



(12) 发明专利

(10) 授权公告号 CN 107480467 B

(45) 授权公告日 2020. 11. 03

(21) 申请号 201610399510.X

(22) 申请日 2016.06.07

(65) 同一申请的已公布的文献号
申请公布号 CN 107480467 A

(43) 申请公布日 2017.12.15

(73) 专利权人 王忠
地址 100700 北京市东城区东直门内南小
街16号
专利权人 李兵

(72) 发明人 李兵 王忠

(74) 专利代理机构 北京瑞恒信达知识产权代理
事务所(普通合伙) 11382
代理人 张伟 张佰鹏

(51) Int. Cl.
G16B 5/00 (2019.01)
G16B 40/00 (2019.01)

(56) 对比文件

CN 103049678 A, 2013.04.17

CN 103065066 A, 2013.04.24

陈寅莹. 多维度多组分干预脑缺血药理机制
的比较研究.《中国博士学位论文全文数据库医
药卫生科技辑》.2013, (第11期), 第E056-41页.

陈寅莹 等. 网络分析方法在疾病和药物研
究中的应用.《中国中药杂志》.2013, 第38卷(第5
期), 第773-776页.

张莹莹 等. 药物-疾病复杂网络的模块化
解构.《中国药理学通报》.2013, 第29卷(第11
期), 第1499-1502页.

Peter Langfelder et al. Is My Network
Module Preserved and Reproducible.《PLoS
Computational Biology》.2011, 第7卷(第1期),
第1-29页.

审查员 李富贵

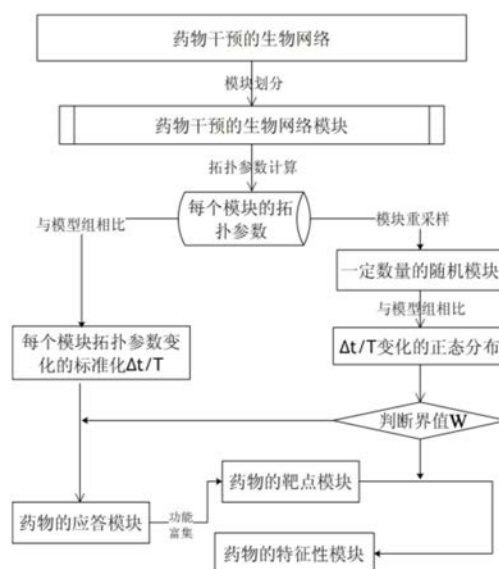
权利要求书2页 说明书11页 附图9页

(54) 发明名称

一种判别或比较药物作用模块的方法

(57) 摘要

本发明提供一种判别或比较药物作用模块的方法,所述方法包括以下步骤:(1)对两个生物网络进行模块的划分,分别得到模块集合;(2)计算两个模块集合中各个模块的一个或多个拓扑结构特征,得到单一或综合的拓扑指标,并由此计算两个模块集合中对应模块的拓扑参数变化;(3)将对应模块的拓扑参数变化与显著性界值进行比较,当两个生物网络分别为药物干预前和干预后的生物网络时,如果拓扑参数变化 \geq 显著性界值,则所述模块为药物的作用模块;当两个生物网络分别为第一药物干预后和第二药物干预后的生物网络时,如果拓扑参数变化 \geq 显著性界值,则所述模块为第一药物和第二药物的作用差异模块。



1. 一种判别或比较药物作用模块的方法,所述方法包括以下步骤:

(1) 对第一生物网络 $G_i(V_i, E_i)$ 和第二生物网络 $G_j(V_j, E_j)$ 进行模块的划分,分别得到模块集合 $G_i(M_i)$ 和 $G_j(M_j)$;其中第一生物网络 $G_i(V_i, E_i)$ 为药物干预前的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为药物干预后的生物网络;或者第一生物网络 $G_i(V_i, E_i)$ 为第一药物干预后的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为第二药物干预后的生物网络,其中所述第一生物网络和第二生物网络中的 V_i 或 V_j 为网络中节点的集合, E_i 或 E_j 为网络中边的集合;

(2) 计算模块集合 $G_i(M_i)$ 和 $G_j(M_j)$ 中各个模块的一个或多个拓扑结构特征 $M(t)$,得到单一或综合的拓扑指标 $T = M(t_1, t_2 \dots t_m)$, m 为拓扑结构特征数目,并由此计算两个模块集合中对应模块的拓扑参数变化 $\Delta t/T$,其中计算拓扑参数变化 $\Delta t/T$ 包括:

计算第一生物网络中每个模块的拓扑参数与第二生物网络中相应模块的拓扑参数相比的变化量,即各个拓扑参数的 $\Delta t/T$ 值,采用离差标准化方法将各个拓扑参数的 $\Delta t/T$ 进行标准化处理,经过离差标准化后,各个拓扑参数的 $\Delta t/T$ 应在 $(0, 1)$ 之间;

(3) 将对应模块的拓扑参数变化 $\Delta t/T$ 与显著性界值 W 进行比较,当第一生物网络 $G_i(V_i, E_i)$ 为药物干预前的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为药物干预后的生物网络时,如果 $\Delta t/T \geq$ 显著性界值 W ,则所述模块为所述药物的作用模块;当第一生物网络 $G_i(V_i, E_i)$ 为第一药物干预后的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为第二药物干预后的生物网络时,如果 $\Delta t/T \geq$ 显著性界值 W ,则所述模块为第一药物和第二药物的作用差异模块,其中显著性界值 W 通过以下方式获得:

利用置换检验或自助重采样方法得到大于1000个与第一生物网络中每个模块对应的随机模块,计算每个随机模块与第二网络相应模块的标准化 $\Delta t/T$ 值,进而计算所有随机模块的标准化 $\Delta t/T$ 值的正态分布概率,显著性界值 W 等于所述正态分布概率中 $p \leq 0.05$ 时对应的标准化 $\Delta t/T$ 值。

2. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

(4) 将所述药物的作用模块进行生物功能富集分析,根据功能富集结果判断所述模块是否是所述药物的靶点模块。

3. 根据权利要求1所述的方法,其特征在于,步骤(1)中,所述药物、第一药物、第二药物为同类型的用以预防和治疗疾病的中药或化学药。

4. 根据权利要求1所述的方法,其特征在于,所述第一药物、第二药物为同类型的中药或化学药的药效组分。

5. 根据权利要求1至4中任一项所述的方法,其特征在于,步骤(1)中,所述第一生物网络和第二生物网络为生物分子网络。

6. 根据权利要求5所述的方法,其特征在于,所述第一生物网络和第二生物网络选自基因共表达网络、蛋白质相互作用网络、基因转录调控网络、非编码基因调控网络及代谢网络。

7. 根据权利要求5所述的方法,其特征在于,所述第一生物网络和第二生物网络为相同类型的基因共表达网络或蛋白质相互作用网络。

8. 根据权利要求5所述的方法,其特征在于,所述第一生物网络和第二生物网络的节点数目 ≥ 200 、 ≥ 500 、 ≥ 1000 或 ≥ 5000 。

9. 根据权利要求1至4中任一项所述的方法,其特征在于,步骤(1)中,采用选自聚类、启

发式搜索、种子扩展、矩阵分解、拓扑网络的网络分析算法进行模块划分。

10. 根据权利要求1至4中任一项所述的方法,其特征在于,步骤(1)中,采用选自分层聚类算法、WGCNA、MCODE、MCL、CFinder、CPM、SPC、G-N algorithm、ModuLand、DME、MINE、SVD的方法进行模块划分。

11. 根据权利要求1至4中任一项所述的方法,其特征在于,步骤(2)中,所述拓扑结构特征 $M(t)$ 选自模块性modularity、度degree、连接度connectivity、密度density、聚类系数clustering coefficient、介数中心性betweenness、网络直径diameter、最短路径shortest path、熵entropy和特征路径长度characteristic path length。

12. 根据权利要求1至4中任一项所述的方法,其特征在于,步骤(2)中,拓扑结构特征数目 m 为1-10。

13. 根据权利要求1至4中任一项所述的方法,其特征在于,拓扑结构特征数目 m 为2,拓扑结构特征 $M(t)$ 包括密度和连接度;或者,拓扑结构特征数目 m 为3,拓扑结构特征 $M(t)$ 包括密度、连接度和熵。

14. 根据权利要求1至4中任一项所述的方法,其特征在于,步骤(4)中,通过基因本体GO功能和/或通路pathway分析进行生物学功能富集分析。

一种判别或比较药物作用模块的方法

技术领域

[0001] 本发明属于生物信息技术领域。具体而言,本发明涉及基于生物网络判别和比较药物靶点模块的方法。

背景技术

[0002] 系统生物学和网络药理学的兴起,使药物机制研究及新药研发从传统的“单组分、单靶点、单疾病”研究模式向“多组分、多靶点、多途径”的方向发展。整合基因组、转录组、蛋白质组和代谢组等高通量数据,结合各种数学模型和算法,在网络的背景下研究疾病、药物、靶点数据及其相互作用关系也成为疾病和药物作用机制研究的热点。

[0003] 许多复杂疾病的发生、发展都与一系列相互作用的基因或蛋白相关,疾病表型是不同的生物学过程在一个复杂网络中相互作用的表现。而药物则是通过作用于疾病网络中的多个靶点,对各靶点的作用产生协同效应,从而对疾病的发生、发展进行干预,最终达到治疗效果。因此,在疾病研究和药物研究领域,网络分析方法越来越多地被应用于识别复杂疾病的生物标志物及药物的作用靶点等。生物网络的分析也为更好的理解复杂疾病的病理机制、系统揭示药物的药理机制及新药的研发提供了新的思路。

[0004] 模块性(modularity)是系统生物网络的一个重要特性。在系统生物网络的背景下,通过各种计算模型和算法对网络进行模块化分析,在模块的层次下对网络进行解构和分析成为网络药理学和系统生物学新的研究方向。模块指生物网络中紧密连接的功能集团,相关文献中亦表示为聚类(cluster)、亚网络(subnetwork)、社团结构(community)、子集(subset)、模体(motif)等。模块在生物网络中具有紧密连接的拓扑结构特征,同时也具有稳定性和功能性的特点。研究者基于网络聚类、启发式搜索、种子扩展、网络拓扑结构、矩阵分解等原理发展了大量的模块划分方法,模块化研究成为生物网络研究的热点。

