

- [9] LEE HW, LEE YJ, KIM SJ, et al. Comparing tuberculin skin test and interferon γ release assay (T-SPOT.TB) to diagnose latent tuberculosis infection in household contacts [J]. Korean J Intern Med, 2017, 32(3): 486-496.
- [10] 陈芳, 项忠元. T淋巴细胞干扰素释放试验在检测结核杆菌感染中的应用[J]. 实用预防医学, 2017, 24(3): 272-274.
- [11] OZEKINCI T, OZBEK E, CELIK Y. Comparison of tuberculin skin test and a specific T-cell-based test, T-Spot.TB, for the diagnosis of latent tuberculosis infection [J]. J Int Med Res, 2007, 35(5): 696-703.
- [12] 王东, 陆普选, 袁虹, 等. HRCT检查在泌尿系结核影像诊断与指导治疗中的应用价值[J]. 首都医科大学学报, 2016, 37(4): 477-480.
- [13] 游瑞雄, 曹代荣, 李银官, 等. 多层螺旋CT曲面重建在诊断泌尿系结核的价值[J]. 中国医学影像学杂志, 2010, 18(2): 115-118.
- [14] 陈义加, 王志娟, 陈伦刚, 等. CTU在泌尿系结核的临床应用价值[J]. CT理论与应用研究, 2014, 23(6): 995-1000.
- [15] 于澄钊, 张弋, 金石华, 等. 输尿管肾盂镜检是否适用于泌尿系结核的诊断和治疗?[J]. 现代泌尿外科杂志, 2015, 20(4): 269-272, 292.
- [16] GUO NY, KIM WJ, JIN HR, et al. Nerve injury-induced protein 1 (Ninjurin-1) is a novel therapeutic target for cavernous nerve injury-induced erectile dysfunction in mice [J]. J Sex Med, 2013, 10(6): 1488-1501.

(收稿日期: 2019-06-01, 修回日期: 2019-08-15)

doi: 10.3969/j.issn.1009-6469.2020.08.015

◇临床医学◇

数据挖掘技术在雌激素对心血管的分子调控网络中的应用

李艳红^{1a}, 杨亚莉^{1a}, 陈芬², 万翔^{1b}, 龚芳³作者单位:¹长江大学附属仙桃市第一人民医院,^a心电图室,^b心内科,湖北 仙桃 433000;²华中科技大学同济医学院附属协和医院心内科,湖北 武汉 430022;³襄阳市中心医院,湖北文理学院附属医院心内科,湖北 襄阳 441021

通信作者: 陈芬, 女, 主任医师, 研究方向为心血管疾病诊疗, E-mail: 54113447@qq.com

基金项目: 湖北省自然科学基金(2016CFB344)

摘要: **目的** 探讨数据挖掘技术在雌激素对心血管的分子调控网络中的应用。 **方法** 利用GEO2R在线分析工具进行差异基因分析, 采用DAVID 6.8进行基因功能和通路富集分析, 采用STRING数据库进行蛋白互作网络分析, 采用MCODE算法提取核心网络。 **结果** 共筛选出192个差异基因, 将差异基因进行GO与信号通路富集分析后发现, 这些基因与脂代谢、血管生成、PPAR信号通路、脂肪酸降解信号通路等有关。蛋白互作网络分析得到一个包含131基因的互作关系网络, MCODE算法分析提取出其中5个核心子网络, 分别与胆固醇平衡、胰岛素分泌、铁离子平衡、血管新生以及生长激素受体信号通路等生物学功能和信号通路有关。 **结论** 应用数据挖掘技术可以发现雌激素对心血管系统的影响与脂质代谢、胰岛素分泌、血管新生和细胞生长等途径有关。

关键词: 雌激素类/生理学; 数据挖掘; 基因调控网络; 心血管; GEO数据库; 蛋白互作网络

Application of data mining technology in molecular regulation network of estrogen on cardiovascular

LI Yanhong^{1a}, YANG Yali^{1a}, CHEN Fen², WAN Xiang^{1b}, GONG Fang³

Author Affiliations: ^{1a}ECG Room, ^{1b}Department of Cardiology, Xiantao First People's Hospital Affiliated to Yangtze University, Xiantao, Hubei 433000, China; ²Department of Cardiology, Union Hospital Affiliated to Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430022, China; ³Department of Cardiology, Xiangyang Central Hospital, Affiliated Hospital of Hubei College of Arts and Sciences, Xiangyang, Hubei 441021, China

