

· 数据挖掘 ·

基于数据挖掘和网络药理学的 2 型糖尿病 用药规律及机制分析

白明, 刘保松, 彭孟凡, 刘田园, 姜权, 苗明三*
(河南中医药大学, 郑州 450000)

[摘要] **目的:**挖掘中医药治疗 2 型糖尿病(T2DM)的组方用药规律,并探讨高频药物作用机制。**方法:**以中国知网、万方医学网收录的期刊文献为资料来源,使用 Excel 2013,SPSS Modeler 14.1,SPSS Statistics 19.0 软件,对纳入标准的中药进行关联规则分析、因子分析和系统聚类分析,使用 BATMAN-TCM 系统对高频药物进行京都基因与基因组百科全书(KEGG)通路分析,构建靶点-通路-疾病关联图,揭示潜在作用机制。**结果:**在纳入标准的 92 首方剂中,单药黄芪最为常见,药物四气以温、寒为主,五味中甘、苦味为主,归经中肝经居多,使用最多的药物类别为补虚药、活血化瘀药。关联规则分析中发现关联强度最高的药对组合为 6 种,因子分析中提取 7 个公因子,系统聚类分析中归为 6 大类。KEGG 通路分析发现高频药物黄芪富集疾病与 T2DM 有关的有 6 种,生物学通路共有 24 条。**结论:**T2DM 的治疗多以甘温药黄芪(补虚药)为主,临床使用可与甘平药茯苓、苦寒药丹参、甘寒药生地黄等组合用药,以及可随证配伍活血化瘀、清热等药物。高频药物黄芪治疗 T2DM 的作用机制可能是通过干预磷脂酰肌醇 3 激酶(PI3K)/蛋白激酶 B(Akt)、淀粉和糖代谢等相关信号通路,作用于核苷酸外焦磷酸酶/磷酸二酯酶(PDEs),腺苷 A1 受体基因 1(ADORA1),腺苷 A2A 受体基因(ADORA2),神经肽第二受体(NPY2R),内皮一氧化氮合成酶(NOS3)靶点有关。

[关键词] 2 型糖尿病(T2DM); 用药规律; 数据挖掘; 中药组方; 网络药理学

[中图分类号] R28;R587.1 **[文献标识码]** A **[文章编号]** 1005-9903(2019)09-0182-07

[doi] 10.13422/j.cnki.sjfx.20190941

[网络出版地址] <http://kns.cnki.net/kcms/detail/11.3495.R.20190117.1349.016.html>

[网络出版时间] 2019-01-18 14:09

Mechanisms and Medication Rules of Type 2 Diabetes Mellitus Based on Data Mining and Network Pharmacology

BAI Ming, LIU Bao-song, PENG Meng-fan, LIU Tian-yuan, JIANG Quan, MIAO Ming-san*
(Henan University of Chinese Medicine, Zhengzhou 450000, China)

[Abstract] **Objective:** To analyze the medication rules of traditional Chinese medicine (TCM) in treating type 2 diabetes mellitus (T2DM), and explore the mechanism of action. **Method:** The articles about TCM treatment for T2DM in CNKI and Wanfang Med Online were retrieved. The data was entered into Excel 2013 to set up a database, and the factor analysis, association rules and cluster analysis were used to analyze core medication combinations by SPSS Statistics 19.0 and SPSS Modeler 14.1. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed using BATMAN-TCM to reveal the mechanism of action. **Result:** A total of 92 prescriptions were included, the high frequency medicines were Astragali Radix. The major properties of TCM were cold and plain. The major flavors of TCM were sweet and bitter. The major meridian tropism of TCM was liver meridian. These high frequency medicines were deficiency-nourishing herbs, and blood circulation-invigorating and stasis-eliminating herbs. Totally 6 core medication combinations were mined through association rules. Seven

[收稿日期] 20181115(021)

[基金项目] 中原学者项目(162101510003);国家国际合作基地项目[国科外函(2016)65号];河南省科技攻关项目(162102310181)