[0005] 疾病病理和药物作用机制同样具有模块化的特征。“单组分、单靶点”的药物研究模式忽视了靶点基因或蛋白之间的相互作用关系,因此,一些研究尝试从模块的角度来筛选具有特异的靶点和较低副作用的药物。模块药理学(Modular pharmacology,MP)的概念框架也被提出,从模块化角度揭示药物与疾病间关系,提出通过药物的多靶点、模块化设计来治疗复杂疾病,这种以模块而不是单个基因或蛋白为靶点的方法为研究药物作用机制及新药研发提供了新的策略,而如何在复杂的生物网络中识别药物的靶点模块就成为一种挑战。因此,结合生物网络模块的拓扑结构、稳定性和功能性特点,探寻药物靶点模块识别方法具有重要的意义。

发明内容

[0006] 本发明提供了一种基于生物网络判别和比较药物作用模块的方法,本方法整合了模块的拓扑结构分析、统计分析来综合判别药物干预后生物网络的作用模块(又称“应答模块”),组合生物功能分析则可以识别其是否为药物的靶点模块;此外,本发明的方法还可以通过比较不同药物干预后的生物网络的差异模块,来对比分析不同药物的应答差异。

[0007] 定义

[0008] 本文所用术语“生物网络”是指由节点与边组成的网络,由 $G(V,E)$ 表示, V 表示网络中的节点的集合,节点可以是基因、蛋白、化合物、代谢物等药物可能作用的靶点; E 表示网络中的边的集合,边是节点之间的相互作用关系,例如可以是基因的共表达或相互作用关系、蛋白的相互作用关系、转录调节关系等靶点间可能的相互作用关系。常见的生物网络包括基因共表达网络、蛋白质相互作用网络、代谢网络、基因转录调控网络、非编码基因调控网络及各种跨组学的调控及相互作用网络等。以基因共表达网络为例,基因共表达网络是以基因之间表达谱数据的相关性为基础而构建的,常使用图模型来描述基因之间的关系,图中的节点代表基因,边表示两个基因之间的共表达相互作用关系。

[0009] 本文所用术语“模块”是指生物网络中由节点和边组成的紧密连接的功能集团,其涵义等同于生物网络中的“子集”、“聚类”、“社团”、“子网”、“模体”等。

[0010] 本文所用术语“药物的作用模块”是指与药物作用之前的生物网络中对应模块相比较,拓扑结构参数发生显著性变化的模块,其可与术语“药物的应答模块”互换使用。

[0011] 本文所用术语“药物的靶点模块”是指将药物的作用模块经功能富集而确定确实具有相关功能的模块。

[0012] 本文所用术语“药物的差异模块”是指将两种药物作用之后获得的生物网络中对应模块相比较,拓扑结构参数发生显著性变化的模块。

[0013] 本文所用术语“药物的保守模块”是指将两种药物作用之后获得的生物网络中对应模块相比较,拓扑结构参数未发生显著性变化的模块。

[0014] 本发明的具体技术方案如下:

[0015] 本发明提供一种判别或比较药物作用模块的方法,所述方法包括以下步骤:

[0016] (1) 对第一生物网络 $G_i(V_i, E_i)$ 和第二生物网络 $G_j(V_j, E_j)$ 进行模块的划分,分别得到模块集合 $G_i(M_i)$ 和 $G_j(M_j)$;其中第一生物网络 $G_i(V_i, E_i)$ 为药物干预前的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为药物干预后的生物网络;或者第一生物网络 $G_i(V_i, E_i)$ 为第一药物干预后的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为第二药物干预后的生物网络;

[0017] (2) 计算模块集合 $G_i(M_i)$ 和 $G_j(M_j)$ 中各个模块的一个或多个拓扑结构特征 $M(t)$,得到单一或综合的拓扑指标 $T=M(t_1, t_2, \dots, t_m)$, m 为拓扑结构特征数目,并由此计算两个模块集合中对应模块的拓扑参数变化 $\Delta t/T$;

[0018] (3) 将对应模块的拓扑参数变化 $\Delta t/T$ 与显著性界值 W 进行比较,当第一生物网络 $G_i(V_i, E_i)$ 为药物干预前的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为药物干预后的生物网络时,如果 $\Delta t/T \geq$ 显著性界值 W ,则所述模块为所述药物的作用模块;当第一生物网络 $G_i(V_i, E_i)$ 为第一药物干预后的生物网络,而第二生物网络 $G_j(V_j, E_j)$ 为第二药物干预后的生物网络时,如果 $\Delta t/T \geq$ 显著性界值 W ,则所述模块为第一药物和第二药物的作用差异模块。

[0019] 优选地,本发明的方法可以用于判别所述药物的作用模块是否是药物的靶标模块,即该方法还包括:

[0020] (4) 将所述药物的作用模块进行生物功能富集分析,根据功能富集结果判断所述模块是否是所述药物的靶点模块。

[0021] 在本发明方法的步骤(1)中,所述药物、第一药物、第二药物为同类型的用以预防和治疗疾病的中药或化学药。优选地,所述第一药物、第二药物为同类型的中药或化学药的

药效组分。

[0022] 所述第一生物网络和第二生物网络中的 V_i 或 V_j 为网络中节点的集合, E_i 或 E_j 为网络中边的集合。优选地,所述第一生物网络和第二生物网络为生物分子网络,例如选自基因共表达网络、蛋白质相互作用网络、基因转录调控网络、非编码基因调控网络及代谢网络。更优选地,所述第一生物网络和第二生物网络为相同类型的基因共表达网络或蛋白质相互作用网络;更优选地,所述生物分子网络的节点数目 ≥ 200 、 ≥ 500 、 ≥ 1000 或 ≥ 5000 。

[0023] 优选地,采用选自聚类、启发式搜索、种子扩展、矩阵分解、拓扑网络等网络分析算法进行模块划分;更优选地,采用选自分层聚类算法、WGCNA、MCODE、MCL、CFinder、CPM、SPC、G-N algorithm、ModuLand、DME、MINE、SVD的方法进行模块划分。

[0024] 在本发明方法的步骤(2)中,所述拓扑结构特征 $M(t)$ 选自模块性(modularity)、度(degree)、连接度(connectivity)、密度(density)、聚类系数(clustering coefficient)、介数中心性(betweenness)、网络直径(diameter)、最短路径(shortest path)、熵(entropy)和特征路径长度(characteristic path length)。

[0025] 优选地,拓扑结构特征数目 m 为1-10,例如可以是1-10中的任何整数个。

[0026] 优选地,拓扑结构特征数目 m 为2,拓扑结构特征 $M(t)$ 包括密度和连接度;或者优选地,拓扑结构特征数目 m 为3,拓扑结构特征 $M(t)$ 包括密度、连接度和熵。

[0027] 优选地,计算拓扑参数变化 $\Delta t/T$ 包括:

[0028] 计算第一生物网络中每个模块的拓扑参数与第二生物网络中相应模块的拓扑参数相比的变化量,即各个拓扑参数的 $\Delta t/T$ 值,采用离差标准化方法将各个拓扑参数的 $\Delta t/T$ 进行标准化处理,经过离差标准化后,各个拓扑参数的 $\Delta t/T$ 应在(0,1)之间。

[0029] 在本发明方法的步骤(3)中,利用置换检验(permutation test)或自助重采样(bootstrap resampling)程序计算模块拓扑参数变化的显著性界值 W ;

[0030] 优选地,显著性界值 W 通过以下方式获得:

[0031] 利用置换检验或自助重采样方法得到大于1000个与第一生物网络中每个模块对应的随机模块,计算每个随机模块与第二网络相应模块的标准化 $\Delta t/T$ 值,进而计算所有随机模块的标准化 $\Delta t/T$ 值的正态分布概率,显著性界值 W 等于所述正态分布概率中 $p \leq 0.05$ 时对应的标准化 $\Delta t/T$ 值。

[0032] 在本发明方法的步骤(4)中,对药物的作用模块进行生物功能富集分析,以判断其是否是药物的靶点模块。模块的功能富集分析工具较多,可根据模块具体情况选择适合的工具。根据模块的功能富集结果可以判断药物的作用模块是否具有相应的生物学功能。例如,可以通过基因本体(GO)功能和/或通路(pathway)分析进行生物学功能富集分析。一般根据超几何分布计算功能富集的 p 值。GO的功能富集是计算模块组成基因在GO分类中某一(几)个特定的分支的超几何分布关系,返回模块的GO功能富集 p 值,小的 p 值则表示这个模块在该GO中出现了富集。同样,通路(pathway)富集分析是计算模块组成基因在同一条Pathway中的超几何分布关系,返回模块在该通路的富集 p 值,小的 p 值则表示这个模块在该通路中出现了富集。

[0033] 本发明的方法的示意图见图1。

[0034] 在生物网络中,对于判别与比较药物的作用模块一直缺少有效的方法,如何判断药物干预前后生物网络中模块的变化及其显著性一直存在挑战。本发明提出了一种利用网

络的拓扑参数、统计和功能分析方法来判别与比较药物的作用靶点模块的方法,该方法能够有效地发现药物的作用模块,比较不同药物的作用模块的异同,为以模块为药物靶点的药物研发模式提供了新的策略。

[0035] 具体地,药物干预后必然会对生物网络产生影响,而怎么判断哪些模块对药物产生应答,也就是怎样选择药物的靶点模块就是其中的关键问题。模块本身也是由节点和边组成的小网络,所以模块本身具有网络的拓扑结构特征,如模块会有特定的密度、连接度、最短路径、特征路径长度、模块性等。因此,一方面,本发明的方法利用了特定药物干预前和干预后的网络中模块的拓扑结构变化,通过对比分析来判断该模块是否是药物的作用模块,进而通过功能富集分析,可以从中选择药物的靶点模块。另一方面,本发明的方法还可以利用在不同药物干预后的网络中模块的拓扑结构异同,通过对比分析来判断不同药物的不同作用模块,从而判断药物的作用异同。由此,本发明的方法是通过观察药物干预前后模块的拓扑结构变化(而不是单个基因、蛋白表达水平的变化)来反应药物的作用机制。