Abstract: **Objective** To explore the application of data mining technology in molecular regulation network of estrogen on cardiovascular. **Methods** The online analysis tool GEO2R was used to analyze differential gene, DAVID 6.8 was used for gene function and pathway enrichment analysis, STRING database was used for protein interaction network analysis, and the core network was extracted by MCODE algorithm. **Results** A total of 192 differential genes were screened. GO and signaling pathway enrichment analysis found that these genes were related to lipid metabolism, angiogenesis, PPAR signaling pathway, and fatty acid degradation signal-

ing pathway. Protein-protein interaction network analysis obtained an interaction network containing 131 genes, and MCODE algorithm analysis extracted five core subnetworks, which were respectively related to cholesterol balance, insulin secretion, iron ion balance, angiogenesis and growth hormone receptor signaling pathway. **Conclusion** The effects of estrogen on cardiovascular system are related to lipid metabolism, insulin secretion, angiogenesis and cell growth by using data mining technology.

Key words: Estrogens/physiology; Data mining; Gene regulatory networks; Cardiovascular; GEO database; Protein-protein interaction network

心血管疾病(cardiovascular diseases, CVDs)是世界范围内常见的高病死率疾病,严重威胁人们的生命健康^[1]。与男性相比,女性病人的发病年龄虽然较晚,但病死率及并发症的发生率均高于男性^[2]。流行病学研究发现绝经前女性心血管疾病发病率较低,绝经后明显增高,这是由于女性体内的雌激素可以影响血脂代谢,对心血管系统具有一定的保护作用^[3-4]。此外,有研究指出在时间窗口期应用雌激素对心血管系统具有保护作用。但是,也有研究^[5-6]发现应用口服雌孕激素替代治疗对心血管并没有保护作用。因此,关于雌激素对心血管系统的作用引起广泛的关注,成为目前研究的热点。本研究2019年1—3月通过筛选雌激素作用后的差异基因,进一步明确雌激素在心血管系统中的调控靶点及作用网络,从而为进一步明确雌激素在心血管系统中的作用提供方向。

1 资料与方法

1.1 数据检索 通过NCBI GEO Datasets 检索雌激素在心血管中的相关研究,筛选出GSE9371数据集^[7]。该数据来源于美国波士顿大学生物部门Hansens实验室,研究者将小鼠分为安慰剂组与雌激素组,每组5个重复,在接受长期(1周)的雌激素治疗后,提取了小鼠主动脉组织中的总RNA进行了基因表达谱分析,以确定长期雌激素治疗对综合基因组的表达的调控。样本数据信息见表1。

表1 雌激素在心血管中的相关研究GSE9371
数据样本信息

样本编号	分组	平台	种属	性别	周龄 范围	样本 来源
GSM238478	安慰剂组1	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238479	安慰剂组2	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238480	安慰剂组3	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238481	安慰剂组4	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238482	安慰剂组5	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238483	雌激素组1	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238484	雌激素组2	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238485	雌激素组3	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238486	雌激素组4	GPL1261	C57BL/6	雌性	10~18	总RNA
GSM238487	雌激素组5	GPL1261	C57BL/6	雌性	10~18	总RNA

1.2 差异表达基因分析 采用NCBI GEO2R 在线分析工具^[8]进行差异表达基因的分析,以安慰剂组

为对照组,雌激素组为干预组。将差异表达倍数大于2倍,同时 $P < 0.05$ 的基因筛选为差异表达基因。

1.3 GO富集分析及作图 采用DAVID 6.8从生物学过程,细胞组分和分子功能三个方面进行GO富集分析^[9],以 $P < 0.05$ 筛选为差异有统计学意义的GO富集结果。采用R语言ggplot2程序包进行可视化绘图。

1.4 通路富集分析及作图 采用DAVID 6.8在REACTOME, BIOCARTA, KEGG_pathway 三个通路数据库中,进行KEGG富集分析^[9],以 $P < 0.05$ 筛选为差异有统计学意义的通路富集结果。采用R语言ggplot2程序包进行可视化绘图^[10]。

1.5 蛋白质相互作用网络分析 采用STRING数据库^[11]对筛选出的差异基因进行蛋白质相互作用的分析,隐藏不存在相互作用的节点,最小相互作用证据评分设置为0.4分。此外,采用R语言统计与其它蛋白相互作用关系最多的节点,取前30个进行可视化作图。采用Cytoscape中的MCODE算法分析^[12],提取其中的核心基因模块,同样利用DAVID 6.8对核心模块基因进行GO富集分析及通路富集分析。