[第一作者] 白明,教授,从事中药药理学教学与研究,E-mail:baiming666@126.com

[通信作者] *苗明三,教授,博士,从事中药药理学教学与研究,E-mail:miaomingsan@163.com

common factors were extracted from factor analysis. Clustering analysis was used to get 6 kinds of herbs. KEGG pathway analysis revealed 24 biological pathways and 6 diseases relating to T2DM. **Conclusion:** The most commonly used Astragali Radix for treating T2DM. In clinical, it can be combined with Poria, Salviae Miltiorrhizae Radix et Rhizoma and Rehmanniae Radix for treating T2DM. At the same time, it can be combined with such herbs as deficiency-nourishing herbs, and blood circulation-invigorating and stasis-eliminating herbs. The mechanism of action of high-frequency herbs may be related to the targets of exophosphatase/phosphodiesterase (PDEs), adenosine A1 receptor gene 1 (ADORA1), adenosine A2A receptor gene (ADORA2A), neuropeptide second receptor (NPY2R), and endothelial carbon monoxide synthase (NOS3) by interfering with PI3K/Akt, starch and sucrose metabolism signaling pathways.

[**Key words**] type 2 diabetes mellitus (T2DM); medication rule; data mining; traditional Chinese medicine (TCM) prescription; network pharmacology

2 型糖尿病 (T2DM) 是有遗传倾向的以血糖偏高为特征的代谢性疾病,其发病率高,病程较长,易引发多种糖尿病并发症^[1]。现代医学认为 T2DM 是一种慢性疾病,一旦确诊,将伴随长期服用降糖药,且病程后期易发肾脏疾病、视网膜病变等多种并发症,由于其病因较多,如遗传、环境、年龄、生活方式等多种因素均可诱发 T2DM,因此临床治疗多仅以降糖为主,只能治标并不能治本。T2DM 属于中医消渴症,《黄帝内经》将其分为“上、中、下消”,随着医学发展,人们对其病因病机的认识逐渐加深,现已形成补虚为主、并兼滋阴降火、活血化瘀等基本治法^[2-5]。中医药传承悠久,数据庞大,运用现代信息技术对已有数据进行加工,挖掘组方规律,发现潜在共性并提出创新,最终再反馈给临床,是现代中医药传承新模式之一,现已有较多研究者将其应用到名老中医经验传承^[6],用药用量规律^[7],脉诊分析^[8]等方面。同时基于现代信息技术将网络药理学与中药研究相结合,也是近年研究热点,充分利用现代信息技术对于中药的靶点发现、机制预测是一种快速、有效的现代手段,为中药创新提供了新的途径^[9-11]。基于此,本文利用现代信息技术,对中医临床用药文献进行研究总结,挖掘中医临床治疗 T2DM 的用药规律及潜在的药物组合,以期发掘新的中药药对和主治疾病,充分发挥中药配伍优势,并挖掘高频中药与疾病相关的作用靶点及相关生物学通路,构建靶点-通路-疾病关联图,发现药物潜在作用机制,为 T2DM 的临床用药提供理论基础。

1 资料与方法

1.1 数据来源 在中国知网数据库“高级检索”的“专业检索”中填入 SU = (‘2 型糖尿病’ + ‘II 型糖尿病’) × ‘中药’ - ‘糖尿病并发症’。时间范围 2010 年 1 月—2018 年 11 月,共检索到 1 268 篇期刊

文献。在万方医学网数据库“高级检索”中填入:(中药)AND 2 型糖尿病 AND II 型糖尿病 NOT 糖尿病并发症。时间范围 2010 年 1 月—2018 年 11 月,共检索到 117 篇期刊文献。

1.2 纳入标准 中药组方治疗 T2DM 的临床研究,名中医用药经验且明确提供方剂组成的文献,文献中有具体全部中药组成的方剂。

1.3 排除标准 剔除重复文献,剔除细胞、动物等基础研究,综述,理论探讨等非临床研究,中西医联合用药,无具体药物或单味中药等文献。

1.4 数据处理 初步筛选出 92 篇期刊文献,将文献中 92 首方剂的基本方进行录入处理,加减方舍去,以此建立中药数据库。药物名称,分类,性味归经等均参照 2015 年版《中国药典》^[12] 和《中药学(成人高等教育药学专业教材)》^[13] 进行统一。