附图说明

[0036] 以下,结合附图来详细说明本发明的实施方案,其中:

[0037] 图1示出本发明所述药物靶点模块识别方法的流程图;

[0038] 图2示出实施例1中BA组数据的基因共表达模式;

[0039] 图3示出实施例1中BA组的基因共表达模块;

[0040] 图4示出实施例1中BA组与模型组相比得到的应答模块 ($W > 0.2$);

[0041] 图5示出实施例1中BA组与JA组相比得到的差异模块 ($W > 0.2$);

[0042] 图6示出实施例1中BA组的16号模块示例及富集到的生物学功能;

[0043] 图7示出实施例1中BA组的靶点模块示例;

[0044] 图8示出实施例1中BA组的特征性模块示例;

[0045] 图9示出实施例2中判别出的BA药物靶点模块中各节点在已知数据库中的相互作用关系。

[0046] 图10示出实施例2中判别出的药物靶点模块内基因的表达情况。

[0047] 图11示出实施例2中判别出的药物靶点模块内基因的相互作用关系情况。

[0048] 图12示出实施例3中效应组的模块划分结果。

[0049] 图13示出实施例3中效应组与对照组相比的抗PD-1治疗黑色素瘤的应答模块。

[0050] 图14示出实施例4中效应组的模块划分结果。

[0051] 图15示出实施例4中效应组与对照组相比的英夫利昔单抗治疗RA的应答模块。

具体实施方式

[0052] 以下参照具体的实施例来说明本发明。本领域技术人员能够理解,这些实施例仅用于说明本发明,其不以任何方式限制本发明的范围。

[0053] 下述实施例中的实验方法,如无特殊说明,均为常规方法。下述实施例中所用的药材原料、试剂材料等,如无特殊说明,均为市售购买产品。

[0054] 实施例1清开灵有效组分干预小鼠脑缺血模型的靶点模块识别

[0055] 本实例以精制清开灵有效组分黄芩苷(BA)、栀子苷(JA)、胆酸(UA)干预小鼠脑缺

血模型的基因表达谱数据为例,构建不同药物干预前和干预后的基因共表达网络,并且以BA组的数据来具体说明本方法的实施。

[0056] 整个实施过程的示意图见图1。

[0057] 1. 获得基因共表达数据

[0058] 应用精制清开灵有效组分黄芩苷(BA)、栀子苷(JA)、胆酸(UA)分别对小鼠脑缺血模型(MCAO)进行药物干预。

[0059] 小鼠脑缺血模型的构建:小鼠MCAO模型制备采用Zea-Longa线栓法,稍加改进,致备小鼠脑缺血1.5h再灌注24h的模型。小鼠术前禁食12h,自由饮水,经以10%水合氯醛溶液为麻醉剂,采用腹腔注射的方式,以0.4ml/100g体重的剂量,麻醉小鼠,仰卧固定,颈部正中切口,分离并暴露左侧颈总动脉,颈内外动脉,用镊子小心剥离出颈总动脉、颈外动脉,分离迷走神经。用动脉夹夹住颈总动脉,暂时阻断颈总动脉与颈内动脉的血流。用消毒后的缝合线在颈外动脉远心端结扎,并剪断颈外动脉,确保颈外动脉残端长度长于0.5cm。在颈内动脉近端备线,远端放置动脉夹,颈总动脉分叉膨大处切口,松开动脉夹,向颈内动脉插入栓线18~20mm,在大脑中动脉起始端堵塞大脑中动脉,然后将颈总动脉连同鱼线一起结扎,外留10mm线头,缝合皮肤。脑缺血1.5h时退出线栓,使血流再通。24h后观察小鼠的神经功能评分判断造模是否成功。

[0060] 将构建得到的脑缺血模型小鼠随机分成5组,包括sham组、vehicle组、BA组、JA组和UA组,缺血后2h,用药组按2ml药液/kg体重分别经尾静脉注射BA(5mg/ml)、UA(7mg/ml)和JA(25mg/ml)。Sham组和vehicle组小鼠接经尾静脉注射0.9%NaCl(2ml/kg体重)。然后按一步法分别抽提小鼠海马RNA,利用Bioanalyzer基因芯片对RNA质量进行评估。利用小鼠Oligo片段的芯片对干预后小鼠的基因表达谱进行分析。所用芯片由与脑缺血相关的374个c-DNA组成。本实例所用基因表达谱数据共有4组,即模型组(Vehicle)、黄芩苷治疗组(BA)、栀子苷治疗组(JA)及胆酸治疗组(UA)。

[0061] 将模型组基因表达谱数据作为药物干预前的数据,BA组、JA组和UA组基因表达谱数据作为药物干预后的数据。每组基因表达谱数据都由12个样本的374个基因组成(Tbp、Zeb1、Pou2f1、Foxb1、Creb1、Camk2g、Csf1、F5、Hspd1、Matn2、Mt1、Adamts1、Klf6、Dffa、Rgs18、Rhoa、Kcnmb1、Pdc11、Pdpk1、Casp8ap2、Mogat1、Rps26、Akt1、Csnk2a2、Dkk2、Ppm1e、Tnfrsf22、Trp53i11、Smpd3、Grin1、Cdk5、Jund、E2f1、ApoE、I11b、Prkar1b、I17r、Ngfb、Rela、Ifnar1、Adcy6、Bak1、Fzd6、Prkch、Rgs4、Actg1、Gck、Rgs9、Sox9、Rgs1、Dgke、Rgs20、Map2k2、Pin1、Prkcn、Dgkz、Csnk1g1、Dusp4、I111、Grb2、Shc1、Syk、Sim2、Ywhah、Fgf13、Bid、Gstm2、Rarg、Pou3f1、Camk2b、Mapkapk2、Tcf4、Sos1、Stat5a、Vegfb、Bad、Etv3、Id1、Lcat、Nf1、Gsn、Bbc3、Clu、Capn9、Ercc5、Comt、Cts1、Amph、Vegfc、Bax、Cyp51、Sox10、Nfyc、Gata2、Id3、Lef1、Pou6f1、6330503C03Rik、Ech1、Ccl4、Itm2a、Hspa1a、Cbx3、Klf10、Idh3g、Gpx2、Map2k5、Daxx、E2f3、Fgf12、Ikbkg、Btrc、Ikbkap、Ifnar2、Cdk5、Psmbl1、Sufu、Gab1、Sox30、Pxn、Pygo2、Ctnnb1、Grin2a、I15ra、Cdk4、Bcl2l1、Actb、Myb、Prkca、Csf2rb2、Gnaq、B-raf、Wnt6、Adcy7、Cacna1b、Fzd7、Prkcm、Rock1、Adcy8、Prkcc、Sub1、Tuba1b、Rgs6、Plcb1、Mknk1、Diablo、Mef2c、Lrp1b、Dgkg、Rgs12、Serpina5、Hspb1、Ppm1b、Dlk1、Cdc42、Fadd、Mdfi、Fgf11、Map3k4、Klk1b3、I16ra、Tgfb2、Wnt11、Ccna1、Map2k6、Htr1f、Zmat3、Bnip3、Tsg101、Vim、Srf、D14Abble、Cdh11、Vdac2、Tfdp1、Gak、Ccna2、Vegfa、Vegfa、Hdac1、Srebf1、Stch、

E2f1、Nfatc1、Gna12、Gna13、Caenb3、Zic1、Pou4f3、Tcf12、Ldb1、Capns1、Fxyd2、Gcgr、LOC100304588、Syt11、Gadd45a、Pbx2、Ier3、Mapk9、Ctnnbip1、Fgf15、Smad3、Nlk、Mecp2、Sigirr、Rgs18、Ptk2b、Sap30bp、Pcmt1、Tcf3、Braf、Ankrd6、Rgs5、Rap1gap、Adcy1、Grin2b、Gap43、Map2k1、Mapk10、Tgfb1、Lta、Rps6ka1、Wnt3、Rara、Prkcd、Atf4、Adcyap1r1、Cycs、Hint1、Rdx、Src、Adcy9、Prkce、Shcbp1、Elk3、Rgs14、Rgs17、Dusp10、Tubb3、Cyc1、Dusp16、Plcg2、Fzd10、Dgkd、Stat3、Mapk14、Map2k4、Htr1a、Map3k2、Frat1、Casp7、Eef2k、Thbd、Rarb、Camk4、Htr2c、E2f5、Met、Htr7、Camk2b、Stat6、Sod1、Efna4、Vdac3、Adora1、Bmp1、Vdac1、Grb2、Igf2bp2、Top2b、Rpl35、Bdnf、Ppp3cb、Raf1、Cpe、Caenb3、0610007C21Rik、Gna14、Gna11、Tuba1a、Zic3、Mlx、Id4、Ldb2、Sepp1、Prodh、S100a9、Pgam2、Rcan1、Abcc5、Ccr5、Ap1m1、Map3k5、Csnk1e、Axin1、Freq、Sh2b1、Rps6ka4、Wif1、Nkd1、Pam、Crem、Tgm2、Barhl1、Tradd、Plcd4、Ppp2r4、Otdud7b、Rgs7、Casp2、Junb、Il2rg、Bad、Il1a、Egr1、Pdgfra、Gapdh、Eif4e、Apc、Prkcz、Parp1、Egfr、Prkcb1、Rgs2、Traf2、Ccr3、Rgs16、Smpd1、Tbp、Dgka、Mos、B230120H23Rik、Eif4e2、Rgs19、Adcy3、Creb5、Taf7、Pik3ca、Stat1、Il15、Atf3、Dvl3、Map3k3、Casp4、Kcnq1、Ptp4a3、fosB、Wnt3a、Calm1、Htr3a、Crkl、Casp3、Lhx1、Camk4、Selenbp2、Tcfe2a、Scg5、Pold3、Mmp2、Farp2、Pold2、Pold1、Gpx4、App、Mlh3、Rbl2、Tpp2、Cdh3、Fmo2、Pold4、Arf1、Sox1、Arhgef1), 数据构成了 374×12 的基因表达矩阵, 部分数据元素的格式如下:

