

基于双聚类的药物重定位研究热点分析

张 庆

轩 扬

吕少妮

(济宁医学院医学信息工程学院
日照 276826)(济宁医学院管理学院
日照 276826)(济宁医学院医学信息工程学院
日照 276826)

〔摘要〕 以 PubMed 收录的药物重定位相关文献为数据来源,提取主要主题词进行频次统计,生成高频主题词-论文矩阵进行聚类分析,结果表明当前药物重定位主要研究热点为基于疾病的药物重定位与基于生物信息学的药物重定位两个方面,具体包括 5 个主题方向。

〔关键词〕 药物重定位;研究热点;聚类分析

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.08.015

Analysis on Study Hotspots of Drug Repositioning Based on Biclustering ZHANG Qing, School of Medical Information Engineering, Jining Medical University, Rizhao 276826, China; XUAN Yang, School of Management, Jining Medical University, Rizhao 276826, China; LV Shaoni, School of Medical Information Engineering, Jining Medical University, Rizhao 276826, China

〔Abstract〕 The paper uses literatures related to drug repositioning collected in PubMed as data sources, and extracts major thesauruses to perform frequency count, generates high-frequency thesauruses-literatures matrix to conduct cluster analysis. Results show that the major study hotspots of drug repositioning comprise two bases of diseases and bioinformatics with five subject directions in detail.

〔Keywords〕 drug repositioning; study hotspots; clustering analysis

1 引言

双聚类算法是在 1971 年由 Hartigan 提出的,又被称作同时聚类^[1],是对数据矩阵中的样本和变量同时进行聚类,在对象及其属性两个方向上,针对对象及其属性提取它们的联合信息,发现潜在的布局模式,在应用上比许多单向传统聚类方法更具有优势^[2]。药物重定位(Drug Repositioning)指开发药物新的适应症,即利用相关技术方法对已有药物进行重新筛选、组合或改造从而发现其新用途的过

程^[3]。制药公司试图寻求低成本的替代品以避免在新药研发中带来的风险以及昂贵的研发经费。在临床安全性研究已经完成的条件下,重新定位有助于以更低的成本及更短的时间,发现疾病新的治疗方法。因此近年来药物重定位越来越多地引起研究人员和开发商的关注。基于此,本研究对 Medline 数据库收录的药物重定位相关论文中的高频主题词进行共现聚类分析,总结药物重定位的研究热点并对其进行分析。

2 数据来源与方法

2.1 数据来源

数据来源于 Medline 数据库,由美国国家医学

〔收稿日期〕 2019-01-18

〔作者简介〕 张庆,讲师,发表论文 10 余篇。

图书馆编辑出版，是国际上生物医学领域最熟知及普遍应用的权威数据库，迄今收录文献达 2 900 万篇。所用的检索策略为"Drug Repositioning" [Ma-jr]，检索时间为 2018 年 12 月 6 日，共得到相关文献 627 篇。

2.2 方法

以 XML 格式套录检索结果，利用书目共现分析系统 BICOMB^[4]统计并抽取文献记录中的主要主题词与副主题词，按照出现频次由高到低进行排序，选取频次≥6 的 47 个主题词/副主题词作为高频词。其中出现频次最高的前 3 位主题词/副主题词分别为：药物重定位/方法（Drug Repositioning / methods）、计算生物学/方法（Computational Biology/methods）、抗肿瘤药物/药理学（Antineoplastic

Agents / pharmacology）。该 47 个高频词占有与药物重定位相关主题词的累计比例为 45.68%。根据高频词在每篇文献记录中出现次数对其进行统计，形成高频词词篇矩阵。将词篇矩阵输入 gCLUTO 软件，采用系统聚类法对所得词篇矩阵进行聚类分析，聚类分析的结果可以反映出高频词之间的亲疏关系，根据高频主题词的聚类结果以及主题词之间的语义关系，总结出药物重定位的研究热点。