1.5 数据分析 使用 SPSS Modeler 14.1 对纳入标准的中药进行网络化可视分析展示,并利用 Apriori 建模进行关联规则分析;使用 SPSS Statistics 19.0 提取公因子,进行因子分析,以及采用组间联接的方法进行系统聚类分析;使用 BATMAN-TCM 系统^[14] 进行京都基因与基因组百科全书 (KEGG) 通路分析,构建靶点-通路-疾病关联图。

2 结果

2.1 单味中药使用情况 将文献中所涉及到的 92 首方剂中全部中药进行排序,共有 162 种中药,累计频率出现 1 025 次。在 92 首方剂中,使用频数 ≥ 12 的药物共 22 种,累计出现 535 次。使用频数最高的前五味中药分别是,黄芪 (54 次, 58.7%), 丹参 (42 次, 45.6%), 茯苓 (36 次, 39.1%), 生地黄 (35 次, 38.0%), 山药 (31 次, 33.7%)。使用频数 ≥ 12 的药物见表 1。

2.2 药物功效分类使用情况 依据数据库中记录

表 1 单味中药使用情况 (频数 ≥ 12)

Table 1 Application of single traditional Chinese medicine

(frequency ≥ 12)

名称	使用频数	百分比/%	名称	使用频数	百分比/%
黄芪	54	58.7	川芎	21	22.8
丹参	42	45.6	白术	21	22.8
茯苓	36	39.1	红花	19	20.1
生地黄	35	38.0	麦冬	15	16.3
山药	31	33.7	玄参	15	16.3
当归	31	33.7	枸杞子	14	15.2
大黄	30	32.6	山茱萸	14	15.2
泽泻	27	29.3	蒲公英	13	14.1
葛根	27	29.3	菟丝子	13	14.1
天花粉	27	29.3	五味子	13	14.1
党参	25	27.2	苍术	12	13.0

的 92 首方剂,对其全部中药的功效进行分类统计分析。162 味中药涵盖 17 种功效分类,使用频数最多是补虚药 (349 次,34.0%),活血化瘀药 (116 次,11.3%),清热药 (98 次,9.5%),收涩药 (92 次,9.0%)。见图 1。

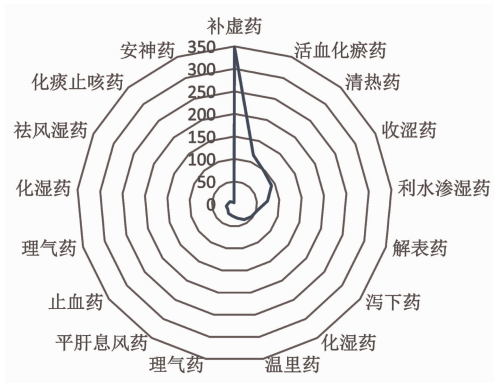


图 1 药物分类归纳统计

Fig.1 Classification and statistics of herbs

2.3 药性药味使用情况 依据《中药学(成人高等教育药学专业教材)》和 2015 年版《中国药典》分类标准,对纳入标准的高频药物(使用频数 ≥ 12)进行药性药味统计分析。如果一味中药有多个药性药味,则全部统计在内。在 22 味中药里,累计出现 34 次药味,可以看出甘 (15 次,40.5%),苦 (8 次,21.6%) 药味最为常见,药性温 (9 次,28.1%),寒 (8 次,25.0%) 居多。见图 2。

2.4 药物归经使用情况 依据《中药学(成人高等教育药学专业教材)》和 2015 年版《中国药典》分类标准,对纳入标准的高频药物(使用频数 ≥ 12)进行药物归经统计分析,如果一味中药有多个归经,则全部统计在内。在总共 22 味中药里,累计出现 59 次

