

基于监督学习算法的延胡索成分-靶点-疾病网络的预测研究

苑婕¹,王珍珍²,宋丽娟¹,薛媛¹,张维金¹

1.中国人民解放军联勤保障部队第九八八医院第三派驻门诊部,河南 郑州 450001;

2.新乡医学院第三附属医院肿瘤内科,河南 新乡 453000

【摘要】 目的 应用监督学习算法中的随机森林算法构建药物-靶点模型,预测延胡索治疗心脑血管类疾病的关键靶标。方法 训练集从KEGG数据库获取,包括药物-酶、药物-离子通道、药物-G蛋白、药物-核蛋白四类药靶数据,采用随机森林算法构建药靶模型并预测延胡索体内作用靶点,应用Cytoscape软件构建延胡索成分-靶点-疾病网络;其中模型精度利用十折交叉验证进行评价。结果 四类药靶模型的预测正确率分别是71.36%、67.08%、73.71%、68.22%;延胡索的20个化学成分被预测出作用于ADRA2A、ADRA2C、ADRB2、ADRA1D、ADRB1、ADRA2B、DRD2、CACNA1B、GABRG2、SCN1A等多个与心脑血管疾病相关的靶点并得到较好的文献验证,每个化合物的平均靶点数为9.8。结论 随机森林算法所建模型具有较好的预测正确率,能预测出延胡索治疗心脑血管类疾病的多个关键靶点,其还可用于预测其他中药化学成分的潜在作用靶点。

【关键词】 延胡索;监督学习;网络药理学;心脑血管疾病;肿瘤

【中图分类号】 R93 **【文献标识码】** A **【文章编号】** 1003-6350(2020)13-1638-06

Study on prediction of compound-target-disease network of *Corydalis yanhusuo* based on supervised learning.

YUAN Jie¹, WANG Zhen-zhen², SONG Li-juan¹, XUE Yuan¹, ZHANG Wei-jin¹. 1.The Third Outpatient Department, 988 Hospital of PLA Joint Logistics Support Force, Zhengzhou 450001, Henan, CHINA; 2.Department of Oncology, the Third Affiliated Hospital of Xinxiang Medical University, Xinxiang 453000, Henan, CHINA

【Abstract】 Objective To build the drug-target interaction model based on random forest algorithm of a supervised learning, and to predict key targets of *Corydalis yanhusuo* on cardio-cerebral vascular diseases. **Methods** The datasets of drug compounds and the related enzymes, ion channels, G-protein-coupled receptors, nuclear receptors downloaded from KEGG database were used as the training set. Random forest algorithm was applied to build the drug-target interaction model and to predict the potential targets interacted with the reported compounds of *Corydalis yanhusuo*. Cytoscape was used to construct a compound-target-disease network. The models accuracies were evaluated by 10-fold cross-validation tests. **Results** The overall success rates of the four models were 71.36%, 67.08%, 73.71%, and 68.22%, respectively. The predicted targets of 20 compounds of *Corydalis yanhusuo* on cardio-cerebral vascular diseases, including ADRA2A, ADRA2C, ADRB2, ADRA1D, ADRB1, ADRA2B, DRD2, CACNA1B, GABRG2, SCN1A, were validated by literatures. The average number of targets for each compound was 9.8. **Conclusion** The models established in this paper based on random forest algorithm have a good prediction accuracy, which can predict key targets of *Corydalis yanhusuo* on cardio-cerebral vascular diseases successfully. This method can be used to discover potential targets in other traditional Chinese medicine ingredients.

【Key words】 *Corydalis yanhusuo*; Supervised learning; Network pharmacology; Cardio-cerebral vascular diseases; Tumor

延胡索为罂粟科(Papaveraceae)植物延胡索(*Corydalis yanhusuo* W.T. Wang)的干燥块茎,又名元胡、玄胡等,始载于汉代《神农本草经》,性味温、辛、苦,归心、肝、脾经等^[1]。“心痛欲死,速觅延胡”出自《雷公炮炙论》,《本草纲目》也记载:“活血利气,止痛,通小便”^[2],说明从汉代开始延胡索已经被用于治疗各种疾病,至今已有几千年历史。近现代延胡索的镇痛镇静、扩张冠状血管、改善心肌缺血、抗肿瘤、治疗胃溃疡

