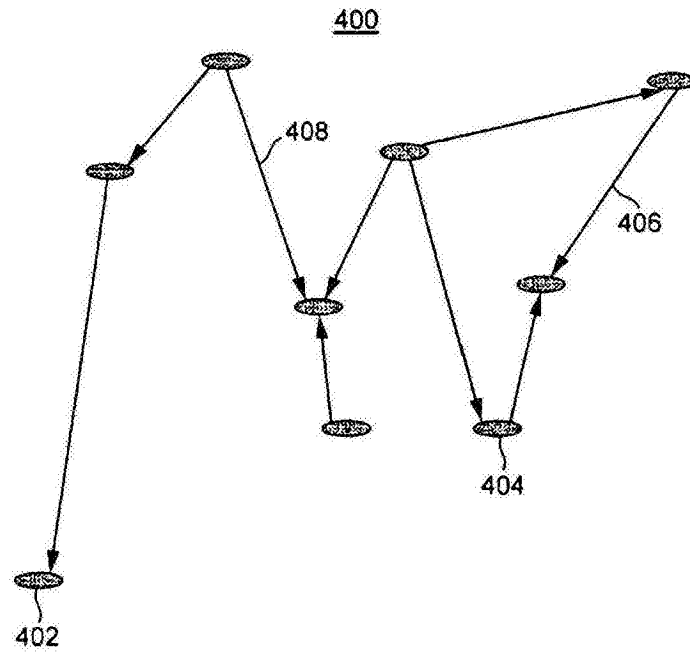


图4

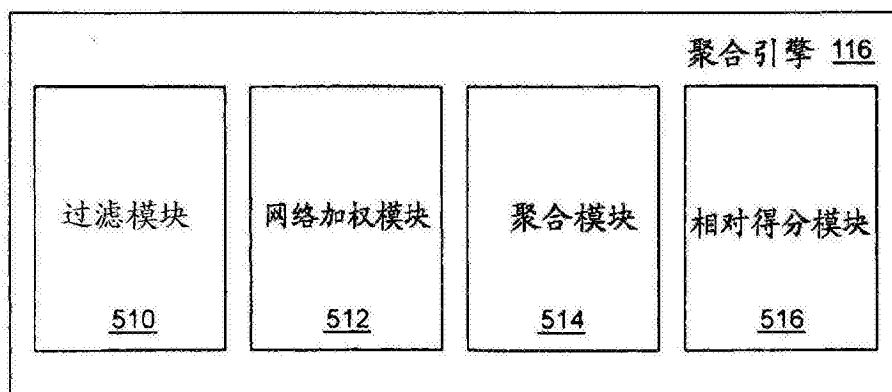


图5

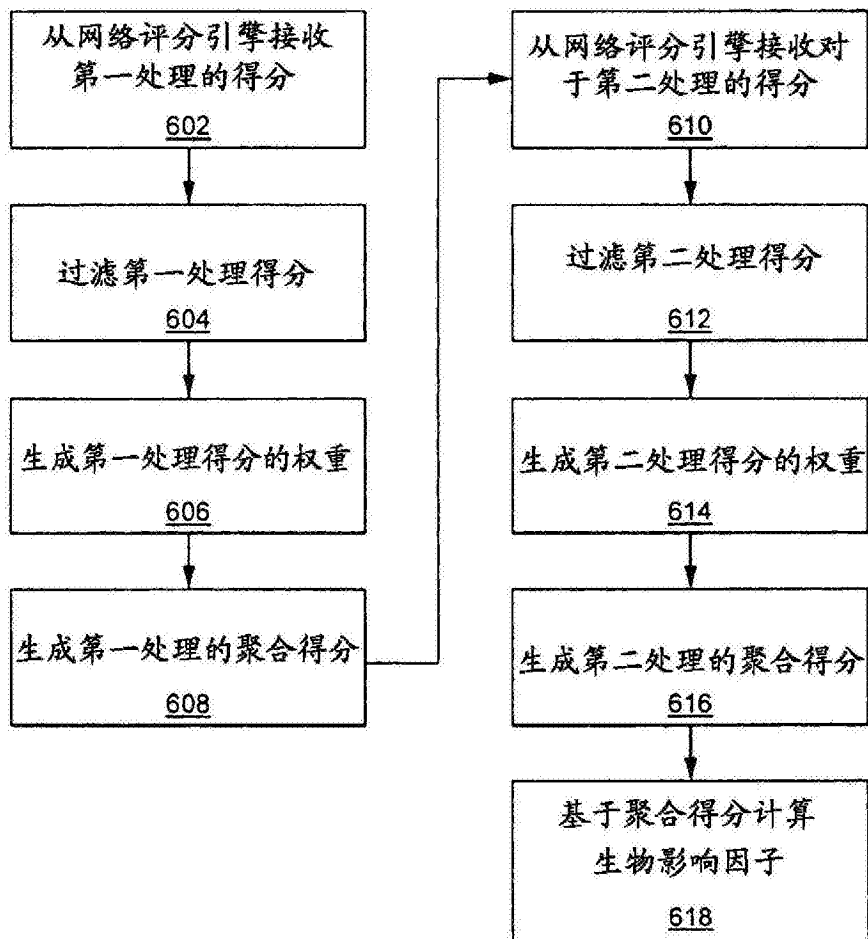


图6

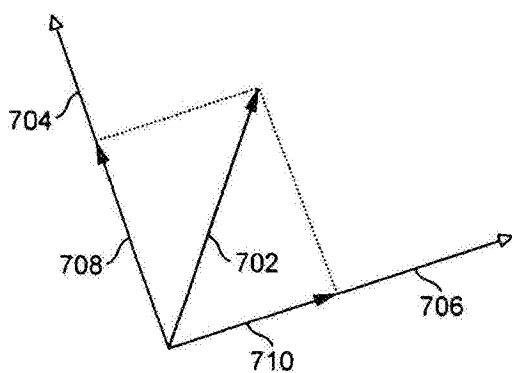


图7

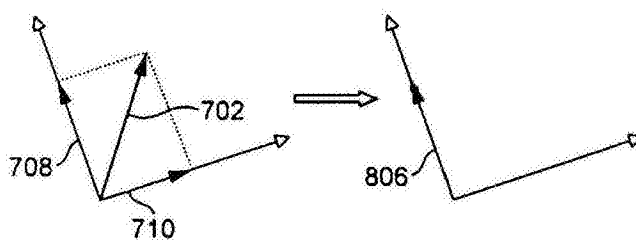


图8A

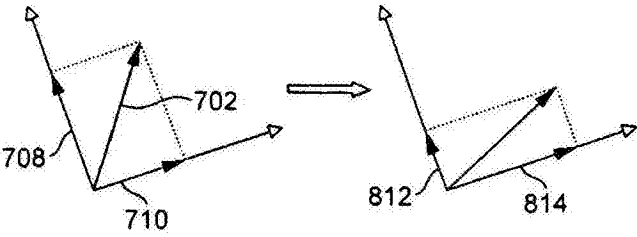


图8B

900

	NET-1 <u>908</u>	NET-2 <u>910</u>	NET-3 <u>912</u>	
<u>短</u> <u>902</u>	$S_{短}^1$	$S_{短}^2$	$S_{短}^3$	
<u>中</u> <u>904</u>	$S_{中}^1$	$S_{中}^2$	$S_{中}^3$	
<u>长</u> <u>906</u>	$S_{长}^1$	$S_{长}^1$	$S_{长}^3$	
	C_1	C_2	C_3	} 914