3 结果

3.1 聚类结果（图 1）

其中横轴代表文献，纵轴代表聚类的主题词/副主题词。两词聚集到一起的距离越短，关系越密切。

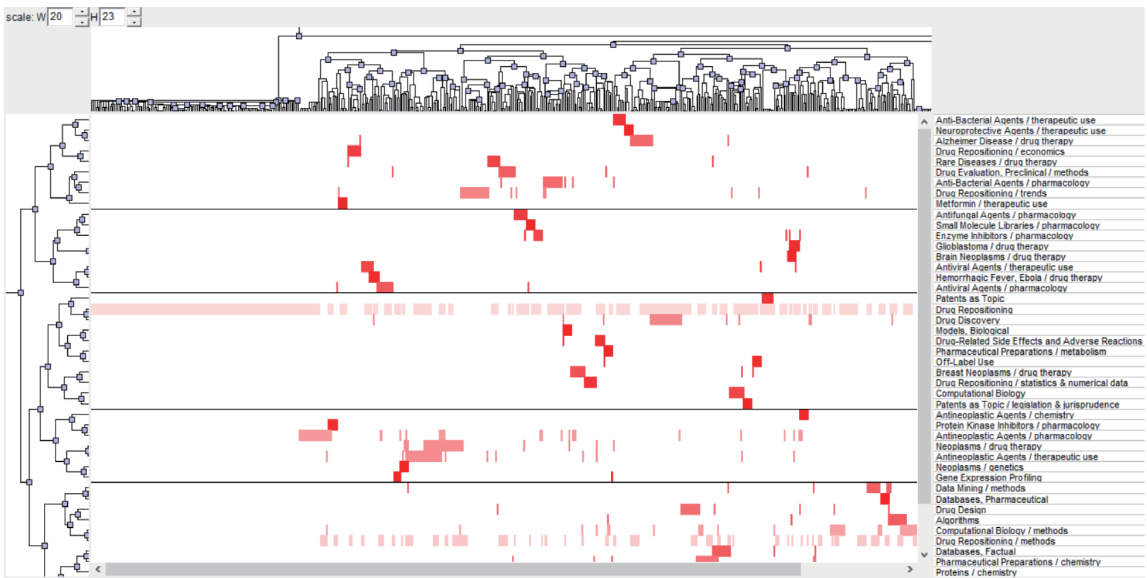


图 1 药物重新定位研究高频主题词的共现聚类

3.2 类成员对象对聚类贡献率

首先根据每一类高频词的含义以及这些主题词之间的语义关系总结出每一类主题词所代表的研究热点，即药物重定位的研究热点，如主题词神经保护药物/治疗应用（Neuroprotective Agents/therapeutic use）和阿尔茨海默病/药物治疗法（Alzheimer Disease/drug therapy）距离较近，关系密切，首先聚成一类；抗菌药/治疗应用（Anti - Bacterial Agents/

therapeutic use）再与前面两个词合成一类，依此类推。通过分析这些主题词的语义关系就能得出它们所代表的类团含义标签，综合各个类别的类标签可以得出该主题的研究热点。其次，利用 gCLUTO 软件计算各类成员对聚类贡献率的指标（描述度和区分度），选择对每一类形成贡献最大的来源文献作为表示该类内容的类标签文献^[5]。通常选取描述度分值最高者作为该类的类标签文献，见表 1。确定类标签文献后再对文献内容进一步分析，进而阐释

该类研究方向的具体内容。通过文献的高频主题词和副主题词进行共现聚类分析,可将该领域的研究热点分为两大方面,5 个主题。

表 1 类成员对聚类贡献率指标

Cluster	Size	ISim	ESim	Descriptive (%)		Descriminating (%)	
0	8	0.198	0.009	27941882	9.6	27941882	5.1
				28423558	9.6	28423558	5.1
				26374633	4.9	25588723	2.5
				27697594	4.9	24277040	2.3
1	9	0.16	0.014	26690765	6.4	26690765	3.5
				26262336	4.4	28818005	2.3
				28818005	4.1	26262336	1.8
				27579542	2.8	27579542	1.7
2	10	0.153	0.015	27716836	7.7	27716836	4.4
				24603772	7.1	24603772	4.1
				22780961	5.3	22780961	2.5
				21357612	4.3	21357612	2.4
3	10	0.153	0.016	25821813	6.6	25821813	3.2
				24460210	3.6	24460210	2.2
				26197808	3.4	28035920	2.1
				28035920	3.4	26197808	2.1
4	10	0.147	0.018	21416632	6.8	21416632	4.3
				28449705	4.2	24097821	2.4
				24097821	3.8	23227958	2.2
				23227958	3.5	22205936	2.1