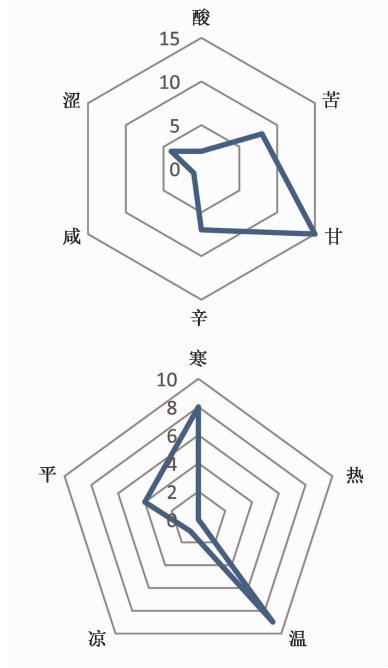


图 2 药物药味药性归纳统计

Fig.2 Statistics flavour and properties of herbs

归经。可以看出肝经 (11 次,18.6%),脾经 (9 次,15.2%),心经 (9 次,15.2%),胃经 (9 次,15.2%) 最为常见。见图 3。

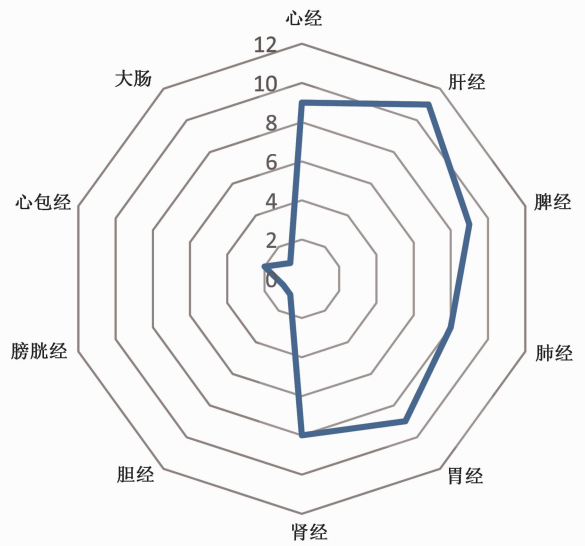


图 3 药物归经归纳统计

Fig.3 Channel tropism and statistics of herbs

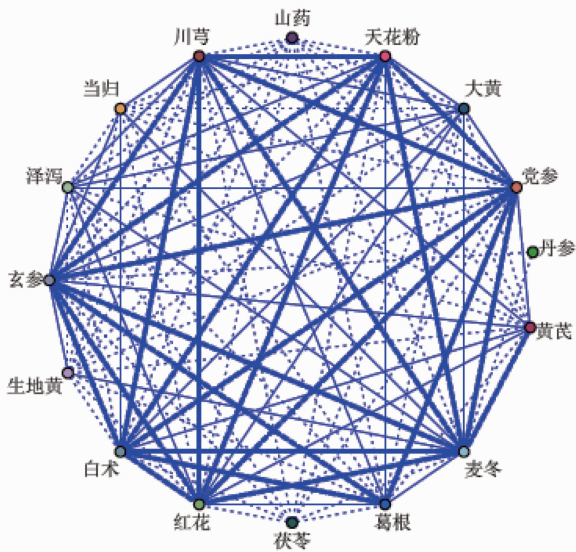
2.5 高频药物关联规则分析 对高频药物(使用频数 ≥ 15),使用 SPSS Modeler 14.1 软件进行关联规则分析^[15],并利用 Apriori 建模进一步挖掘出药物之间的配伍关系,设置支持度为 15%,置信度为 90%,最大前项数为 5,提升 ≥ 1 等条件挖掘出常用方剂中的潜在药物组合,得到核心药对组合共计

6 个。全部药物组合的提升度都大于 1,说明这些药物组合在统计学上均有意义^[16]。以黄芪为核心药物,6 个核心药对组合的统计结果见表 2,关联“网络化展示”见图 4。

表 2 高频药物之间的关联规则分析

Table 2 Analysis of association rules between herbs

药物组合	支持度/%	置信度/%	提升度
黄芪-泽泻-茯苓	17.391	100.000	1.296
黄芪-茯苓-丹参	20.652	94.737	1.228
黄芪-大黄-泽泻	16.304	93.333	1.209
黄芪-白术	15.217	92.857	1.203
黄芪-泽泻-生地黄	15.217	92.857	1.203
黄芪-泽泻	28.261	92.308	1.196