图9

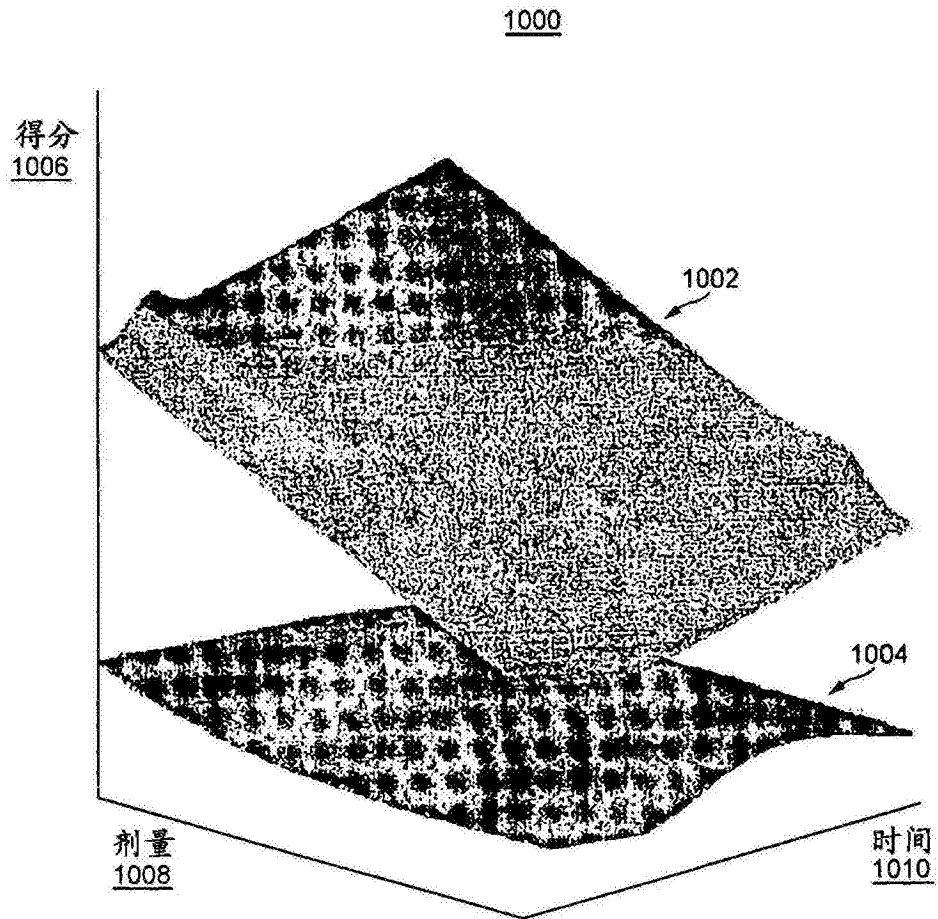


图10