等药理作用已得到广泛研究,但其物质基础与作用机制尚待更深入系统的挖掘。

网络药理学是基于系统生物学、高通量组学、计算机模拟计算及医药大数据等多学科理论技术的强大支撑才得以快速开展应用的,其将生物网络与药物作用网络融合一体,从网络平衡的整体性出发进行网络分析,探索药物和机体的复杂相互作用^[3]。中医药整体观、辨证论治的特点、组方配伍的原则,与网络药

通讯作者:张维金,E-mail:yuanjie2364@163.com

理学的整体性和系统性不谋而合。应用网络药理学进行中医药研究不仅能筛选中药有效成分、阐释中药整体作用机制,而且能探究中药复方配伍规律,为中医药理论传承、中药新药创制提供科学依据^[4]。本文以延胡索为研究对象,利用有监督预测分类方法中的随机森林算法^[5]构建延胡索“药物成分-靶标-心脑血管相关疾病”复杂网络,观察延胡索药物作用网络在整体生物网络上的投射,对关键节点进行网络分析,

预测并印证延胡索治疗心脑血管类疾病的有效成分及潜在靶点,为指导临床应用及开发更多候选药物提供参考和思路。

1 材料与方法

1.1 数据集 阳性样本集是从KEGG数据库^[6]直接下载的小分子化学药及其相应靶蛋白受体的分子结构信息,包括四类药物-靶点配对数据,如图1;阴性样本集^[7]获得如图2。



图1 KEGG数据库下载的阳性样本集

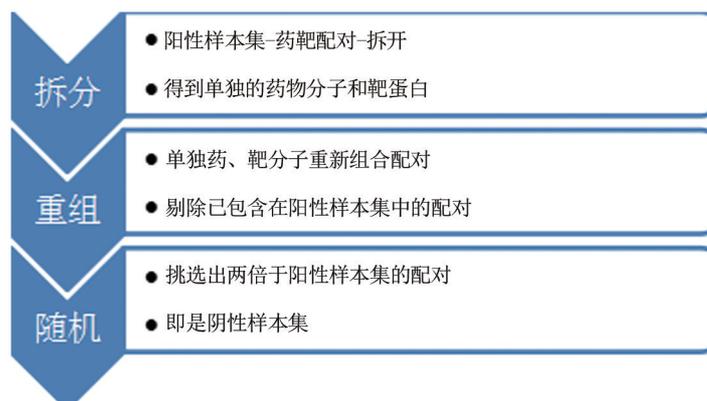


图2 阴性样本集

1.2 分子描述符 计算药物和靶蛋白的分子描述符分别应用PowerMV (www.niss.org/PowerMV)和ProFeat (<http://bidd2.nus.edu.sg/cgi-bin/profeat2016/main.cgi>),共获得7202维的配对数据^[7]。稳定的模型需要具备一定的模型精度和模型运行速度,因此本文采用主成分分析法保留原始数据99.7%的信息量进行降维处理。将本文四组阴、阳数据集分别对应合并组成四组训练集,降维后四组矩阵为8133×17、4095×4、1860×24、258×17;进行算法学习前,再按如下公式进行归一化处理: $x(i)=[x_0(i) - \text{Min}(x_0)] / [\text{Max}(x_0) - \text{Min}(x_0)]$ (式1)。其中 $x_0(i)$ 是原始数据; $x(i)$ 是归一化后数据,其值为-1~1。

1.3 方法 随机森林算法(random forest, RF)最早是由BREIMAN^[8]提出并应用于高维数据处理的一类有监督学习分类器,运用一组分类回归树(Classification and Regression Tree, CART)进行预测,结合了决策树分类模型、Bootstrap aggregating (Bagging)模型组合方法和特征随机选取思想等,具有分类精度高、泛化能力强等优点。随机森林模型建立步骤^[9]如图3。本文采用K折交叉验证(K-fold cross-validation)检验RF的分类能力和预测精度。假设可选的模型为 $S=\{S_1, S_2, S_3, \dots, S_d\}$,本研究采用10折交叉验证($K=10$),将训练集的1/10作为测试集,每个模型训练10次,测试10次,错误率为10次的平均,最终选择平均率最小的模型 S_i 。

