

BCTD: A Drug-Target Relation Database for Drug Repositioning

Keqiang Wang¹, Hongyu Wu^{1,2}, Guodong Li², Qingshan Huang^{1,2*}

¹State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai

²Shanghai High-Tech United Bio-Technological R&D Co., Ltd., Shanghai

Email: 13210700151@fudan.edu.cn, qshuang@fudan.edu.cn

Received: Aug. 28th, 2015; accepted: Sep. 14th, 2015; published: Sep. 17th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Discovering and developing new drug is an arduous, costly and risky process. Drug repositioning, by discovering new indications of old drugs out of their original indications, is a time-saving, cost-efficient and low-risk manner. This made it more and more important in drug discovery and development. It has appeared numerous methods for drug repositioning in recent years. Combining the information of interactions between drugs and targets with omics data is a reasonable and effectively approach of drug repositioning. Dozens of databases including information of drug targets were constructed with varying scopes by different research groups in the past decade. However, it is inconvenient to researchers who need such information because of the lack of coordination in naming rules among these databases and the existed data omissions or redundancies. So we create a new database of drug-target interactions by manually inspecting and standardizing the data collected from these databases and scientific literatures. In the current version, there are 766 drug/compound entries, 746 target entries and 2862 items of drug-target interactions in the database. It will facilitate researchers to get the information of drug targets easily and quickly, providing clues for drug repositioning.

Keywords

Target, Database, Drug Repositioning

BCTD: 一个药物重定位研究用药物靶点数据库

王克强¹, 吴宏宇^{1,2}, 李国栋², 黄青山^{1,2*}

*通讯作者。

¹复旦大学生命科学学院遗传学研究所遗传工程国家重点实验室, 上海

²上海高科联合生物技术研发有限公司, 上海

Email: 13210700151@fudan.edu.cn, qshuang@fudan.edu.cn

收稿日期: 2015年8月28日; 录用日期: 2015年9月14日; 发布日期: 2015年9月17日

摘要

新药研发具有周期长、成本高、风险大等特点, 往往是漫长而艰难的。药物重定位, 即发现原有药物的新适应症, 可有效缩短周期、降低成本、规避风险, 正在成为药物研发的重要策略。近些年药物重定位发展迅速, 其中联合药物靶点相互作用信息和组学数据进行药物重定位是一种可行的方法。但是现有的众多药物靶点数据库中的部分药物靶点信息缺乏实验证明, 而且大多存在或遗漏或冗余的现象, 各库中信息亦采用不同命名规则, 难以直接用作药物重定位研究。因此我们综合整理多个数据库内容和大量文献, 排除缺乏实验证明的信息, 并标准化所得信息, 建立了一个全新的在线药物靶点数据库。该数据库现含有766个化合物、746个靶点和2862条药物/化合物和靶点相互作用的信息。该库的建立大大简化了研究人员查询与寻找药物靶点信息的流程, 能够为药物重定位提供方向和线索。

关键词

靶点, 数据库, 药物重定位

1. 引言

新药研发具有高投入、高风险、费时耗力等特点。一种药物从被发现到批准上市, 期间需要经过分子探索研究、动物实验研究、临床试验研究等多个环节, 平均需投入 10~15 年[1]和 8 亿美元[2], 且成功率极低。药物重定位(Drug Repositioning), 亦称老药新用, 是指发掘已上市药物的新适应症或新用途[3], 毋需考虑安全性和有效性, 可大大降低药物研发的风险和成本。传统的老药新用多是基于偶然发现, 如原打算用来治疗心血管疾病的西地那非在临床试验中偶然发现可用来治疗男性勃起功能障碍[4]。随着生物技术的快速发展和生物信息学的兴起, 出现了基于转录组[5]、全基因组关联分析(genome wide association study, GWAS) [6]、药物副作用[7]等相关数据的多种药物重定位方法。这些方法中, 联合运用药物靶点相互作用信息和组学数据进行药物重定位研究是一种行之有效的策略。因此, 建立一个能够用于药物重定位研究用的合适的药物靶点数据库是必须的。

近些年, 国内外研究者建立了众多包含药物靶点信息的数据库。其中, TTD (Therapeutic Target Database)属于较早建立并持续更新的药物靶点数据库。该数据库目前含有 2360 个可成为药物靶点的分子信息, 包括 388 个已有相应药物上市的、461 个处于临床试验阶段的以及其它正在研究的和已停止的药物的靶点信息[8]。DrugBank 是一个囊括物理化学性质、药理机制、临床适应症和副作用及用法用量等全面信息的综合数据库[9], 该库目前含有 1602 个美国食品药品监督管理局(FDA)通过的小分子化学药物。京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)是一个整合了化学信息、基因组信息和系统信息的综合数据库, 其中的 KEGG COMPOUND 和 KEGG DRUG 数据库中含有众多已被证实的药物靶点信息。PROMISCUOUS 则通过结合药物、靶点及药物副作用构建了一个可用于药物重定位的数据库[10]。此外, 还有 PDTD、CTD、Drug2Gene 等[11]-[13]数据库均有涉及药物靶点信息。