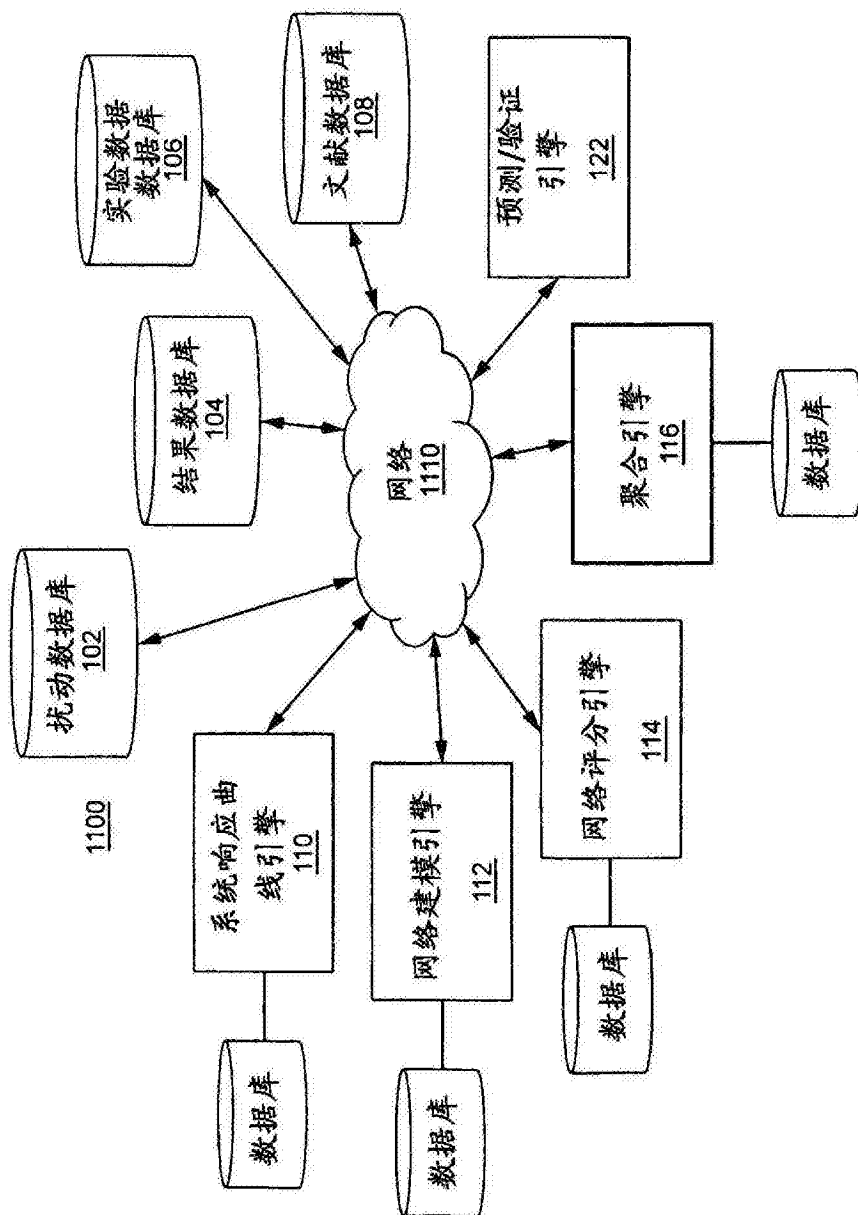


图11

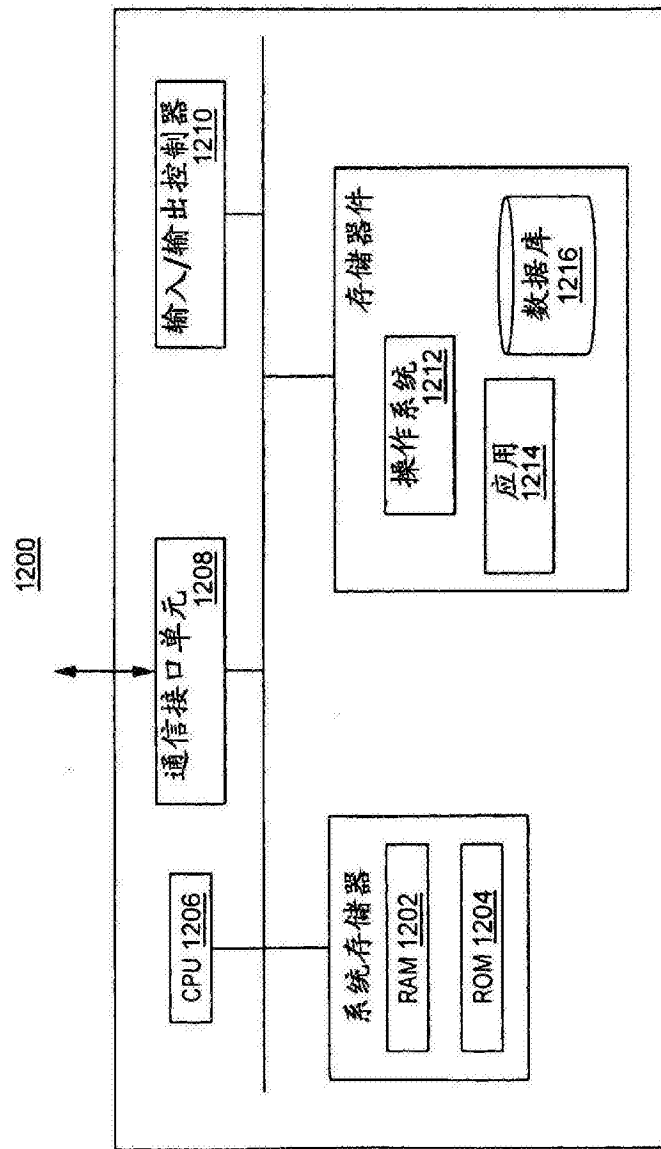