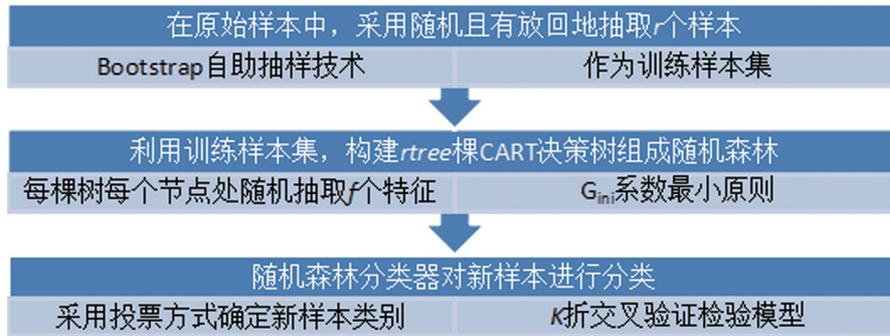


图3 构建随机森林模型

2 结果

2.1 模型优化 在随机森林算法构建过程中, 决策树的个数(*rtree*)和节点分裂过程中抽取的特征个数(*f*)是需要优化的两个关键参数。本文采用网格搜索法对 *rtree* 和 *f* 两参数进行优化, 以获得最高的模型交叉验证正确率(*Ac*), 结果见表 1。 $Ac=(1-WY/Y) \times 100\%$ (式 2)。其中 *WY* 是分类错误的样本个数, *Y* 则是所有样本个数。

表 1 本文模型的参数优化

序号	决策树(棵)	特征数(个)	正确率(%)
1	500	1	71.36
2	400	2	67.08
3	200	1	73.71
4	100	1	68.22

2.2 延胡索作用靶点预测 已报道的延胡索化学成分从化学专业数据(<http://www.organchem.csdb.cn/scdb/>)下载, 共计 20 个。将上述 20 个化合物与四种已知的靶点进行组合并构建预测集, 投入 RF 建立的

分类预测模型前, 先按训练集数据处理方法对预测集计算分子描述符、降维和归一化等处理。表 2 列出了与延胡索化学成分作用频次较高且有代表性的 15 个靶点; 应用 Cytoscape_v3.7.2 软件构建延胡索成分-靶点-疾病网络^[10], 见图 4; 延胡索中 5 号化合物(延胡索乙素)-靶点-疾病网络模型, 见图 5。两个网络中三角形、椭圆形、长方形节点分别代表延胡索化学成分、潜在靶点、相关疾病; 某化合物与潜在作用靶点、靶点与相关疾病, 以边相连。从图 4 可见, 延胡索化学成分与靶点间存在复杂的网络关系; 表 3 中连接度是指药物成分节点与靶点节点所形成的连接边数总和, 即表 3 列出了延胡索 20 个成分在该网络模型中潜在作用靶点数目, 每个化合物的平均靶点数为 9.8; 由于本文篇幅所限, 在此仅以一个化学成分的可视图举例说明, 从图 5 可见, 5 号成分(延胡索乙素)作用于体内 11 个靶点, 此 11 个靶点与图中 44 种疾病相关联。以上网络分析反应了中药与疾病间“多成分-多靶点-多途径”的作用特点。