4 分析

4.1 基于疾病的药物重定位

4.1.1 药物重定位在胶质母细胞瘤的应用 胶质母细胞瘤是一种 5 年生存率有限的脑肿瘤,需要开发新的治疗方案提高患者的存活率。现有用于其他症状的药物可能具有潜在的抗癌作用,为患者的生存带来显著益处。该类涉及的主题词主要包括 Glioblastoma / drug therapy、Brain Neoplasms / drug therapy、Antiviral Agents / therapeutic use、Antiviral Agents / pharmacology 等。研究内容为利用现有用于其他症状的药物为胶质母细胞瘤的治疗提供新的选

择。如利巴韦林是用于丙型肝炎治疗的抗病毒药物,也是目前唯一已知用于靶向 eIF4E 的临床批准药物,实验证明利巴韦林与替莫唑胺或放疗联合可增加神经胶质瘤细胞的细胞死亡^[6];重新利用一些穿过血脑屏障并具有潜在抗癌活性的旧药物,为胶质母细胞瘤提供新的治疗选择^[7];糖原合酶激酶(GSK) -3 β 已成为胶质母细胞瘤的治疗靶标。通过抑制 GSK -3 β 的活性研究当前批准的药物对胶质母细胞瘤的治疗效果,在实验环境和对复发性胶质母细胞瘤患者的临床研究中,将现有药物与替莫唑胺重新定位^[8]。

4.1.2 药物重定位在罕见病领域的应用 包括主题词 Rare Diseases/drug therapy、Drug Design、Pharmaceutical Preparations/metabolism 等。研究内容为通过对已经批准用于其它疾病的药物治疗罕见病。临床对大部分罕见病的致病机制尚不清楚,而罕见病的治疗药物则更少,仅有约 5% 的罕见病有治疗药物^[9]。药物重新定位是最大化药物潜力的一种方法。面对为罕见病开发新疗法的巨大困难,制药公司正在寻求“重新利用”已经批准用于其他疾病的药物治疗罕见病^[10-12]。促进罕见病治疗的一种有效策略是找到现有药物与罕见病之间的联系,如通过利用系统的药物重定位方法,发现肿瘤药物在治疗罕见病方面具有很大的再利用潜力^[12]。

4.2 基于生物信息学的药物重定位

4.2.1 基于计算生物学进行药物重定位 涉及的主题词包括 Computational Biology/methods、Drug Repositioning/methods、Drug Discovery/methods、Computer Simulation 等。研究内容为通过计算生物学的方法进行药物重新定位。如构建分类器,基于化学相互作用和化学相似性预测药物适应症^[13];提出系统的计算方法,建立药物相似性网络,通过整合化学结构相似性和基因语义相似性产生的“表达谱”来重新定位药物,以确定给定药物的潜在新型适应症^[14];在药物发现中应用共表达推断将遗传多样性细胞组中的药物反应和基因表达与靶肿瘤中的基因表达模式相关联,识别和重新利用化合物^[15];神经退行性疾病如阿尔茨海默病药物再利用的生物信息

学方法,使用基于机制的药物-靶标相互作用建模方法,重新定位候选药物^[16-17]。

4.2.2 应用数据库和专业软件发现有价值信息

该类包括主题词 Pharmaceutical Databases、Data Mining / methods、Algorithms、Computer Simulation、Software 等。研究内容为利用药物相关的大数据如药物和靶点数据库、关于信号通路的数据库等提高药物重定位的效率。如开发细胞遗传学平台,将国际人类细胞遗传学命名系统中的核型转换为可用于计算分析的机器可读模型。通过计算药物-基因相互作用数据产生药物-基因-疾病 3 联体,识别药物再利用的潜在机会^[18]。通过与所研究药物相互作用或受其影响的药物靶点分析阐明其作用机制和临床功能。基于公共数据库的药物靶标信息的途径富集分析可以提供阐明药物作用机制和重新定位的新方法,因此有利于发现新的疾病治疗方法^[19]。