2 结果

2.1 差异表达基因的分析 将雌激素组与安慰剂组基因表达水平进行比较,按照差异表达倍数大于2倍,同时 $P < 0.05$ 进行差异表达基因的筛选,共筛选到差异表达基因192个,按照 P 值排序,其中前10位的基因为Grem2、Ano4、Mmd2、Apod、Galnt15、Heph、Vldlr、Stc1、Vwf和Zmiz1,均为上调表达(表2)。

表2 雌激素在心血管中的相关研究中雌激素组
与安慰剂组前10位差异表达基因

基因名	安慰剂组 表达值	雌激素组 表达值	差异倍数 log2值	P 值
Grem2	4.526 493	9.171 912	4.645 42	0.000 108
Ano4	2.506 665	6.799 716	4.293 05	0.000 150
Mmd2	3.331 12	6.300 121	4.435 17	0.000 171
Apod	4.219 454	4.971 080	1.491 66	0.001 578
Galnt15	5.851 995	8.273 049	2.264 98	0.004 365
Heph	5.572 095	7.611 388	2.039 29	0.004 423
Vldlr	7.948 688	9.354 011	1.574 17	0.004 423
Stc1	3.883 332	4.962 789	2.115 11	0.015 263
Vwf	8.805 763	10.73 562	1.929 86	0.020 589
Zmiz1	4.570 739	5.736 501	1.282 93	0.026 412

2.2 基因功能富集分析 基因功能富集分析将上述 192 个差异基因映射到生物学过程,细胞组分和生物学功能三个基因功能数据库中,以 $P < 0.05$ 筛选有统计学意义的富集结果,共筛选到 100 个显著相关的基因功能条目。其中 59 个(59%)条目与生物学过程有关,涉及脂代谢过程、巨噬细胞活化调控过程、红细胞分化、脂肪细胞分化调控过程、视黄醇代谢过程、血管生成调控、细胞生长调节以及细胞黏附等生物学功能。20 个(20%)条目与细胞组分有关,主要与胞外区、细胞外隙、蛋白质的细胞外基质、细胞外分泌体、线粒体等有关。21 个(21%)条目与分子功能有关,涉及的分子功能有 Wnt 蛋白结合、肝素结合、脂质结合、胰岛素样生长因子结合、RNA 聚合酶 II 转录因子活性等有关。按照 P 值排序,其中前 10 位的基因功能富集结果条目见图 1。

2.3 通路富集分析 通路富集分析将上述 192 个差异基因映射到 BBID、BIOCARTA 和 KEGG_pathway 三个通路数据库中,以 $P < 0.05$ 筛选有统计学意义的富集结果,共筛选到 13 个显著富集的信号通路,分别是乙醇氧化信号通路、PPAR 信号通路、脂肪酸降解信号通路、视黄醇类代谢和运输信号通路以及红细胞吸收氧气并释放二氧化碳信号通路等有关,其中前 10 位的通路富集结果条目见图 2。

2.4 蛋白质相互作用分析 我们采用 STRING 数据库对 192 个差异基因进行蛋白质(基因)互作网络分析,结果发现其中 131 个基因构成了互作网络(图 3)。对网络中每个基因的互作基因的个数进行统计后发现,与 Col1a1、Ppargc1a、Cdh2、Ppara、Thbs1 等基因存在相互的基因较多,其中前 30 位的互作基因见图 4。随后我们使用 MCODE 算法进一步提取了其中的核心子网络,结果发现该蛋白质互作网络中有 5 个核心子网络(图 5),对每个核心子网络的基因进行富集分析后发现,核心子网络 1 主要与细胞外基质、胆固醇平衡、细胞外间隙、内质网以及 PPAR 信号通路有关。核心子网络 2 主要与钙调蛋白结合和胰岛素分泌等有关,核心子网络 3 主要与铁离子平衡和红细胞分化有关,核心子网络 4 主要与血管新生的调控有关,核心子网络 5 主要与生长激素受体信号通路和 Jak-STAT 信号通路有关(表 3)。

3 讨论

近年来,我国心血管疾病的发病率及死亡率均持续上升,其死亡率居于首位,同时其在居民死亡的疾病构成中占比达到 40% 以上,严重威胁居民的生命健康^[13]。就性别而言,女性心血管疾病病人的