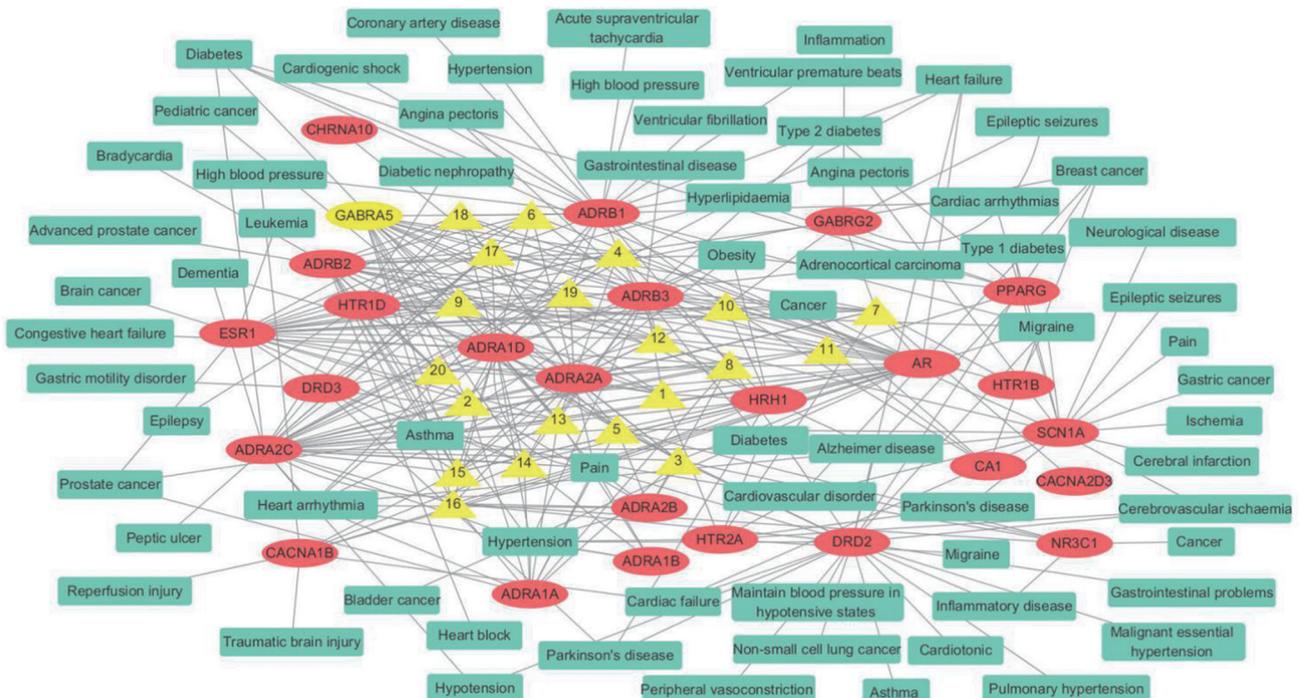


图 4 延胡索 20 个成分-靶点-疾病网络模型

表2 本文预测的延胡索15个潜在作用靶点

序号	靶点名称	名称缩写	频次	类别	心脑血管相关疾病
1	α -2A 肾上腺素受体(alpha-2A adrenergic receptor α -2A)	ADRA2A	20	G 蛋白	心律失常(heart arrhythmia)、肥胖(obesity)、疼痛(pain)、老年痴呆症(dementia)
2	α -2C 肾上腺素受体(alpha-2C adrenergic receptor α -2C)	ADRA2C	19	G 蛋白	高血压(hypertension)、低血压(hypotension)、心律失常(heart arrhythmia)、心力衰竭(heart failure)、老年痴呆症(dementia)、肥胖(obesity)、糖尿病(diabetes)、哮喘(asthma)、疼痛(pain)、阿尔茨海默病(alzheimer disease)、糖尿病肾病(diabetic nephropathy)、消化性溃疡(peptic ulcer)
3	γ -氨基丁酸受体亚单位 α -5 (gamma-aminobutyric acid receptor subunit alpha-5)	GABRA5	17	离子通道	阿尔茨海默病(Alzheimer disease)
4	β -2 肾上腺素受体(beta-2 adrenergic receptor)	ADRB2	17	G 蛋白	心动过缓(bradycardia)、心肌梗死(heart block)、哮喘(asthma)、肥胖(obesity)、高血压(hypertension)
5	α -1D 肾上腺素能受体(alpha-1D adrenergic receptor)	ADRA1D	14	G 蛋白	充血性心力衰竭(congestive heart failure)、高血压(hypertension)、心率失常(heart arrhythmia)、糖尿病(diabetes)、儿科癌症(pediatric cancer)、疼痛(pain)
6	β -1 肾上腺素受体(beta-1 adrenergic receptor)	ADRB1	8	G 蛋白	急性室上性心动过速(acute supraventricular tachycardia)、心绞痛(angina pectoris)、冠状动脉疾病(coronary artery disease)、心率失常(cardiac arrhythmias)、高血压(hypertension)、心源性休克(cardiogenic shock)、心力衰竭(heart failure)、心室早搏(ventricular premature beats)、心室纤颤(ventricular fibrillation)、糖尿病(diabetes)、肥胖(obesity)、偏头痛(migraine)
7	α -1a 肾上腺素能受体(alpha-1A adrenergic receptor)	ADRA1A	7	G 蛋白	高血压(hypertension)、心律失常(heart arrhythmia)、前列腺癌(prostate cancer)、疼痛(pain)
8	α -2B 肾上腺素能受体(alpha-2B adrenergic receptor)	ADRA2B	5	G 蛋白	心律失常(heart arrhythmia)、高血压(hypertension)、疼痛(pain)