线条粗细程度表示两种药物间关联度的强弱

图 4 高频药物之间关联网络展示

Fig. 4 Association network display between high frequency herbs

2.6 高频药物因子分析 对高频药物(使用频数 ≥ 15),使用 SPSS Statistics 19.0 统计软件进行因子分析^[17],得出 KMO 统计量为 0.546, Bartlett 球形检验, P 值 = 0.000,说明可以做因子分析。采用主成份分析法,提取初始值特征值 > 1 的成份 7 个,因子累计贡献率共 63.914%,即表示可以涵盖大部分信息,通过具有 Kaiser 标准化的全体旋转法得到旋转成分矩阵成分图,见图 5。把药物归入其贡献值最高的公因子内,可得 7 个公因子,见表 3。

2.7 高频药物系统聚类分析 对高频药物(使用频数 ≥ 15),使用 SPSS Statistics 19.0 统计软件进行系统聚类分析^[18],采用组间联接的聚类分析方法,共得到 6 类。类 1,天花粉、川芎。类 2,白术、麦冬、

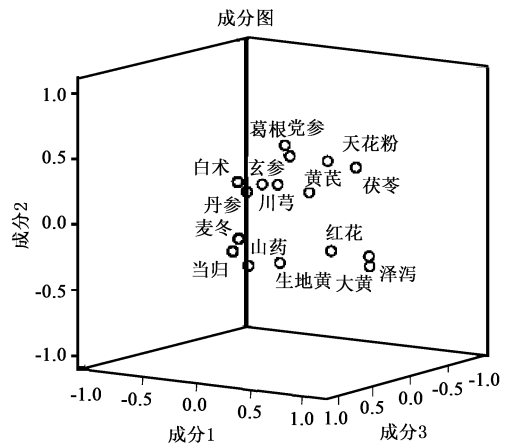


图 5 高频药物之间因子分析成分图

Fig. 5 Factor analysis composition diagram between high frequency herbs

表 3 公因子信息表

Table 3 Information form of commonfactor

公因子	因子贡献率/%	药物
F1	13.201	大黄、泽泻
F2	10.750	葛根、天花粉、党参
F3	10.505	生地黄、山药、当归、川芎
F4	9.073	丹参
F5	7.360	白术、红花、黄芪
F6	6.548	茯苓、麦冬
F7	6.478	玄参

玄参。类 3,党参、红花、葛根。类 4,泽泻、茯苓、大黄。类 5,山药、当归、生地黄。类 6,黄芪、丹参。聚类分析树状图见图 6。

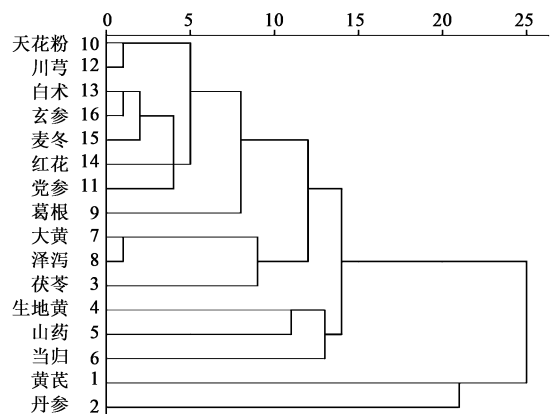


图 6 用药聚类分析树状图

Fig. 6 Dendrogram of cluster analysis

2.8 基于 BATMAN-TCM 分析高频药物黄芪治疗 T2DM 的潜在作用机制^[19-20]

2.8.1 分析高频药物黄芪的疾病富集 在

BATMAN-TCM 系统中,“Score cutoff”选择 80,“Adjusted P-value”选择 0.05,得到高频药物黄芪富集的疾病共 85 种,与 2 型糖尿病最为相关的有 6 种,得到靶点 5 个,相关信号通路 24 条,见表 4。

表 4 高频药物黄芪的疾病富集

Table 4 Enrichment analysis results of Astragali Radix

疾病名称	靶点名称	信号通路个数
糖尿病 (diabetes mellitus)	ENPP1	6
胰岛素抵抗 (肥胖) [insulin resistance (obesity-related)]	ADORA1	2
肾脏疾病 (renal diseases)	ADORA2A	4
肥胖 (obesity)	NPY2R	1
高血压 (hypertension)	NOS3	9
非胰岛素依赖型糖尿病 (noninsulin-dependent diabetes mellitus)	ADORA1	2