图12

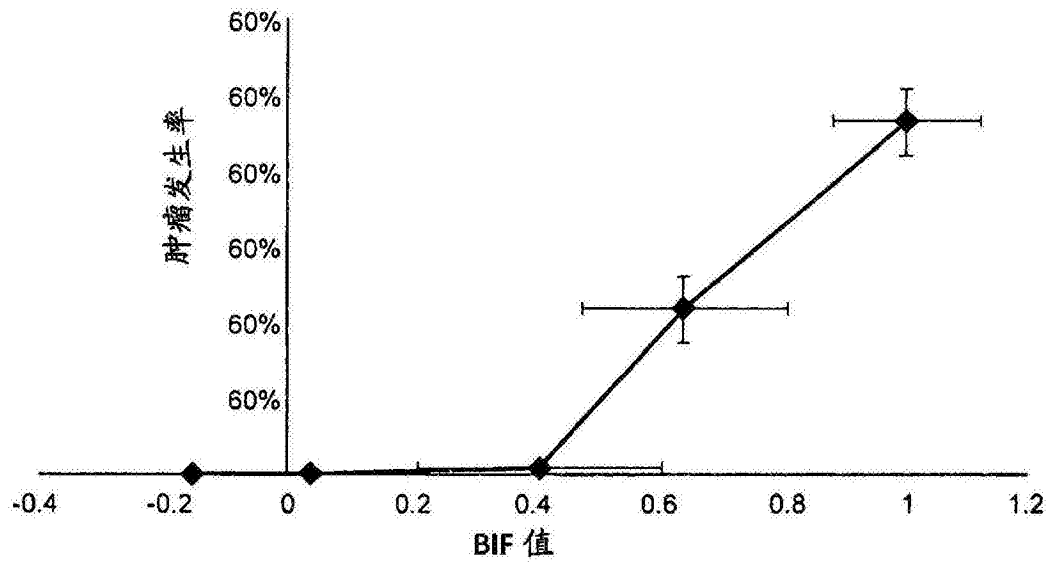


图13

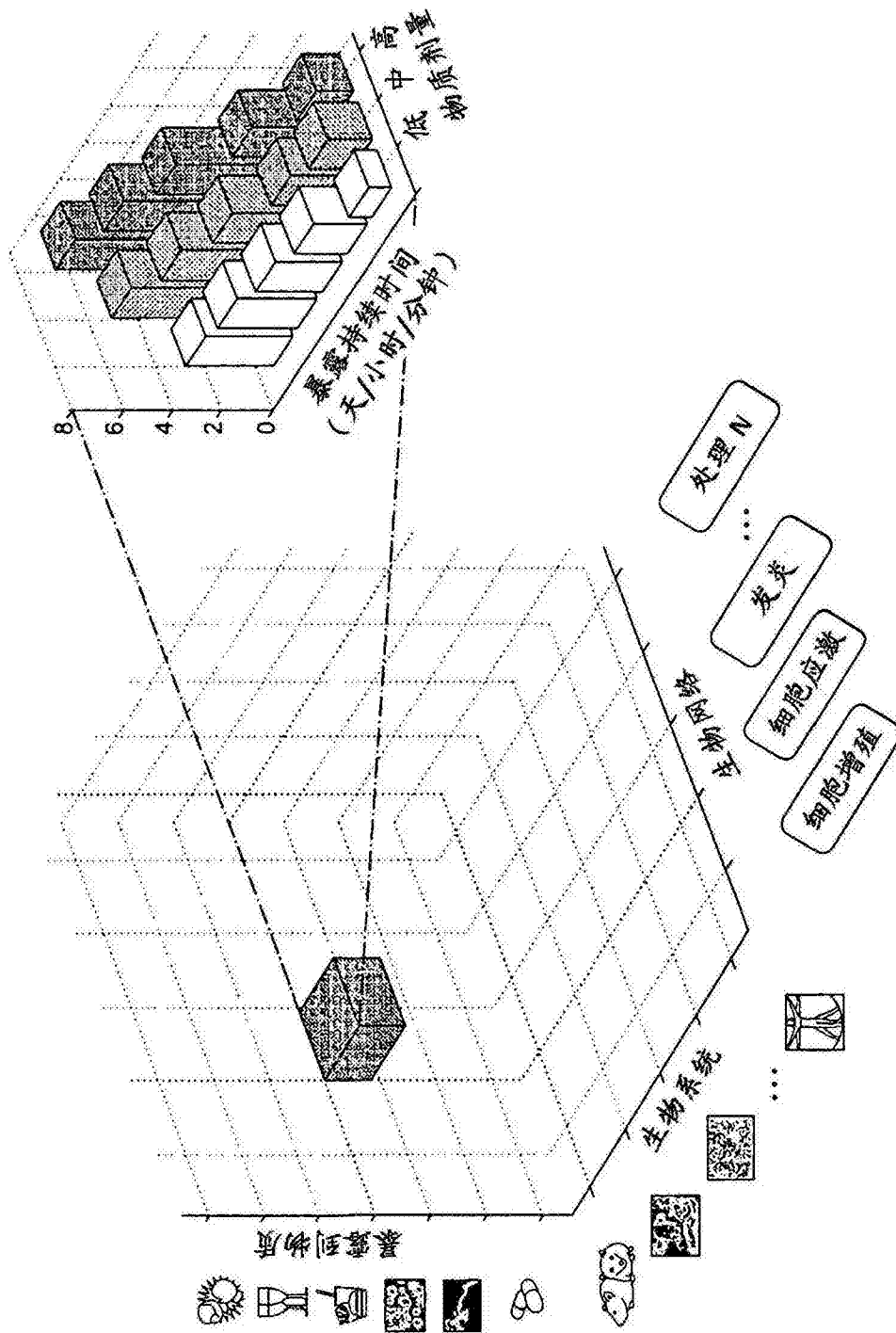