4.2.3 基于数据挖掘的药物重定位 研究内容为利用药物-基因-疾病 3 者数据结合复杂的文本分析识别药物的潜在新用途。如开发并评估一种基于网络的计算方法(MD-Miner)用于预测有效药物并揭示在信号通路水平上的潜在药物作用机制^[18];通过分析“组学”数据,包括基因组学、表观基因组学、蛋白质组学和代谢组学数据,生成抗阿尔茨海默病(AD)蛋白靶标列表,开发排序算法用于筛选抗 AD 靶标的药物^[20]。

5 结语

快速发展的生物信息学向公众提供了药物作用机制关联性和临床信息,这一因素促进了药物重定位的快速发展。药物重定位在当前的药物研究中起着越来越重要的作用。通过对高频主题词进行聚类分析可以总结出药物重利用的研究热点主要集中在疾病治疗业务信息学应用等方面。一是基于疾病的药物重定位:研究内容主要为利用现有用于其他症状的药物为胶质母细胞瘤的治疗提供新的选择以及通过对已经批准用于其他疾病的药物治疗罕见病。二是基于生物信息学的药物重定位:研究内容主要为对药物和疾病相关的高通量组学(如基因组、转

录组、蛋白组和代谢组等)以及医药大数据进行挖掘和分析,进而为药物研发提供有质量的假说和线索。

参考文献

- 1 Hartigan J A. Direct Clustering of a Data Matrix [J]. Publications of the American Statistical Association, 1972, 67 (337): 123-129.
- 2 李向森. 基于双聚类和可视化的重金属污染领域研究热点探析 [J]. 公共卫生与预防医学, 2016 (6): 67-70.
- 3 崔建梅,尹大力. 药物重新定位策略在新药发现中的应用与进展 [J]. 中国药理学杂志, 2005, 39 (20): 1524-1526.
- 4 崔雷,刘伟,闫雷,等. 文献数据库中书目信息共现挖掘系统的开发 [J]. 现代图书情报技术, 2008, 24 (8): 70-75.
- 5 史航,高雯珺,崔雷,等. 生物医学文本挖掘研究热点分析 [J]. 中华医学图书情报杂志, 2016, 25 (2): 27-33.
- 6 Volpin F, Casaos J, Sesen J, et al. Use of an Anti-viral Drug, Ribavirin, as an Anti-glioblastoma Therapeutic [J]. Oncogene, 2017, 36 (21): 3037-3047.
- 7 Rundle-Thiele D, Head R, Cosgrove L, et al. Repurposing some Older Drugs that Cross the Blood-brain Barrier and Have Potential Anticancer Activity to Provide New Treatment Options for Glioblastoma [J]. British Journal of Clinical Pharmacology, 2016, 81 (2): 199-209.
- 8 Furuta T, Sabit H, Dong Y, et al. Biological Basis and Clinical Study of Glycogen Synthase Kinase-3 β -targeted Therapy by Drug Repositioning for Glioblastoma [J]. Oncotarget, 2017, 8 (14): 22811-22824.
- 9 Tao H, Zhang Y, Zeng X, et al. Niclosamide Ethanolamine-induced Mild Mitochondrial Uncoupling Improves Diabetic Symptoms in Mice [J]. Nature Medicine, 2014, 20 (11): 1263-1269.
- 10 Xu K, Coté T R. Database Identifies FDA-approved Drugs with Potential to Be Repurposed for Treatment of Orphan Diseases [J]. Briefings in Bioinformatics, 2011, 12 (4): 341-345.
- 11 Drent M, Bast A, Bootsma HP, et al. Repositioning Old Drugs to Treat Rare Diseases: Arguing from the Mechanism of Action [J]. Sarcoidosis Vasculitis & Diffuse Lung Diseases Official Journal of Wasog, 2016, 33 (2): 191-194.
- 12 Liu Z, Fang H, Slikker W, et al. Potential Reuse of Oncol-