[0062] Vehicle组:

Array ID	Msam-1	Msam-2	Msam-3	Msam-4	Msam-5	Msam-6
Sc02R02C10	0.908	0.972	0.817	0.849	1.077	1.011
Sc03R02C10	1.052	1.029	1.41	1.267	1.152	1.065
Sc04R02C10	0.615	0.862	0.961	0.953	0.614	0.722
Sc01R03C01	1.127	1.096	1.203	1.307	1.181	1.26
Sc02R03C01	1.017	1.134	0.972	1.169	1.255	1.205

Sc03R03C01	1.082	0.93	1.039	1.154	0.741	1.112
Sc04R03C01	1.112	1.408	1.167	1.201	1.291	1.229
Sc01R03C02	1.763	2.578	2.213	2.703	1.518	2.124

[0065] BA组:

Array ID	Sam 1	Sam 2	Sam 3	Sam 4	Sam 5	Sam 6
Sc02R02C10	1.019	0.966	1.019	1.121	1.146	0.967
Sc03R02C10	0.79	0.948	1.106	0.758	1.016	0.995

Sc04R02C10	0.931	1.003	0.873	0.712	1.02	0.833
Sc01R03C01	1.089	1.137	1.368	1.326	1.282	0.991
Sc02R03C01	1.269	1.482	1.451	1.522	1.044	1.321
Sc03R03C01	1.21	0.625	1.082	0.903	0.83	0.876
Sc04R03C01	1.822	2.376	0.804	1.226	2.714	1.944
Sc01R03C02	1.782	2.199	1.929	2.252	4.277	4.575

[0067] 每一个单元格内的数值为该基因(行)在该样本(列)中的表达值,根据网络构建的方法,在这个表达谱中,具有同表达模式的基因可以被聚类在一起,形成共表达基因,这些共表达的基因聚类成为“基因共表达模块”。

[0068] 2. 网络构建和模块划分

[0069] 用加权共表达网络分析(WGCNA)工具分别构建3个药物组数据的基因共表达网络(节点数目均为374个)并划分模块。经过模块识别BA组一共得到23个模块,模块的大小在3-149个节点(基因)之间;JA组得到42个模块,模块大小在3-46个节点之间;UA组得到15个模块,模块大小在3-29个节点之间。

[0070] 以BA组的模块为例,其整个网络所有基因的表达情况如图2所示,可以看到明显的共表达模块的存在。基于分层聚类算法得到BA组的模块,每一种颜色代表一个模块,结果见图3。

[0071] 3. 模块拓扑参数计算

[0072] 计算药物干预前(模型组)、BA干预后(BA组)的不同状态下模块的拓扑结构特征变化。

[0073] 选择密度、连接度作为模块的拓扑指标,分别计算BA干预组和模型组模块的密度、连接度。将BA干预组与模型组进行对照,观察两组模块在密度、连接度指标中的变化。

[0074] 密度:density=mean(vectorize Matrix(A))

[0075] 连接度:
connectivity=
$$k_i = \sum_{j \neq i} a_{ij}.$$

[0076] 4. 计算拓扑结构变化的显著性。

[0077] 用置换检验(Permutation test)对两组模块的拓扑结构变化进行显著性检验,对BA组23个模块随机置换1000次,与模型组相比,计算每个模块的标准化 $\Delta t/T$,进而得到标准化 $\Delta t/T$ 在1000次随机置换中的正态分布概率,当统计p值 ≤ 0.05 时对应的标准化 $\Delta t/T$ 值最小值,即药物干预后模块拓扑参数显著变化的界值。当p ≤ 0.05 时,w=0.2为模块显著变化的界值。与模型组相比,BA干预组中 $\Delta t/T > 0.2$ 的模块即可认为是药物的应答模块。

[0078] 将BA组与模型组相比,共发现11个BA的应答模块,结果见图4。

[0079] 还发现不同药物相对于模型组的应答模块存在差别,这说明不同药物干预后产生的生物网络中存在的模块具有拓扑结构差异。因此,本发明的方法还可用于不同药物组之间的相比, $\Delta t/T > 0.2$ 的模块为两组之间的差异模块。将BA组与JA组相比,可发现9个差异模块,结果见图5。

[0080] 5. 通过功能分析判别与比较药物的靶点模块

[0081] 利用功能富集分析工具(DAVIDE等工具)对三个药物组的应答模块进行GO功能和

KEGG通路的功能富集分析,以 $p < 0.05$ 为富集分析显著性阈值。以BA组为例,BA组的11个应答模块中,BA-16模块富集到最多功能,包括41个GO功能和9条KEGG通路,BA-6模块没有富集到任何功能。如BA-16模块富集到的生物学功能如下:

	GO 功能	P-value
[0082]	GO:0007267~细胞-细胞信号	0.001834725
	GO:0001890~胎盘发育	0.002565481
	GO:0006793~磷代谢过程	0.005366167
	GO:0006796~磷酸盐代谢过程	0.005366167
	GO:0001701~宫内胚胎发育	0.02229598
	GO:0042325~磷酸化调控	0.026015859
	GO:0019220~磷酸盐代谢过程调控	0.027879898
	GO:0051174~磷代谢过程调控	0.027879898
	GO:0005216~离子通道活性	0.032346557
[0083]	GO:0022838~底物特异性通道活性	0.034250297
	GO:0022803~跨膜转运蛋白活性	0.035130286
	GO:0015267~通道活性	0.035130286
	GO:0000187~MAPK 活性激活	0.043290991
	GO:0004714~跨膜受体蛋白酪氨酸激酶活性	0.045409762
	KEGG 通路	P-value
	mmu05210:结肠直肠癌	0.005863221
	mmu04010:MAPK 信号通路	0.049483325

[0084] 可以看到BA-16模块与细胞信号、胎盘发育、磷代谢过程等生物学功能相关,而且BA-16模块与结肠癌和MAPK信号通路相关,因此可以推断BA干预BA-16应答模块的机制与如上的功能有关。图6为BA-16模块及其富集到的功能。

[0085] 对得到的BA、JA、UA三个药物组的应答模块都进行生物学功能的富集。发现,与模型组相比,BA组得到11个应答模块,其中3个模块未能富集到任何的生物学功能,其余的8个模块为BA的靶点模块;与模型组相比,JA组得到22个应答模块,其中3个模块未能富集到任何的生物学功能,其余的19个模块为JA的靶点模块;与模型组相比,UA组得到8个应答模块,其中1个模块未能富集到任何的生物学功能,其余的7个模块为UA的靶点模块。BA组的靶点模块(BA-5、BA-15)示例见图7。

[0086] BA、JA、UA三组间两两比较,如果一组的模块与其他两组相比较均不同,则该模块为该组的特征性模块。经比较发现,BA有5个特征性模块,特征性模块(BA-9、BA-21)示例见图8;JA有8个特征性模块,UA有2个特征性模块。

[0087] 通过模块的功能富集分析可以得到模块相关的生物学功能,即可能是药物的作用机制。

[0088] 实施例2对判别出的药物的靶点模块进行验证

[0089] 1. 基于数据库已知知识的验证

[0090] 通过String(functional protein association networks)数据库可以查找已知的模块内基因(蛋白)的各种相互作用关系。如BA-16模块,将模块内的节点输入到数据库中,可以得到节点的相互作用关系,参见图9。

[0091] 由图9可以看到,BA-16模块内的多个节点有多条边相连,每一条边代表一种相互作用关系的证据类型,证明已知这些节点确实存在着相互作用关系。

[0092] 2. 基于实验的验证

[0093] 采用实验方法,通过基因的表达量来证实靶点模块的组成意义。

[0094] 用Western Blot方法对各治疗组的靶点模块基因进行验证。BA-21模块中,VEGF和B230120H23Rik的表达水平与Vehicle组相比具有显著性差异,分别参见图10A和图10B。

[0095] 利用免疫共沉淀(Co-Immunoprecipitation)证实模块内基因编码蛋白的相互作用关系。BA-16模块内Met和Egfr蛋白的免疫共沉淀结果表明二者有直接的相互作用关系(图11B),BA-9模块内的Fmo2和Jund的免疫共沉淀结果表明二者有相互作用关系(图11A)。

[0096] 实施例3在抗PD-1治疗黑色素瘤基因表达网络中的应用

[0097] 本实例将本发明的方法应用于抗PD-1治疗黑色素瘤的mRNA表达数据,用以判别抗PD-1治疗黑色素瘤的靶点模块。将抗PD-1治疗产生反应的病例分为一组(效应组),将未产生反应的病例分为一组(对照组),利用WGCNA方法分别构建两组的基因共表达网络。

[0098] 本实例中所用数据下载自GEO(Gene Expression Omnibus)数据库(<https://www.ncbi.nlm.nih.gov/geo/>,数据ID为GSE78220),该数据集为抗PD-1治疗黑色素瘤的mRNA表达数据,样本来源为人,共有28例样本,其中15例样本为效应组,13例样本为对照组。数据中共包含mRNA数量为25268。具体实施过程如下:

[0099] 1. mRNA表达数据的整理

[0100] 将下载的数据整理成WGCNA分析所需的数据表格格式,表格的列为mRNA,行为治疗样本,相对应的单元格内数值为该mRNA(行)在该样本(列)中的表达值,基于mRNA的共表达情况,进行网络构建和模块划分。

[0101] 2. 网络构建和模块划分

[0102] 同样用WGCNA工具构建效应组的mRNA共表达网络并划分模块。在网络构建过程中,有3080个mRNA由于过多的缺失值而被移除,剩余的22188个mRNA被用来构建网络(所构建网络的节点数为21976个)和模块划分。在效应组中共得到30个模块,模块的大小在38-2900个节点(mRNA)之间。经WGCNA识别得到效应组的模块划分结果见图12。

[0103] 3. 模块拓扑参数计算

[0104] 计算效应组和对照组模块的拓扑结构特征变化。选择密度、连接度作为模块的拓扑指标,分别计算效应组和对照组模块的密度、连接度。将效应组与对照组进行比较,观察