9	多巴胺 d2 受体[D(2) dopamine receptor]	DRD2	4	G 蛋白	阿尔茨海默病(Alzheimer disease)、乳腺癌(breast cancer)、哮喘(asthma)、心力衰竭(cardiac failure)、强心剂(cardiotonic)、脑血管缺血(cerebrovascular ischaemia)、低血压(hypotension)、高血压(hypertension)、心力衰竭(heart failure)、恶性原发性高血压(malignant essential hypertension)、偏头痛(migraine)、疼痛(pain)、维持血压在低血压状态(maintain blood pressure in hypotensive states)、非小细胞肺癌(non-small cell lung cancer)、帕金森病(Parkinson's disease)、周围血管收缩(peripheral vasoconstriction)、肺动脉高压(pulmonary hypertension)、胃肠问题(gastrointestinal problems)、炎症性疾病(inflammatory disease)
10	电压依赖型 N 型钙通道 α -1B 亚基(voltage-dependent N-type calcium channel subunit alpha-1B)	CACNA1B	2	离子通道	高血压(high blood pressure)、偏头痛(migraine)、疼痛(pain)、再灌注损伤(reperfusion injury)、创伤性脑损伤(traumatic brain injury)、心力衰竭(cardiac failure)
11	γ -氨基丁酸受体亚单位 γ -2 (gamma-aminobutyric acid receptor subunit gamma-2)	GABRG2	2	离子通道	癫痫发作(epileptic seizures)、炎症(inflammation)
12	钠通道蛋白 1- α 亚基(sodium channel protein type 1 subunit alpha)	SCN1A	2	离子通道	心绞痛(angina pectoris)、心力衰竭(cardiac failure)、心血管疾病(cardiovascular disorder)、脑梗死(cerebral infarction)、心律失常(cardiac arrhythmias)、脑血管缺血(cerebrovascular ischaemia)、缺血(ischemia)、糖尿病(diabetes)、癫痫发作(epileptic seizures)、胃癌(gastric cancer)、心律失常(heart arrhythmia)、偏头痛(migraine)、疼痛(pain)、神经系统疾病(neurological disease)、帕金森病(Parkinson's disease)、1 型糖尿病(type 1 diabetes)
13	α -1b 肾上腺素能受体(alpha-1B adrenergic receptor)	ADRA1B	2	G 蛋白	高血压(hypertension)、疼痛(pain)、心律失常(heart arrhythmia)
14	5-羟色胺 2a 受体(5-hydroxytryptamine receptor 2a)	HTR2A	2	G 蛋白	高血压(hypertension)、心血管疾病(cardiovascular disorder)、脑血管缺血(cerebrovascular ischaemia)、偏头痛(migraine)、糖尿病(diabetes)、帕金森病(Parkinson's disease)
15	过氧化物酶体增殖物激活受体 γ (peroxisome proliferator-activated receptor gamma)	PPARG	1	核蛋白	高脂血症(hyperlipidaemia)、阿尔茨海默病(Alzheimer disease)、糖尿病(diabetes)、2 型糖尿病(type 2 diabetes)、1 型糖尿病(type 1 diabetes)