- ogy Drugs in the Treatment of Rare Diseases [J]. Trends in Pharmacological Sciences, 2016, 37 (10): 843 – 857.
- 13 Guohua H, Yin L, Changhong L, et al. Prediction of Drug Indications Based on Chemical Interactions and Chemical Similarities [J]. BioMed Research International, 2015 (2015): 1 – 14.
- 14 Tan F, Yang R, Xu X, et al. Drug Repositioning by Applying ‘Expression Profiles’ Generated by Integrating Chemical Structure Similarity and Gene Semantic Similarity [J]. Molecular Biosystems, 2014, 10 (5): 1126 – 1138.
- 15 Gustafson DL, Fowles JS, Brown KC, et al. Drug Selection in the Genomic Age: application of the coexpression extrapolation principle for drug repositioning in cancer therapy [J]. Assay & Drug Development Technologies, 2015, 13 (10): 623 – 627.
- 16 Siavelis JC, Bourdakou MM, Athanasiadis EI, et al. Bioinformatics Methods in Drug Repurposing for Alzheimer’s Disease [J]. Briefings in Bioinformatics, 2016, 17 (2): 322 – 335.
- 17 Emon MA, Kodamullil AT, Karki R, et al. Using Drugs as Molecular Probes: computational chemical biology approach in neurodegenerative diseases [J]. Journal of Alzheimers Disease, 2016, 56 (2): 677 – 686.
- 18 Abrams ZB, Peabody AL, Heerema NA, et al. Text Mining and Data Modeling of Karyotypes to aid in Drug Repurposing Efforts [J]. Studies in Health Technology & Informatics, 2015 (216): 1037.
- 19 Pan Y, Cheng T, Wang Y, et al. Pathway Analysis for Drug Repositioning Based on Public Database Mining [J]. Journal of Chemical Information & Modeling, 2014, 54 (2): 407 – 418.
- 20 Wu H, Miller E, Wijegunawardana D, et al. MD – Miner: A Network – based Approach for Personalized Drug Repositioning [J]. BMC Systems Biology, 2017, 11 (S5): 86.

(上接第 94 页)

4 结语

移动互联网能够彻底打破时间和空间限制，催生线上与线下、系统化与碎片化结合、多方协作和自主管理的崭新学习环境，使得学习方式由数字学习（E – Learning）延伸至移动学习（M – Learning）。本研究面向移动学习的实际需求，提出中药学习平台设计方法和思路。最后，针对学生、教师等校内用户和学校附近社区的社会公众，进行平台功能测试，能够实现在任何时间、任何地点，便捷快速地获取中药学习资源，支持中医药文化传播，有助于提升学生的研究性学习兴趣，提高学习效率，验证中药学习平台的可用性和有效性。

参考文献

- 1 钟赣生. 中药学 [M]. 北京: 中国中医药出版社, 2016.
- 2 蒋若冰, 杨锴. 中药标本数据库研发管理 [J]. 医学信息学杂志, 2014, 35 (1): 33 – 36.
- 3 薛兴亚, 徐青, 章飞芳, 等. 中药信息数据系统构建 [J]. 世界科学技术 – 中医药现代化, 2006, 8 (3): 91 – 94.
- 4 李丹, 孙艳秋, 闫朝升. 中药信息数据集市的设计与应用 [J]. 中华中医药学刊, 2015, 33 (7): 1682 – 1686.
- 5 李丹, 闫朝升, 姜德友, 等. 基于多维数据模型的中药信息组织方式研究 [J]. 世界科学技术 – 中医药现代化, 2015, 17 (7): 1336 – 1340.
- 6 匡海学, 杨琳, 李永吉, 等. 从首届全国中药学类专业学生知识技能大赛看中药学类专业教育改革 [J]. 中医教育, 2016, 35 (4): 4 – 8.