两组模块在密度、连接度指标中的变化。

[0105] 4. 计算拓扑结构变化的显著性。

[0106] 用自助重采样程序 (bootstrap resampling) 对两组模块的拓扑结构变化进行显著性检验, 对效应组的模块组成基因进行10000次重采样, 与对照组相比, 计算每个模块的标准化 $\Delta t/T$, 进而得到标准化 $\Delta t/T$ 在10000次重采样过程中的正态分布概率, 当统计 p 值 ≤ 0.05 时对应的标准化 $\Delta t/T$ 值最小值, 即模块拓扑参数显著变化界值。当 $p \leq 0.05$ 时, $W = 0.26$ 为模块显著变化的界值。与对照组相比, 抗PD-1效应组中 $\Delta t/T > 0.26$ 的模块即可认为是药物的应答模块。

[0107] 将效应组与对照组相比, 共发现8个抗PD-1治疗黑色素瘤的应答模块, 结果见图13。

[0108] 5. 通过功能分析判别抗PD-1治疗黑色素瘤的靶点模块

[0109] 利用功能富集分析工具 (DAVIDE 6.7工具) 对效应组的应答模块进行GO功能和KEGG通路的功能富集分析, 以 $p < 0.05$ 为富集分析显著性阈值。在抗PD-1治疗效应组中的8个应答模块中都能够富集到相应的功能, 抗PD-1治疗黑色素瘤的靶点模块。

[0110] 实施例4在英夫利昔单抗 (infliximab) 治疗类风湿性关节炎 (RA) 的基因表达网络中的应用

[0111] 本实例将本发明应用于英夫利昔单抗治疗RA的mRNA表达数据, 用3个拓扑参数 (密度、连接度、熵) 来判别英夫利昔单抗治疗RA的靶点模块。将病例按照英夫利昔单抗治疗前、后分为对照组和治疗组, 利用WGCNA方法构建治疗组和对照组的基因共表达网络。

[0112] 本实例中所用数据下载自GEO (Gene Expression Omnibus) 数据库 (<https://www.ncbi.nlm.nih.gov/geo/>, 数据ID为GSE8350), 该数据集为英夫利昔单抗治疗RA的mRNA表达数据, 样本来源为人, 本实例选用其中36例样本, 其中18例为对照组, 另18例为治疗组。具体实施过程如下:

[0113] 1. mRNA表达数据的整理

[0114] 数据整理过程参见实施例3。

[0115] 2. 网络构建和模块划分

[0116] 同样用WGCNA工具构建效应组的mRNA共表达网络并划分模块。共有776个mRNA被用来构建网络 (所构建网络的节点数为776个) 和模块划分, 在治疗组中共得到46个模块, 模块的大小在3-73个节点 (mRNA) 之间。经WGCNA识别得到效应组的模块划分结果见图14。

[0117] 4. 模块拓扑参数计算

[0118] 计算效应组和对照组模块的拓扑结构特征变化。选择密度、连接度、熵作为模块的拓扑指标, 分别计算效应组和对照组模块的密度、连接度、熵。将治疗组与对照组进行比较, 观察两组模块在密度、连接度、熵指标中的变化。

[0119] 4. 计算拓扑结构变化的显著性。

[0120] 用置换检验 (Permutation test) 对两组模块的拓扑结构变化进行显著性检验, 对治疗组的模块进行1000次置换, 与对照组相比, 计算每个模块的标准化 $\Delta t/T$, 进而得到标准化 $\Delta t/T$ 在1000次置换过程中的正态分布概率, 当统计 p 值 ≤ 0.05 时对应的标准化 $\Delta t/T$ 值最小值, 即模块拓扑参数显著变化界值。当 $p \leq 0.05$ 时, $W = 0.15$ 为模块显著变化的界值。与对照组相比, 英夫利昔单抗治疗组中 $\Delta t/T > 0.15$ 的模块即可认为是药物的应答模块。

[0121] 将治疗组与对照组相比,共发现14个英夫利昔单抗治疗RA的应答模块,结果见图15。

[0122] 5.通过功能分析判别英夫利昔单抗治疗RA的靶点模块

[0123] 利用功能富集分析工具(DAVIDE 6.7工具)对效应组的应答模块进行GO功能和KEGG通路的功能富集分析,以 $p < 0.05$ 为富集分析显著性阈值。在英夫利昔单抗治疗RA的14个应答模块中,有2个模块未富集到任何生物学功能,其余12个模块为英夫利昔单抗治疗RA的靶点模块。

[0124] 实施例5在黄芩苷治疗小鼠脑缺血模型蛋白相互作用网络中的应用

[0125] 本实例以精制清开灵有效组分黄芩苷(BA)干预小鼠脑缺血模型的蛋白相互作用网络为例进行发明的应用。与模型组对照,得到BA干预小鼠脑缺血模型的差异表达基因,基于String数据库进行差异表达基因的网络映射,从而构建BA干预小鼠脑缺血模型的蛋白相互作用网络,用本发明的方法判别BA干预小鼠脑缺血蛋白相互作用网络的靶点模块。

[0126] 本实施例与实施例1的差别在于所用数据来源数量不同,且网络类型为蛋白相互作用网络,因此与实施例1中的模块划分结果和应答模块判别结果均存在一定差异。下面仅对与实施例1中差异部分进行描述。

[0127] 1.蛋白相互作用网络数据。

[0128] BA干预小鼠脑缺血模型的差异表达基因获取参见实施例1。利用String数据库,通过差异表达基因的映射构建蛋白相互作用网络,所构建的网络共包含2229个蛋白(节点)。

[0129] 2.模块划分

[0130] BA干预小鼠脑缺血模型的蛋白相互作用网络中,用MCODE方法进行模块划分,共得到49个模块。

[0131] 3.模块拓扑参数计算

[0132] 计算BA干预后模块的拓扑结构特征变化。

[0133] 选择密度、连接度作为模块的拓扑指标,分别计算BA干预组和模型组模块的密度、连接度。将BA干预组组成模块与模型组中相应节点进行对照,观察两组模块在密度、连接度指标中的变化。

[0134] 4.计算拓扑结构变化的显著性。

[0135] 用置换检验(Permutation test)对两组模块的拓扑结构变化进行显著性检验,对BA组49个模块随机置换1000次,与模型组相比,计算每个模块的标准化 $\Delta t/T$,进而得到标准化 $\Delta t/T$ 在1000次随机置换中的正态分布概率,当统计 p 值 ≤ 0.05 时对应的标准化 $\Delta t/T$ 值最小值,即药物干预后模块拓扑参数显著变化的界值。当 $p \leq 0.05$ 时, $W=0.11$ 为模块显著变化的界值。与模型组相比,BA干预组中 $\Delta t/T > 0.11$ 的模块即可认为是药物的应答模块。

[0136] 将BA组与模型组相比,共发现13个应答模块。

[0137] 5.通过功能分析有个6应答模块没有富集到相应功能,其余7个模块为BA在蛋白相互作用网络是的靶点模块。

[0138] 通过以上实例分析,证明了本发明能够有效地在生物网络中识别药物的靶点模块,还可以通过不同药物组间的比较得到不同药物的保守模块和差异模块。

[0139] 以上对本发明具体实施方式的描述并不限制本发明,本领域技术人员可以根据本发明作出各种改变或变形,只要不脱离本发明的精神,均应属于本发明所附权利要求的范围。