2.8.2 分析高频药物黄芪潜在作用信号通路 在 BATMAN-TCM 系统中,通过 KEGG 通路分析得到高频药物黄芪富集的生物学通路共有 71 条,与 T2DM 最为相关的 5 个靶点其连接的代表性信号通路见表 5。黄芪的全部靶点-通路-疾病网络化展示,见图 7。

表 5 KEGG 通路分析结果

Table 5 Analysis results of KEGG pathway

靶点名称	通路 ID	信号通路
ENPP1	hsa00500	淀粉和糖代谢 (starch and sucrose metabolism)
ADORA1	hsa04022	环鸟苷酸/蛋白激酶 (GcGMP/PKG)
ADORA2A	hsa04015	端粒结合蛋白 Rap1 (Rap1 signaling pathway)
NPY2R	hsa04080	神经活性配体/受体相互作用 (neuroactive ligand-receptor interaction)
NOS3	hsa04151	磷脂酰肌醇 3 激酶/蛋白激酶 B (PI3K/Akt)

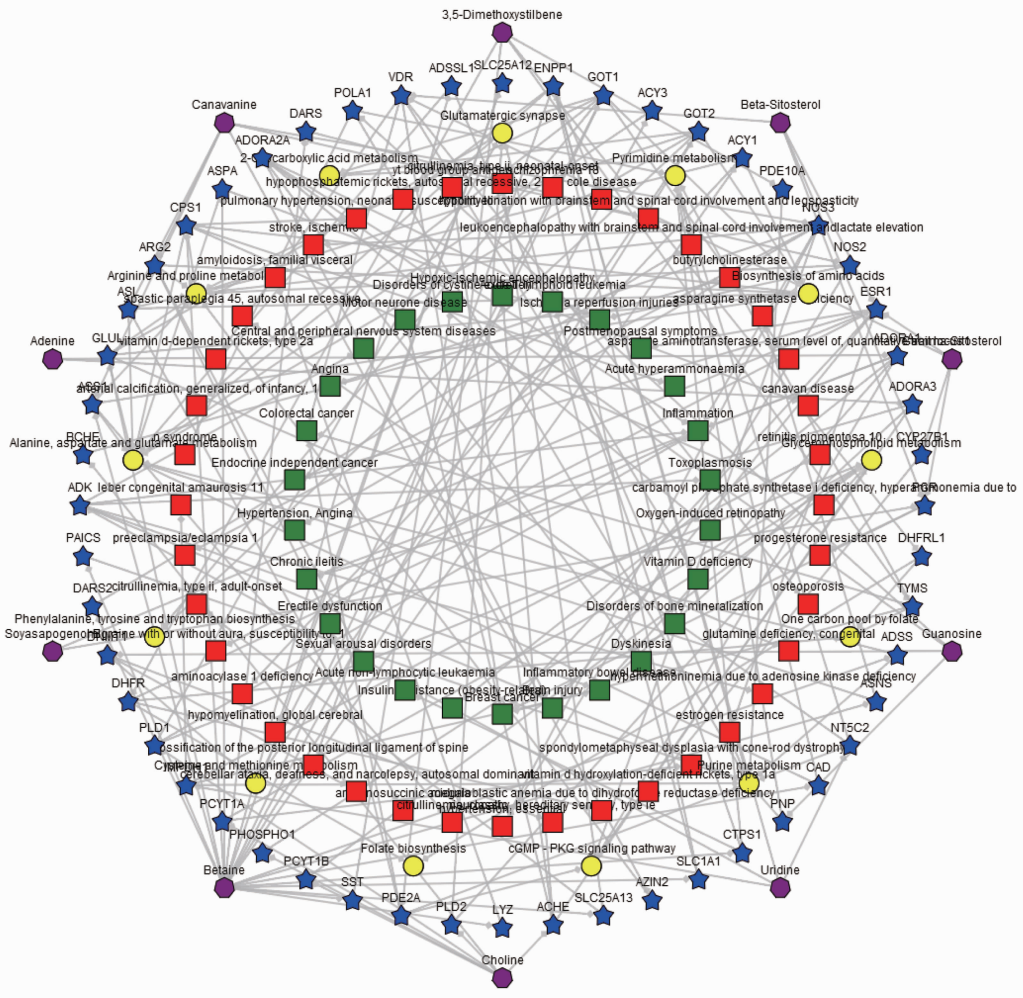


图 7 靶点-通路-疾病网络化展示

Fig.7 Visualization of potential targets, signaling pathways and disease networks