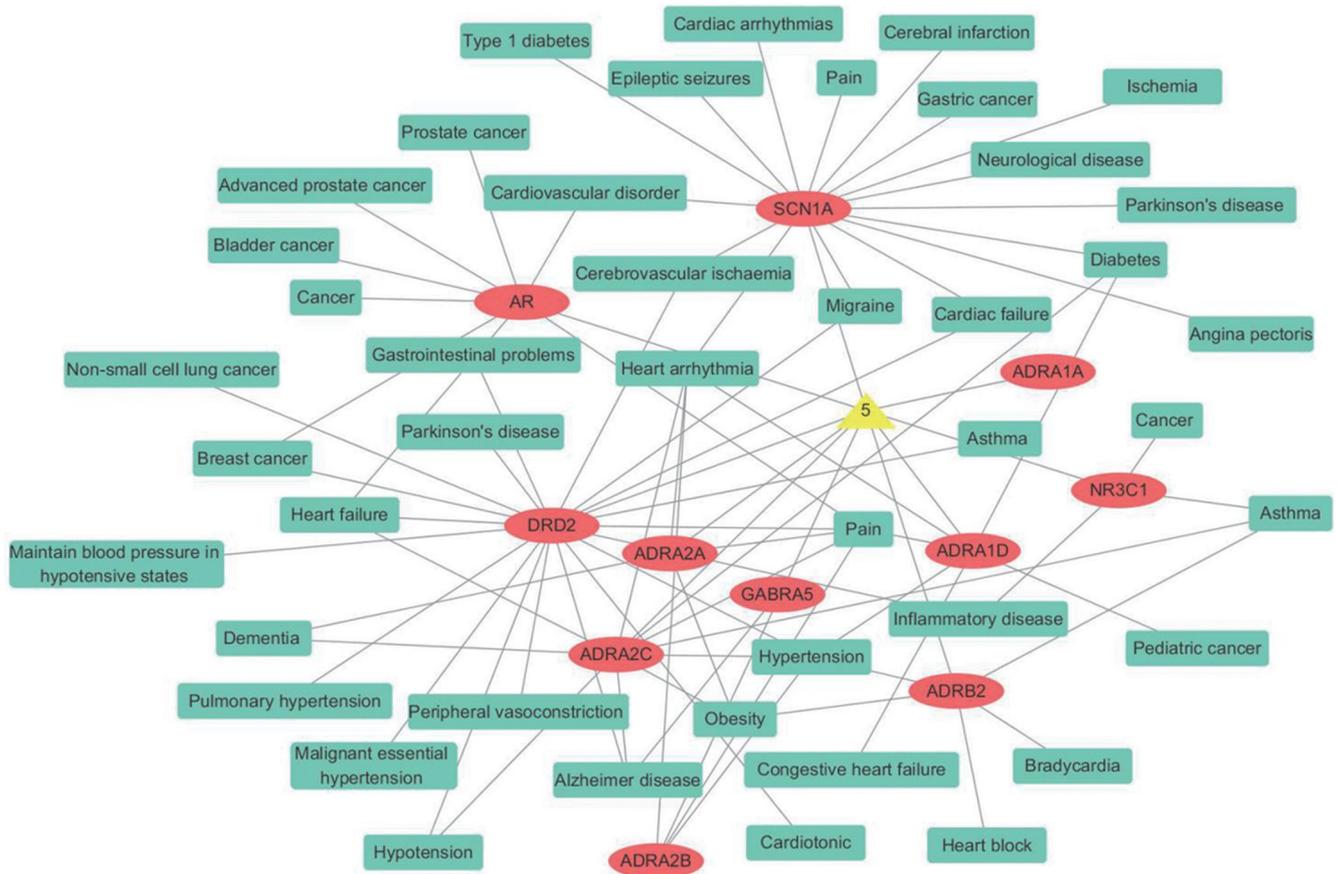


图5 5号化合物(延胡索乙素)-靶点-疾病网络模型

表3 20种化合物的连接度

化合物编号	化合物名称	连接度
1	比枯枯灵碱	10
2	掌叶防己碱(巴马汀)	10
3	去氢海罂粟碱	10
4	四氢黄连碱(延胡索丁素)	11
5	四氢掌叶防己碱(延胡索乙素)	11
6	延胡宁	10
7	原阿片碱(延胡索丙素)	12
8	紫堇碱(延胡索甲素)	8
9	别隐品碱(延胡索癸素)	6
10	异波尔定(碱)	10
11	β -谷甾醇	5
12	非洲防己碱	11
13	南天宁碱	11
14	南天竹种碱甲醚	11
15	狮足草碱	10
16	海罂粟碱	10
17	黄连碱	11
18	隐品碱	9
19	紫堇鳞茎碱	10
20	β -别隐品碱	10

2.3 文献验证 对上文预测与延胡索20种成分相关的心脑血管疾病体内潜在作用靶点进行文献验证。(1)钠离子通道受体(SCN)、钙离子通道受体(CACN)、 γ -氨基丁酸受体(GABA)等是延胡索化学成分在体内主要关联的离子通道受体。姚伟星等^[11]通过