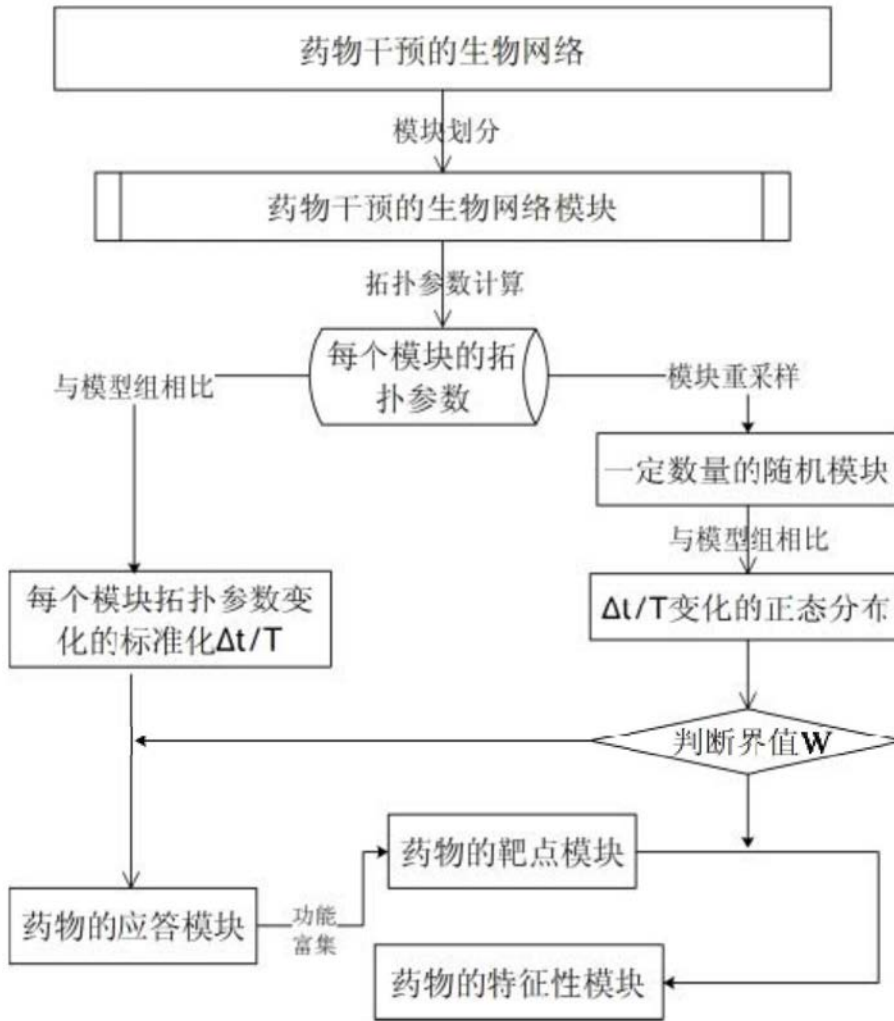


图1

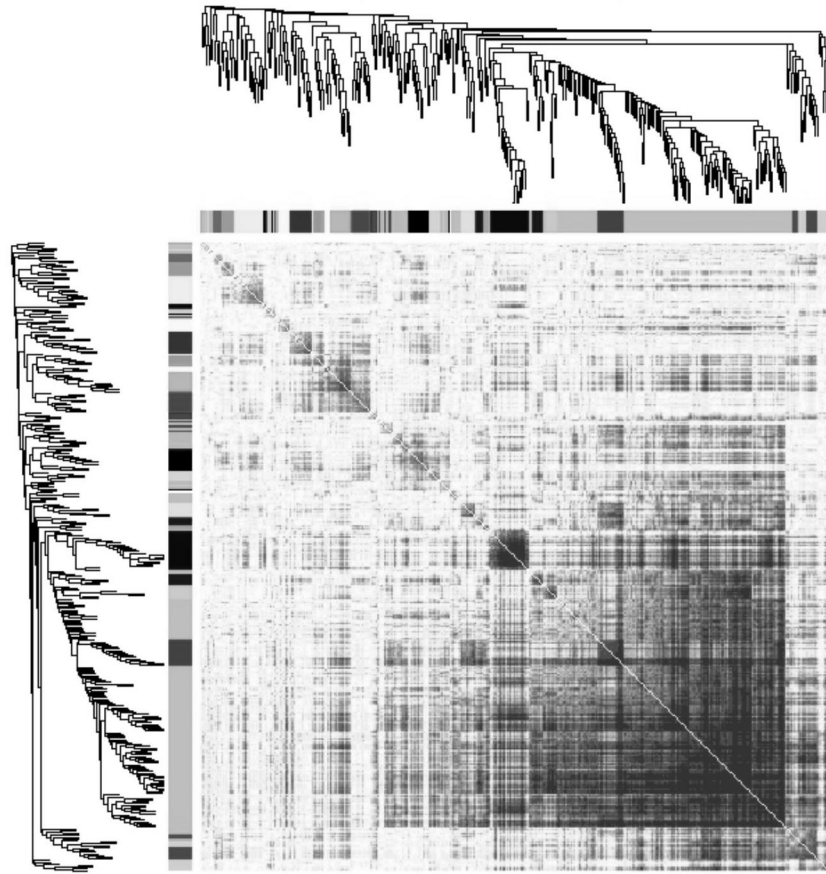


图2

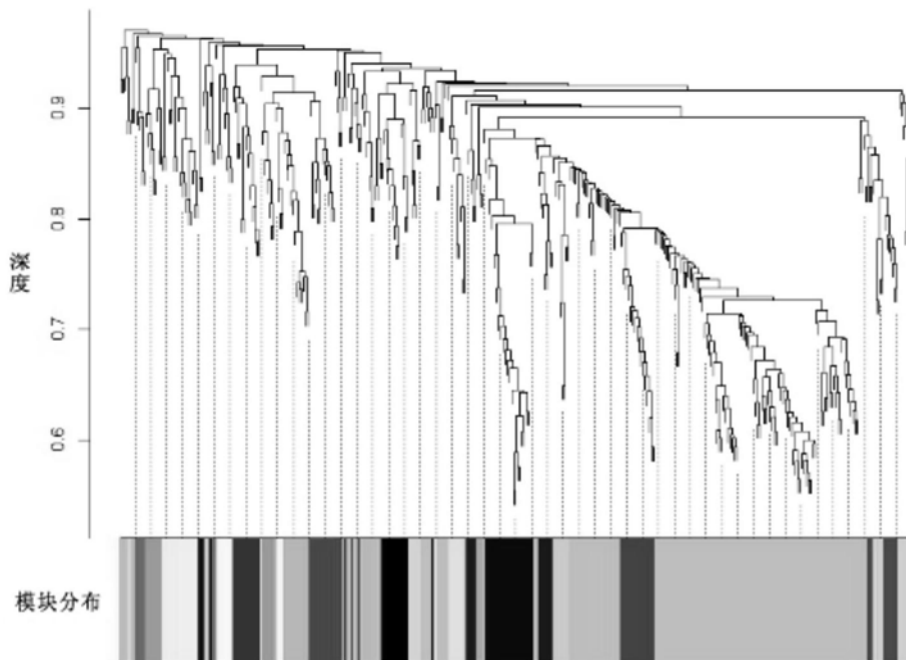


图3

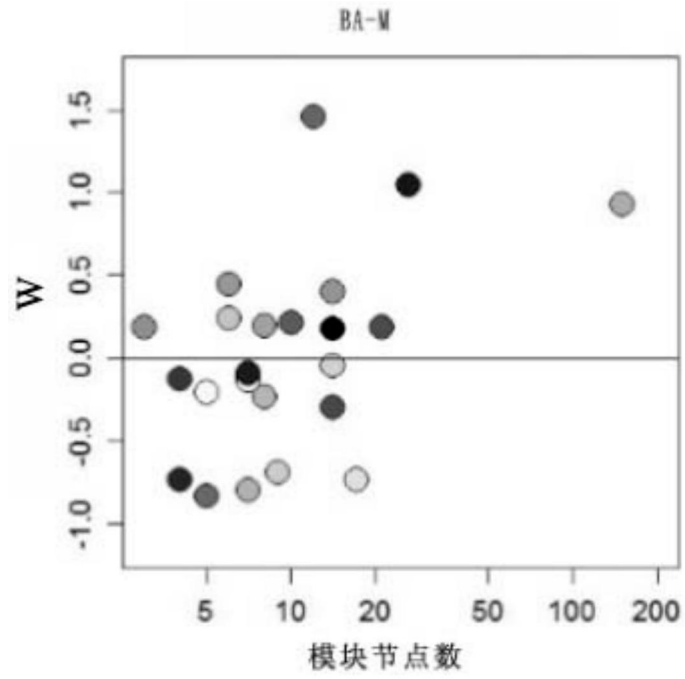


图4

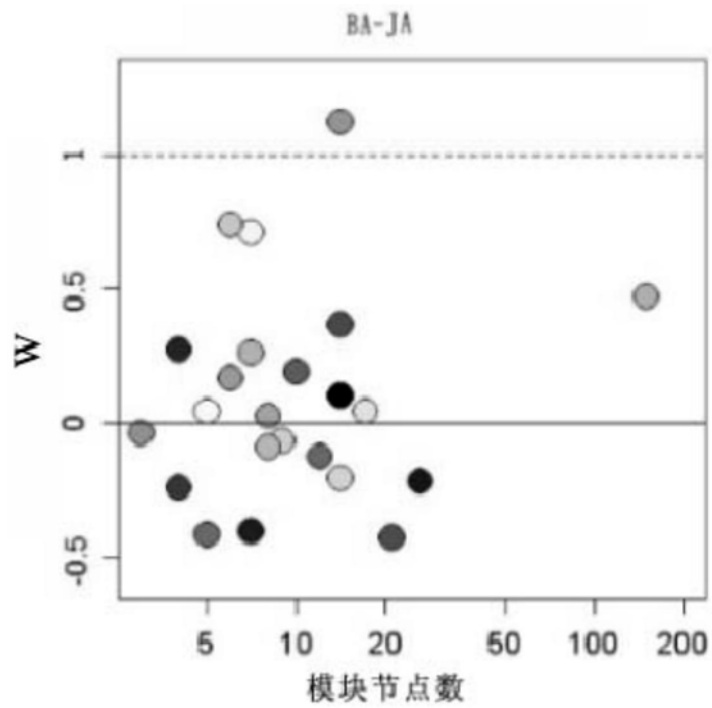


图5

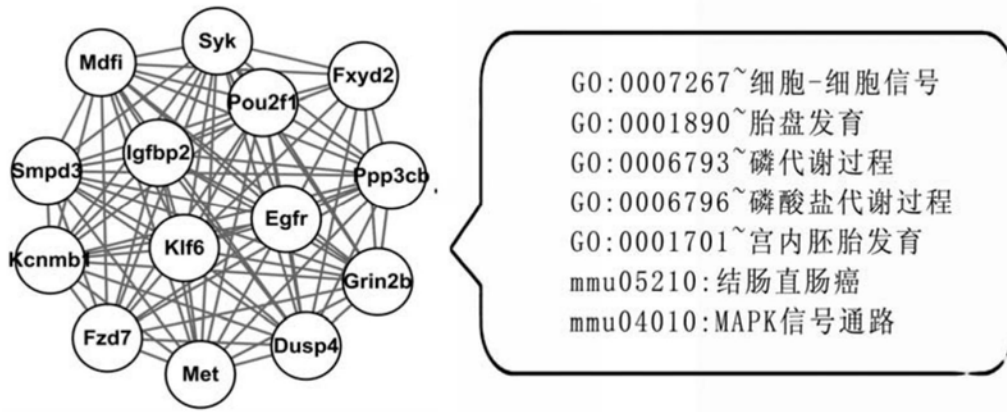


图6

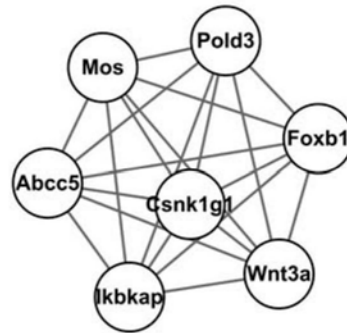
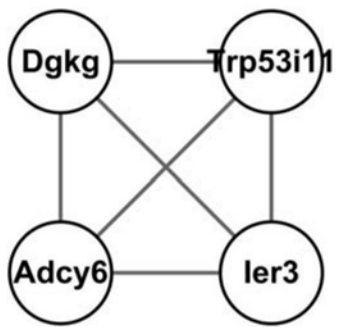