3 讨论

频数分析发现,使用次数最多的单药为黄芪、丹参、茯苓等,药物的四气中温、寒最多,五味以甘味、苦味为主。寒性中药能清热、解毒、泻火等,温性中药能散寒、祛风、温里等,甘具有补、和、缓特性,有补虚、和中的特点,苦具有泄、燥特性,有降泄、通泄、燥湿的特点。使用最多的药物类别为补虚药与活血化瘀药,如黄芪为补虚之圣药,丹参补血活血,川芎活血行气等。综合分析表明,对于治疗 T2DM 的临床用药选择,可注重甘温药物如黄芪,苦寒药物如丹参,甘平药物如茯苓、山药等组合使用,也可兼顾随证配伍一些补虚、活血化瘀、清热等药物。这也符合中医的补虚为主、并兼滋阴降火、活血化瘀等基本治法。

关联规则分析发现,关联强度最强的核心药对组合 6 种,因为置信度能反映出规则预测的准确程度,提示临床用药时应注重黄芪、泽泻、茯苓等药物之间的配伍。如从关联强度较高的药对组合,黄芪-泽泻-茯苓(补虚药-利水渗湿药-利水渗湿药),黄芪-茯苓-丹参(补虚药-利水渗湿药-活血化瘀药)可以看出,对于 T2DM 的研究未来应注重从补虚、利水渗湿与活血化瘀等相互配伍进行研究。

因子分析发现,共分为 7 大类公因子,因子 F1 中含有大黄、泽泻(泻下药-利水渗湿药),因子 F2 中含有葛根、天花粉、党参(解表药-清热药-补虚药),因子 F3 中含有生地黄、山药、当归、川芎(清热药-补虚药-补虚药-活血化瘀药),F4 和 F7 中含有丹参(活血化瘀药)、玄参(清热药),因子 F5 中含有白术、红花、黄芪(补虚药-活血化瘀药-补虚药),因子 F6 中含有茯苓、麦冬(利水渗湿药-补虚药),可见“补虚”药在中医药治疗 T2DM 中发挥重要作用,临床应注重活血化瘀药和清热药类的单独使用,以及补虚药黄芪、党参、白术等与其他类药物的配伍使用。

系统聚类分析发现,采用组间链接的聚类分析方法共得到 6 类。类 1,天花粉(清热泻火、生津止渴)、川芎(活血行气、祛风止痛),提示该类主要适用于燥热阴虚型消渴症,治疗应注重生津活血兼滋阴。类 2,白术(健脾益气、燥湿利水)、麦冬(养阴生津、清心止渴)、玄参(清热凉血、滋阴降火),提示该类主要适用于气阴两虚型消渴症,治疗应注重益气养阴降火。类 3,党参(健脾、益气、生津)、红花(活血、通经、散瘀)、葛根(生津止渴、解肌退热)和类 5,山药(补脾健胃、生津益肺)、当归(活血、调经、

止痛)、生地黄(清热凉血、养阴生津),提示该二类主要适用于脾虚瘀滞型消渴症,治疗应注重活血化瘀,健脾除滞。类 4,泽泻(泄热、化浊、降脂)、茯苓(健脾、宁心、利水渗湿)、大黄(清热泻火、逐瘀通经),提示该类主要适用于湿热中阻型消渴症,治疗应注重清热化湿。类 6,黄芪(补气、生津、行滞)、丹参(通经祛瘀、清心除烦)。提示该类主要适用于血虚气滞型消渴症,治疗应注重活血行气。

通过 BATMAN-TCM 系统进行 KEGG 通路分析,得到高频药物黄芪富集的生物通路共有 71 条,与 T2DM 有关疾病的靶点为 ENPP1, ADORA1, ADORA2A, NPY2R, NOS3, ADORA1, 与 6 个靶点相关的信号通路有 24 条,如 PI3K/Akt, 淀粉和糖代谢等相关信号通路等。提示未来可以从这 6 个靶点及靶点涉及的相关通路对 T2DM 进行进一步的研究。