表 3 核心子网络基因功能及通路富集分析结果

核心子网络	数据库	基因功能与通路	基因数	P值
核心子网络 1	GO_CC	extracellular matrix	6	0.000
	GO_BP	cholesterol homeostasis	4	0.000
	GO_CC	extracellular region	9	0.000
	GO_CC	endoplasmic reticulum	8	0.000
	GO_CC	proteinaceous extracellular matrix	5	0.000
	KEGG	PPAR signaling pathway	4	0.000
	KEGG	ECM-receptor interaction	4	0.000
	KEGG	Focal adhesion	4	0.000
核心子网络 2	GO_MF	calmodulin binding	3	0.000
	GO_BP	long-term synaptic potentiation	2	0.000
	KEGG	Gastric acid secretion	2	0.000
	KEGG	Insulin secretion	2	0.000
	KEGG	Aldosterone synthesis and secretion	2	0.000
核心子网络 3	GO_BP	iron ion homeostasis	3	0.000
	GO_BP	erythrocyte differentiation	3	0.000
	GO_BP	response to arsenic-containing substance	2	0.000
	KEGG	Porphyrin and chlorophyll metabolism	2	0.000
核心子网络 4	GO_BP	positive regulation of angiogenesis	2	0.000
核心子网络 5	GO_BP	growth hormone receptor signaling pathway	2	0.000
	GO_BP	negative regulation of insulin receptor signaling pathway	2	0.000
	GO_BP	regulation of multicellular organism growth	2	0.000
	REACTOME	Growth hormone receptor signaling	3	0.000
	KEGG	Jak-STAT signaling pathway	3	0.000

发病年龄虽然较晚,但病死率及并发症的发生率均高于男性^[2]。流行病学研究发现雌激素可以影响血脂代谢,对心血管系统具有一定的保护作用^[3-4]。既往研究发现雌激素可以通过多种途径保护心血管系统,雌激素可以调节平滑肌细胞的收缩性和基质形成,抑制平滑肌和心肌细胞的凋亡,抑制内质网应激引起的凋亡保护血管内皮细胞^[14-16]。但美国妇女健康倡议活动(WHI)研究^[6]发现应用口服雌孕激素替代治疗对心血管并没有保护作用,甚至会增加绝经后女性患冠心病的风险。因此,关于雌激素是否可以用于女性心血管疾病的治疗以及雌激素对心血管系统的作用机制引起广泛的关注,成为目前研究的热点。

本研究通过对GEO数据库的挖掘,分析了雌激素作用后血管总RNA中存在显著差异表达的基因,通过基因功能与通路富集分析,发现这些基因的功能与脂代谢过程、巨噬细胞活化、脂肪细胞分化、血管生成、线粒体、胰岛素样生长因子等有关(图1),涉及的信号通路有乙醇氧化信号通路、PPAR信号通路、脂肪酸降解信号通路等(图2)。说明雌激素可能通过这些基因功能与信号通路影响脂质代谢、血管新生以及细胞生长等生物学功能。蛋白互作网络分析发现,192个差异基因中有131个基因间存在相互作用关系,构成了互作网络(图3)。其中核心的子网络有5个,这些核心子网络的功能主要与细胞外基质、胆固醇平衡、内质网、钙调蛋白、胰岛素分泌、血管新生、生长激素受体信号通路和Jak-STAT信号通路有关(表3)。核心子网络富集到的基因功能与信号通路是所有基因富集后的结果基本一致,更加证明了雌激素可能通过这些基因功能与信号通路对心血管的功能产生影响。对上述雌激素对心血管系统可能的调控途径进一步分析后发现,雌激素可能通过调节脂代谢^[3-4],促进巨噬细胞活化^[17]、促进血管内皮细胞生长^[18],激活胰岛素样生长因子^[19],激活PPAR信号通路^[20]等途径对心血管发挥保护作用,而其可能通过促进脂肪细胞分化^[21]而对心血管产生一定的危害。这些基因功能或信号通路可作为后续进一步研究的方向,从而明确雌激素在心血管系统中的作用。

(本文图1~5见插图8-3)