四氢巴马汀(延胡索乙素)的衍生物7-氯苄基四氢巴马汀对多种实验性心律失常模型的研究发现,后者对 Na^+ 通道有阻滞作用,使跨膜动作电位及有效不应期延长是其抗心律失常的重要机制。汪大金等^[12]从临床电生理角度和临床疗效方面证实延胡索的生物碱-左旋四氢巴马汀为较广的抗心律失常谱的钙拮抗剂。张萍等^[13]对前人^[14-15]的研究结果进行总结,得出延胡索碱可作用于钙离子通道的结论,抗心律失常主要是抑制钙电流,降低钙超载。邢淑华等^[16]研究发现左旋延胡索乙素通过阻滞 α 受体、降低外周儿茶酚胺含量改善冠脉血流并有利于实验动物血压降低^[17]。(2)肾上腺素受体(ADR)、5羟色胺受体(5-HT)、多巴胺受体(DA)等可能是G蛋白受体主要潜在作用靶标;过氧化物酶体增殖物激活受体 γ (PPARG)是网络分析所得的核蛋白受体潜在的作用靶点。利用网络药理学技术,虞希冲等^[18]预测出延胡索生物碱可调控21种糖尿病及其并发症的34个靶点。原阿片碱-PPARG、小檗碱-AKT、巴马汀-KCNJ11、四氢巴马汀-NOS3等药靶与延胡索抗非胰岛素依赖型糖尿病和糖尿病(未分型)的作用有关;延胡索调控高血压与冠心病则主要通过紫堇碱-ADRA1、四氢小檗碱-DRD1、海罂粟碱-SELE、四氢巴马汀-NOS3等药靶实现。韩彦琪等^[19]通过受体实验得出结果,延胡索可能是通过激动ADRB2、抑制M2和TP受体发挥活血、行气、化痰功

效的。ADRB2受体是与心血管、哮喘等疾病相关的重要靶标之一。原阿片碱和延胡索乙素是延胡索潜在重要的药效物质基础,对D2受体有拮抗作用。张铁军等^[20]研究发现延胡索和白芷配伍后发挥多种生物活性、产生增效作用是通过激活5-HT_{1A}、OPRM₁、ADRB2受体,抑制D₂、M₂和TP受体来调节下游生物通路实现的。胡江元^[21]研究发现延胡索乙素能优先阻滞位于脑区纹状体等的D₂受体,进而加强脑干下行控制的痛觉调制系统的抗痛功能。

除心脑血管类疾病外,所得靶点与其他一些重要疾病也被预测出相关性并得到了文献验证。如LEUNG等^[22]在进行高架十字迷宫模型研究时发现,延胡索的抗焦虑作用,至少部分与GABAA受体有关。延胡索的抗实验性胃溃疡作用,推测机制:一是与其增加胃黏膜血流量有关;二是通过部分阻滞胃黏膜DA受体实现^[23]。徐靖宇等^[24]将吗啡依赖胃肠损伤大鼠作为实验模型,研究延胡索和左旋延胡索乙素保护吗啡依赖继发性胃肠损伤,结论是两者能逆转胃肠多巴胺和D₂R的异常减少和增加。

3 讨论

延胡索作为一味常用传统中药材,以活血化瘀、利气止痛为主。现代药理实验研究证明对心脑血管系统、神经系统、消化系统和内分泌系统等均有明确的药理作用。从延胡索的研究现状中可以看出,对叔胺碱类如延胡索乙素的研究较多,其主要起镇痛作用;季胺碱类如紫堇碱等研究较少,可抗心肌缺血等,其作用机制鲜有报道。本文运用网络药理学的方法使延胡索20种成分-靶点-疾病网络可视化并对网络进行拓扑研究,分析关键节点、预测潜在靶标、联系相关疾病,并得到了较好的文献验证,为延胡索治疗心脑血管等疾病的物质基础及分子作用机制提供了理论依据。本研究方法为虚拟筛选中药药效物质成分及潜在作用靶点,及应用现代科学方法研究中药作用机制提供一定的参考思路。

参考文献

[1] 国家药典委员会. 中华人民共和国药典2015年版一部[S]. 北京: 中国医药科技出版社, 2015: 139.
 [2] 李时珍. 本草纲目(上)[M]. 北京: 人民卫生出版社, 2006: 555.
 [3] 潘家祐. 基于网络药理学的药物研发新模式[J]. 中国新药与临床杂志, 2009, 2(10): 721-726.
 [4] 张彦琼, 李梢. 网络药理学与中医药现代研究的若干进展[J]. 中国

药理学与毒理学杂志, 2015, 29(6): 883-892.