本文通过分析中医治疗 T2DM 的临床研究文献,初步总结了其用药规律与药物潜在作用机制,期望未来在对 T2DM 药物进行筛选的过程中,可使用该数据挖掘所得潜在药对或单味药进行深层次的研究,也可利用网络药理学对药物进行潜在靶点和信号通路的预测,为研究治疗 T2DM 的药物提供新方向,为临床用药提供参考。

[参考文献]

- [1] Marijana T, Cesare C. Type 2 diabetes mellitus and atrial fibrillation: from mechanisms to clinical practice [J]. Arch Cardiovasc Dis, 2015, 108(4):269-276.
- [2] 田硕,白明,苗明三,等. 姜黄素对大鼠糖尿病模型的影响[J]. 中药药理与临床,2017,33(2):53-55.
- [3] 白宇,贺云,杨丽霞. 活血化瘀类中药治疗糖尿病肾病机制的研究进展[J]. 中国实验方剂学杂志,2018, 24(23):200-206.
- [4] 刘培,王鹏飞,王科,等. 基于 PI3K/Akt 通路的中药治疗糖尿病研究进展[J]. 中国实验方剂学杂志,2019, 25(5):220-228.
- [5] 尹晓阳,张定棋,唐翠娟,等. 基于关联规则和复杂系统熵聚类的糖尿病用药规律研究[J]. 中国实验方剂学杂志,2015,21(23):212-215.
- [6] 丁诺诺,朱才丰,蔡圣朝. 数据挖掘技术在名老中医学学术经验总结中的应用与讨论[J]. 中医药临床杂志,2018,30(10):1779-1782.
- [7] 吴地尧,章新友,甘宇汾,等. 数据挖掘算法在中药研究中的应用[J]. 中国药房,2018,29(19):2717-2722.
- [8] 崔骥,许家佗. 人工智能背景下中医诊疗技术的应用与展望[J]. 第二军医大学学报,2018,39(8):846-851.

- [9] 符宇,范冠杰,黄皓月,等.基于大数据名老中医学学术经验传承研究方法的思考[J].中华中医药杂志,2017,32(4):1644-1646.
- [10] 朱立国,王尚全,于杰,等.试谈名老中医经验传承模式与实践[J].中国中医骨伤科杂志,2014,22(1):70-71.
- [11] 杨磊,刘梦娇,张志国,等.基于计算预测的茵陈蒿汤治疗肝硬化的“成分-靶标-通路”相关性分析[J].中国中药杂志,2018,43(7):1345-1351.
- [12] 国家药典委员会.中华人民共和国药典[M].北京:中国医药科技出版社,2015:3-421.
- [13] 窦昌贵.《中药学(成人高等教育药学专业教材)》[M].上海:上海科学技术出版社,2013:38-98.
- [14] LIU Z, GUO F, WANG Y, et al. BATMAN-TCM: a bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine[J]. Sci Rep, 2016, doi: 10.1038/srep21146.
- [15] 王玲玲,付桃芳,杜俊英,等.基于 SPSS Clementine 的关联规则分析在中医药数据挖掘中的应用优势和局限[J].云南中医学院学报,2016,39(6):98-102.
- [16] 熊平.数据挖掘算法与 Clementine 实践[M].北京:清华大学出版社,2011:75-89.
- [17] 房赤,雍苏南,孙克伟.基于因子分析和关联规则挖掘中医治疗黄疸病用药配伍规律[J].时珍国医国药,2018,29(4):993-995.
- [18] 武凤震,赵文光.基于数据挖掘的癫痫用药规律探索[J].中国实验方剂学杂志,2017,23(8):190-195.
- [19] 严瑶琦,孙静,张莉,等.基于数据挖掘和网络药理学的痛风湿热蕴结证用药规律和作用机制分析[J].中国中药杂志,2018,43(13):2824-2830.
- [20] 段锦龙,姚魁武,冯潇潇,等.基于网络药理学方法分析中药临床治疗胸痹的作用机制[J].中国中药杂志,2017,42(17):3424-3429.

[责任编辑 孙丛丛]