图7

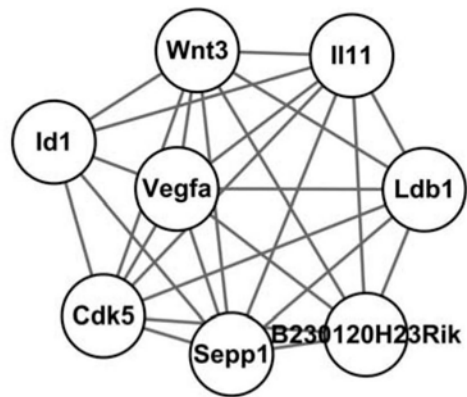
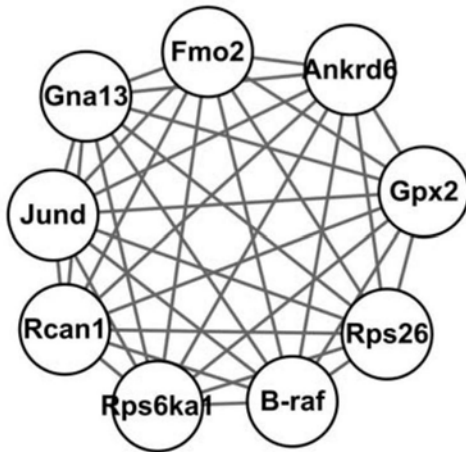


图8

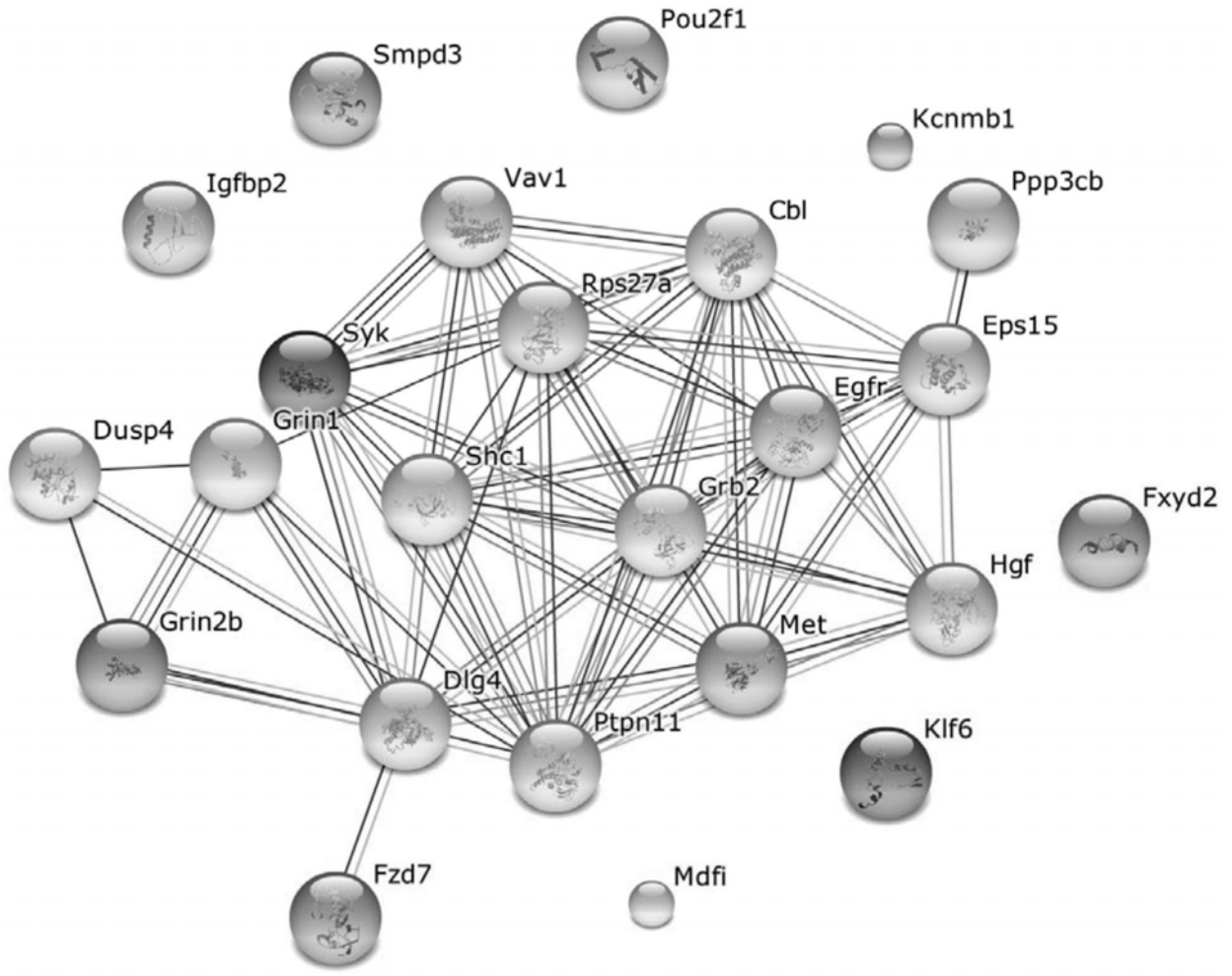
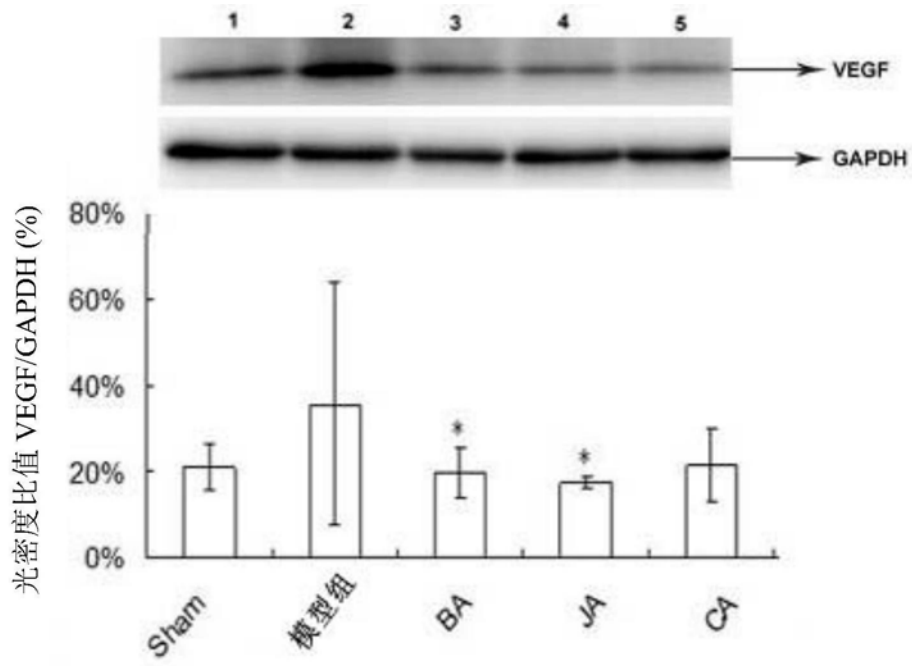
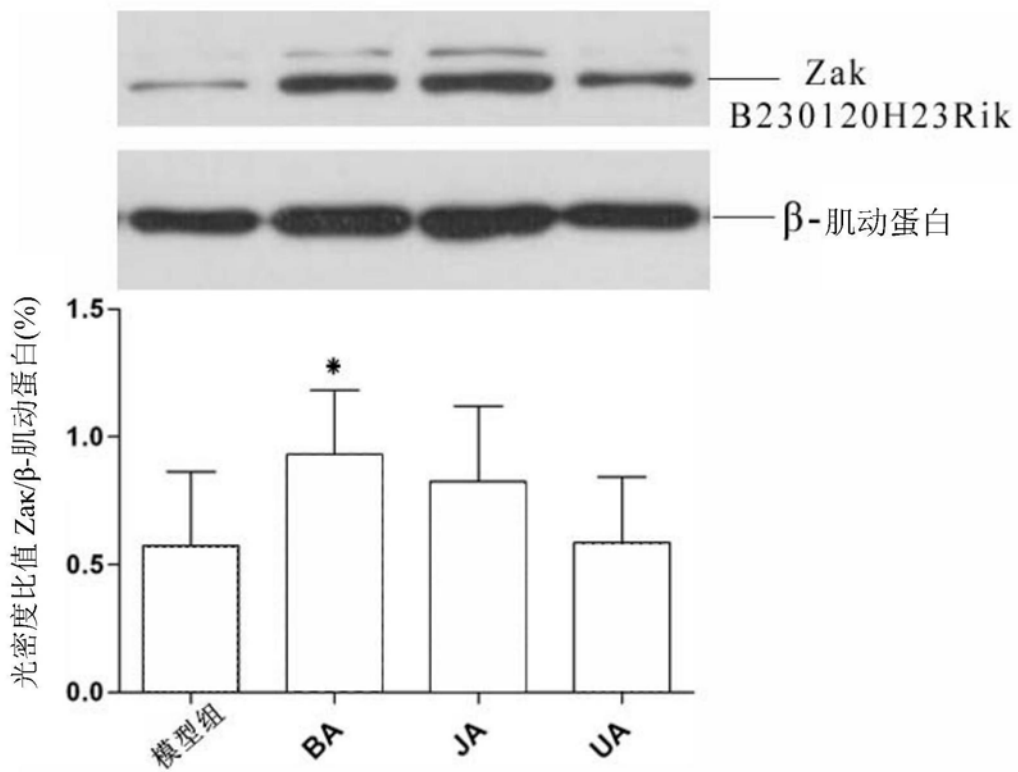