[5] 王奕森, 夏树涛. 集成学习之随机森林算法综述[J]. 信息通信技术, 2018, 12(1): 49-55.
 [6] KANEHISA M, GOTO S, HATTORI M, et al. From genomics to chemical genomics: new developments in KEGG [J]. Nucleic Acids Res, 2006, 34(Database issue): D354-D357.
 [7] 苑婕. 心脑血管类中药的网络药理学初步研究[D]. 广州: 广东药学院, 2014: 20-23.
 [8] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32.
 [9] GISLASON PO, BENEDIKTSSON JA, SVEINSSON JR. Random forests for land cover classification [J]. Pattern Recognition Letters, 2006, 27(4): 294-300.
 [10] 苑婕, 李晓杰, 陈超, 等. 基于随机森林算法的川芎成分-靶点-疾病网络的预测研究[J]. 中国中药杂志, 2014, 39(12): 2336-2340.
 [11] 姚伟星, 夏国瑾, 曾维忠, 等. 7-氯苄基四氢巴马汀的抗心律失常作用[J]. 中国药理学与毒理学杂志, 1995, 9(3): 220-223.
 [12] 汪大金, 毛焕元, 黄从新. 植物性钙拮抗剂频通定的抗心律失常临床疗效及其电生理作用[J]. 同济医科大学学报, 1993, 22(5): 322-324.
 [13] 张萍, 徐凤芹, 马晓昌, 等. 延胡索碱治疗快速性心律失常的研究进展[J]. 中国中西医结合杂志, 2012, 32(5): 713-716.
 [14] 程龙献, 毛焕元, 戴水平, 等. 频通定对单个豚鼠心室肌细胞膜钙通道电流的影响[J]. 临床心血管病杂志, 1995, 11(3): 170-172.
 [15] 赵昕, 汤浩, 王亚杰, 等. 脱氢紫堇碱对正常和低氧豚鼠心肌细胞内钙的影响[J]. 中国应用生理学杂志, 2003, 19(3): 17-20.
 [16] 邢淑华, 郑加麟, 卞春甫, 等. 左旋四氢巴马汀的降压效应及对冠脉循环的影响[J]. 中国现代应用药学, 1999, 16(2): 8-10.
 [17] 李清福, 梁西, 董蜀华. 左旋四氢巴马汀对心脏及血压的影响[J]. 西南国防医药, 2001, 11(4): 303-304.
 [18] 虞希冲, 杨伟, 吴波拉, 等. 网络药理学技术预测延胡索抗糖尿病作用及初步验证研究[J]. 中国药理学杂志, 2014, 49(11): 913-918.
 [19] 韩彦琪, 许浚, 龚苏晓, 等. 基于HPLC-QTOF/MS及G蛋白偶联受体分析的延胡索物质基础及作用机制研究[J]. 药学报, 2016, 51(8): 1302-1308.
 [20] 张铁军, 许浚, 申秀萍, 等. 基于中药质量标志物(Q-Marker)的元胡止痛滴丸的“性-效-物”三元关系和作用机制研究[J]. 中草药, 2016, 47(13): 2199-2211.
 [21] 胡江元. 左旋四氢巴马汀镇痛作用的机制研究[D]. 上海: 中国科学院上海药物研究所, 1999.
 [22] LEUNG W, ZHENG H, HUEN M, et al. Anxiolytic-like action of orally administered dl- tetrahydropalmatine inelevated plus-maze [J]. Prog Neuropsychopharmacol Biol Psychiatry, 2003, 27(5): 775-779.
 [23] 贺凯, 高建莉, 赵光树. 延胡索化学成分、药理作用及质量控制研究进展[J]. 中草药, 2007, 38(12): 1909-1912.
 [24] 徐靖宇, 白威峰, 丘成楷, 等. 延胡索和L-THP对吗啡依赖大鼠胃肠多巴胺系统的影响[J]. 中药材, 2015, 38(12): 2568-2572.

(收稿日期: 2020-01-16)