图14

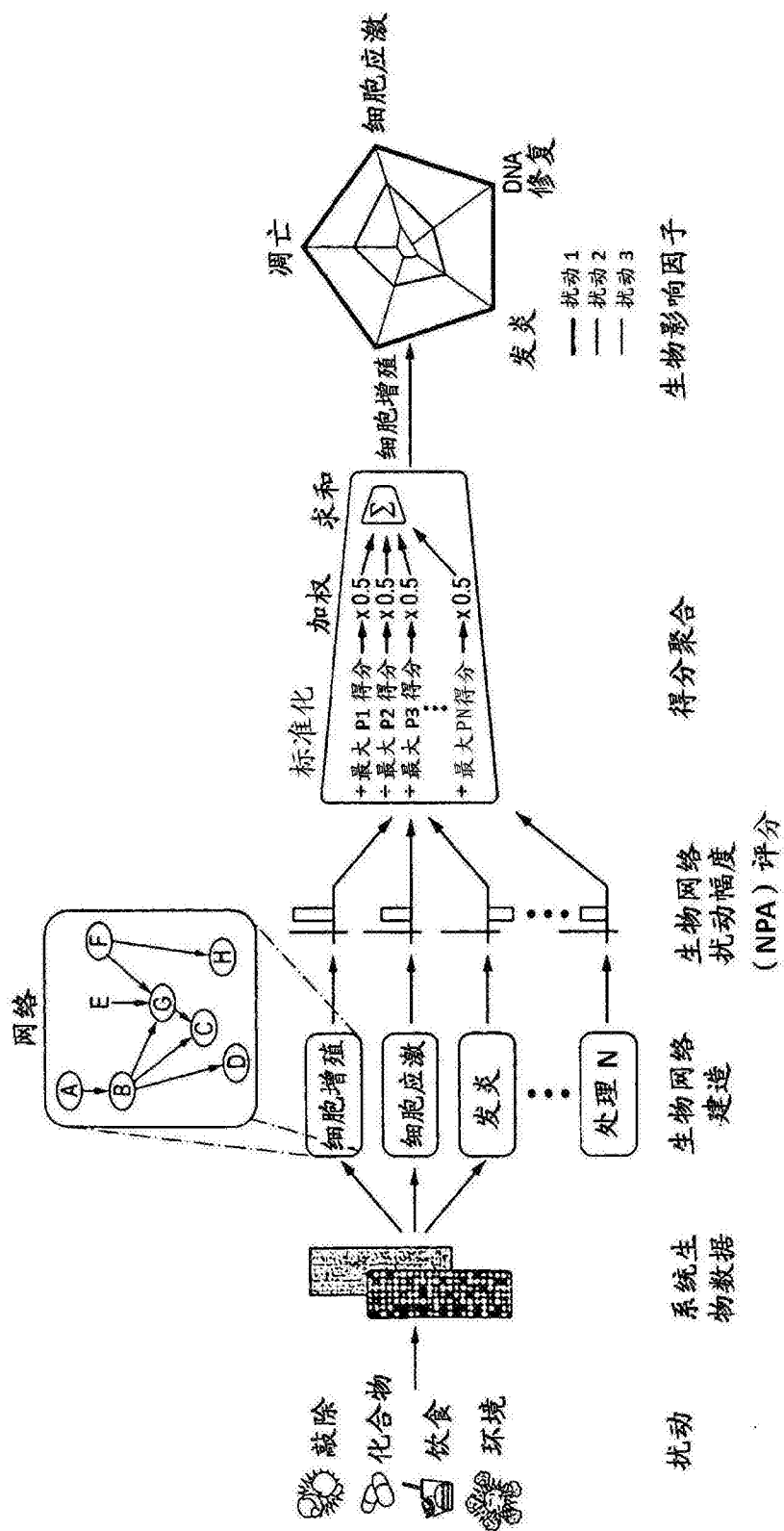


图15