图9

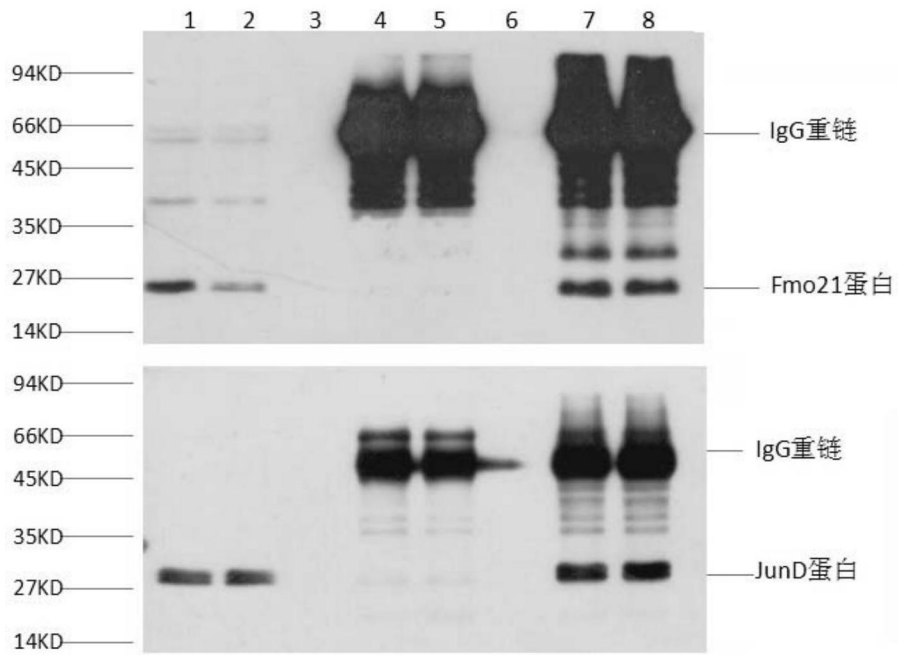


10A

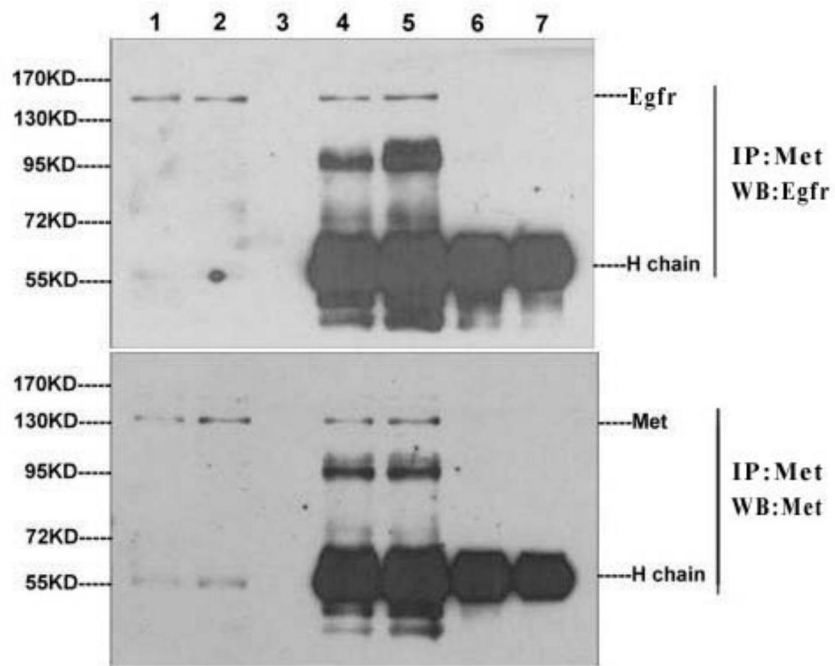


10B

图10



11A



11B

图11

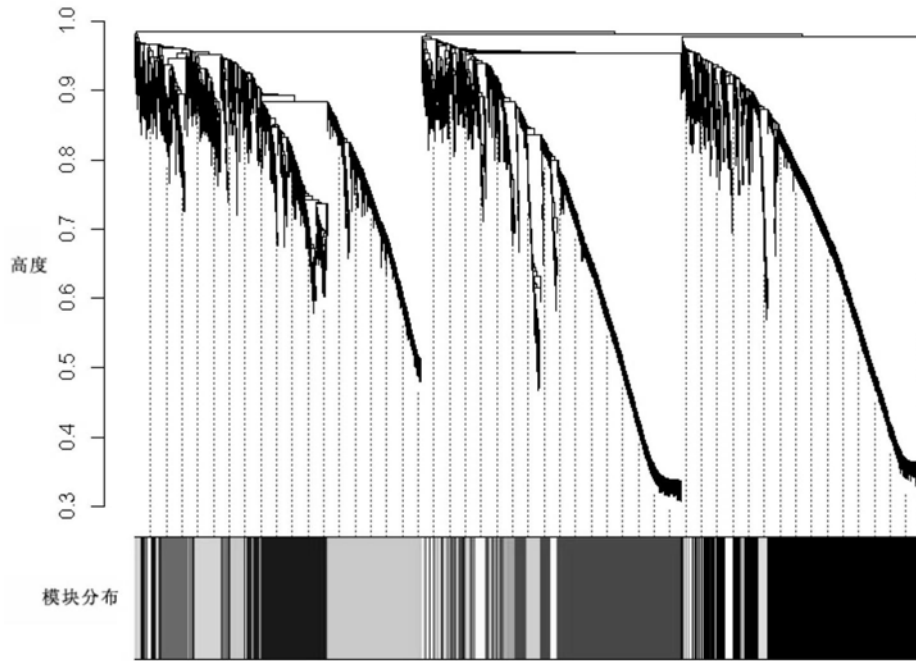


图12

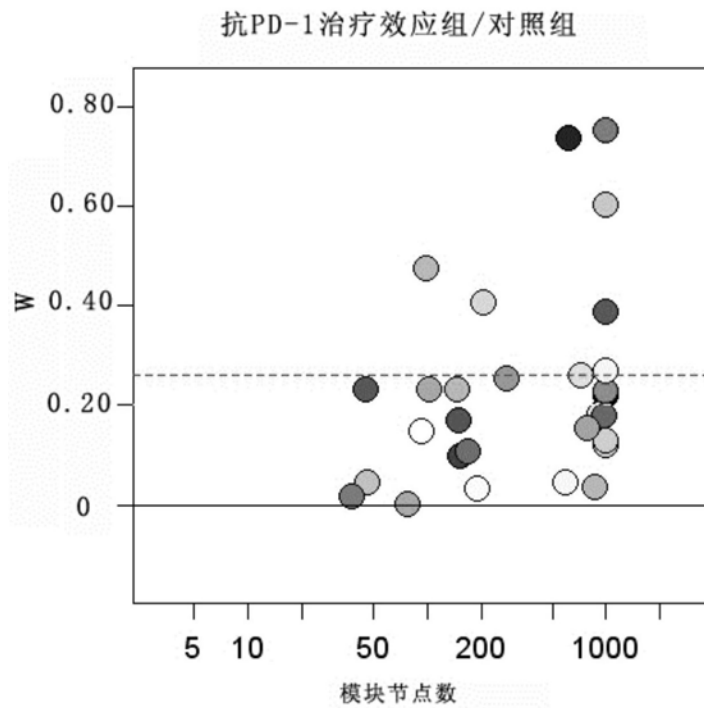


图13

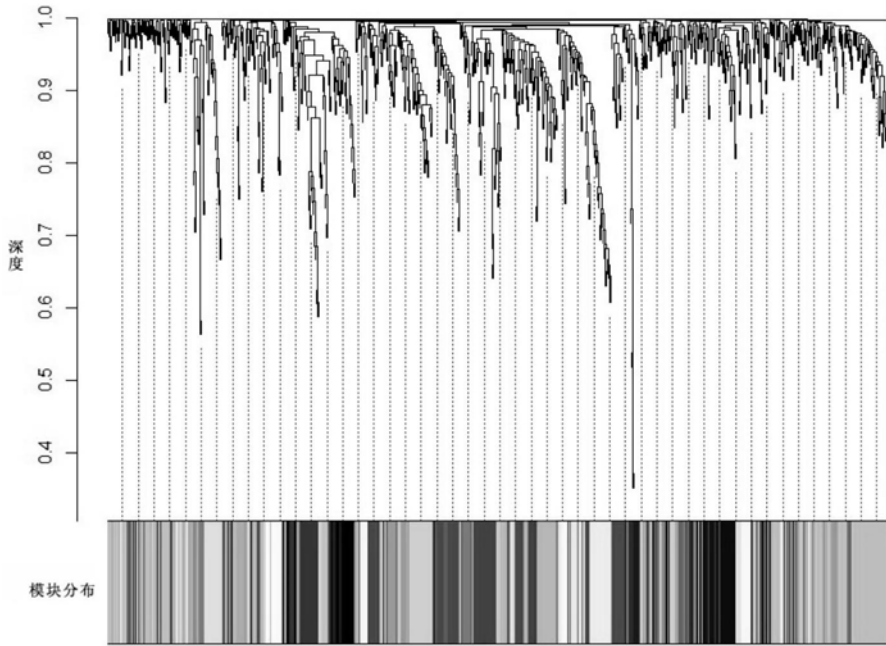


图14

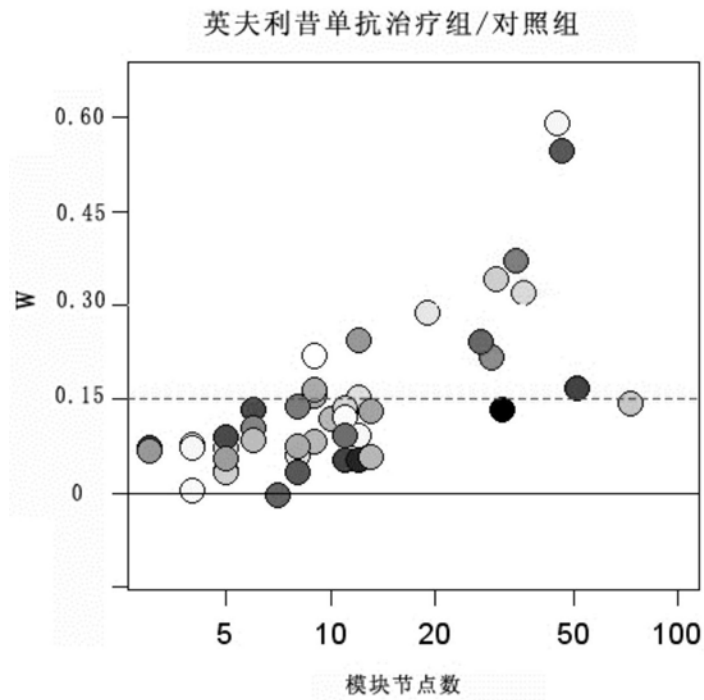


图15