参考文献

- [1] BERINSTEIN E, LEVY A. Recent developments and future directions for the use of pharmacogenomics in cardiovascular disease treatments[J]. *Expert Opin Drug Metab Toxicol*, 2017, 13(9):973-983.
- [2] MOSCA L, MOCHARI-GREENBERGER H, DOLOR RJ, et al. Twelve-year follow-up of American women's awareness of cardiovascular disease risk and barriers to heart health[J]. *Circ Cardiovasc Qual Outcomes*, 2010, 3(2):120-127.
- [3] JANUZZI JL, STERN TA, PASTERNAK RC, et al. The influence of anxiety and depression on outcomes of patients with coronary artery disease[J]. *Arch Intern Med*, 2000, 160(13):1913-1921.
- [4] ANUURAD E, ENKHAMAA B, BERGLUND L. Enigmatic role of lipoprotein(a) in cardiovascular disease[J]. *Clin Transl Sci*, 2010, 3(6):327-332.
- [5] BYINGTON RP, FURBERG CD, HERRINGTON DM, et al. Effect of estrogen plus progestin on progression of carotid atherosclerosis in postmenopausal women with heart disease: HERS B-mode sub-study[J]. *Arterioscler Thromb Vasc Biol*, 2002, 22(10):1692-1697.
- [6] MANSON JE, HSIA J, JOHNSON KC, et al. Estrogen plus progestin and the risk of coronary heart disease[J]. *N Engl J Med*, 2003, 349(6):523-534.
- [7] O' LONE R, KNORR K, JAFFE IZ, et al. Estrogen receptors alpha and beta mediate distinct pathways of vascular gene expression, including genes involved in mitochondrial electron transport and generation of reactive oxygen species[J]. *Mol Endocrinol*, 2007, 21(6):1281-1296.
- [8] BARRETT T, TROUP DB, WILHITE SE, et al. NCBI GEO: archive for high-throughput functional genomic data[J]. *Nucleic Acids Res*, 2009, 37(Database issue):D885-890. DOI: 10.1093/nar/gkn764.
- [9] ASHBURNER M, BALL CA, BLAKE JA, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium[J]. *Nat Genet*, 2000, 25(1):25-29.
- [10] ITO K, MURPHY D. Application of ggplot2 to pharmacometric graphics[J]. *CPT Pharmacometrics Syst Pharmacol*, 2013, 2:e79. DOI: 10.1038/psp.2013.56.
- [11] FRANCESCHINI A, SZKLARCZYK D, FRANKILD S, et al. STRING v9.1: protein - protein interaction networks, with increased coverage and integration[J]. *Nucleic Acids Res*, 2013, 41(Database issue):D808-815. DOI: 10.1093/nar/gks1094.
- [12] BADER GD, HOGUE CW. An automated method for finding molecular complexes in large protein interaction networks[J]. *BMC Bioinformatics*, 2003, 4:2.
- [13] 胡盛寿, 高润霖, 刘力生, 等.《中国心血管病报告2018》概要[J]. *中国循环杂志*, 2019, 34(3):209-219.
- [14] POZDNIAKOVA S, GUITART-MAMPEL M, GARRABOU G, et al. 17 β -Estradiol reduces mitochondrial cAMP content and cytochrome oxidase activity in a phosphodiesterase 2-dependent manner[J]. *Br J Pharmacol*, 2018, 175(20):3876-3890.
- [15] GAO HK, YIN Z, ZHOU N, et al. Glycogen synthase kinase 3 inhibition protects the heart from acute ischemia-reperfusion injury via inhibition of inflammation and apoptosis[J]. *J Cardiovasc Pharmacol*, 2008, 52(3):286-292.
- [16] 王宇, 杨欣, 李晓冬, 等. 雌激素抑制内质网应激引起的血管内皮细胞凋亡及机制[J]. *中国动脉硬化杂志*, 2016, 24(3):217-223.
- [17] AGAH A, KYRIAKIDES TR, LAWLER J, et al. The lack of thrombospondin-1 (TSP1) dictates the course of wound healing in double-TSP1/TSP2-null mice[J]. *Am J Pathol*, 2002, 161(3):831-839.
- [18] PÉREZ-LÓPEZ FR, CHEDRAUI P, GILBERT JJ, et al. Cardiovascular risk in menopausal women and prevalent related co-morbid conditions: facing the post-Women's Health Initiative era[J]. *Fertil Steril*, 2009, 92(4):1171-1186.
- [19] NARAYANAN RP, FU B, OLIVER RL, et al. Insulin-like growth factor-II and insulin-like growth factor binding protein-2 prospectively predict longitudinal elevation of HDL-cholesterol in type 2 diabetes[J]. *Ann Clin Biochem*, 2014, 51(Pt4):468-475.
- [20] NIKOLIC D, CASTELLINO G, BANACH M, et al. PPAR agonists, atherogenic dyslipidemia and cardiovascular risk[J]. *Curr Pharm Des*, 2017, 23(6):894-902.
- [21] 庞祥中, 李俭强, 李为民. 脂肪细胞因子 chemerin 与心血管疾病关系的研究进展[J]. *国际心血管病杂志*, 2016, 43(4):213-216.

(收稿日期:2019-06-12, 修回日期:2019-10-18)