然而以上这些数据库涵盖面不同, 信息来源亦不同, 导致有实验支持的信息和预测的信息混杂在一起, 药物和靶点的命名也未采用统一规则, 同时存在数据遗漏或冗余现象。当研究人员想获得确切的药物靶点作用信息时, 需要在不同数据库间进行查询、比对和鉴定, 并且需要了解相关背景知识, 耗时又费力。因此, 我们对多库检索资源进行再整理, 并且查询大量文献, 对药物和靶点相关信息进行标准化, 建立了一个名为 BCTD (Bioactive Compound Target Database) 的在线药物靶点数据库, 方便研究人员进行药物重定位研究。

BCTD 数据库的建立将帮助相关研究人员快速进行药物重定位研究。通过对 BCTD 数据库的查询, 研究者能够迅速找到作用于同一靶点的不同药物信息, 结合实验验证这些不同药物是否具有相同的功能; 或者依据查询出的不同药物的靶点信息, 通过实验寻找能够对相应靶点起调节作用的分子, 从而推断出此分子可能具有对应药物的功能。该数据库目前可通过下列地址进行自由访问

(<http://biotechlab.fudan.edu.cn/database/bctd/>)。

2. 材料和方法

2.1. 数据来源

BCTD 数据库内容均是通过人工查询和整合 DrugBank、KEGG、ChemBank、CTD、TTD [8] [9] [11] [14] 等数据库和大量文献所得。本研究组选择 1000 多种 FDA 批准的具有生物活性的小分子化合物(其中大多数为药物, 剩余小部分为医药中间体或前体药物, 下文中将统称为药物), 通过 DrugBank 等数据库查询、过滤和整合相应靶点信息。现有数据库中查询不到的药物, 则通过 PubMed 和 Google Scholar 查询该药物相关文献提炼信息。最后剔除未找到靶点信息的药物和缺乏实验验证的药物靶点相互作用信息, 保留有相关文献或实验支持的数据, 构建了 BCTD 数据库。此外, 本数据库提供 DrugBank、UniProt、GenBank、GeneCard、GenAtlas、HGNC 等外部链接供用户进一步了解药物或靶点详细信息。

2.2. 数据库的构建

BCTD 数据库是在 WampServer (V 2.5) 环境下搭建, 该环境是在 Windows 平台上的 Apache (V 2.4.9)、MySQL (V 5.6.17) 和 PHP (V 5.5.12) 集成开发环境。Apache、MySQL 和 PHP 是互为独立的开源程序, 因经常一起搭配使用, 目前已形成具有良好兼容性的高性能网络开发平台, 为广大科研机构和商业公司所使用。

本数据库所有内容均存储于复旦大学信息办服务器内。

数据库的构建采用 B/S 模式, 所有数据存储于 MySQL 数据库中, 通过服务端的 PHP 代码进行数据访问与功能实现。用户使用浏览器进行数据库访问。

3. 结果

3.1. 数据库内容

本数据库现含有 766 个药物、746 个靶点和 2862 条药物/化合物和靶点相互作用的信息。本库中药物名称均采用国际非专有药名(International Nonproprietary Name, INN), 若该药物未列入 INNs 则采用英国批准的名称(British Approved Name, BAN)、美国采用的名称(the United States Adopted Name, USAN)或者默克索引(Merck Index)。本库中所涉及蛋白质和基因均采用 Uniprot 数据库中推荐名称并提供相应链接。

3.2. 数据库 WEB 接口

BCTD 是一个免费在线数据库, 目前可供用户查询和检索药物和靶点的相互作用。网站导航栏如图

1 所示。首页(Home)是对本数据库的概述, 用户可通过该页面对本数据库有初步认识。指导(Guide)页面指导用户如何正确地使用和利用本站内容。用户若在使用过程中对本站内容或网站构建有任何建议或意见, 均可通过联系我们(Contact Us)反馈给我们。浏览(Browse)页下设两个子选项, 药物浏览(Drug Browse)和靶点浏览(Target Browse), 用户可分别从两个角度了解数据库的全部内容。用户可通过位于导航栏右端的搜索框输入药物或靶点名称并选择药物(Drug)或靶点(Target)选项进行简单查询, 亦可以通过查询(Search)页面进行组合查询。

组合查询页面和查询结果如图 2 所示。用户可通过 BCTD 编号(BCTD ID)、药物名称(Drug Name)、CAS 登录号(CAS Number)、DrugBank 编号(DrugBank ID)、靶点(Target)、Uniprot 编号(Uniprot ID)、基因名称(Gene Name)、药理作用(Pharmacological Action)、物种(Species)、作用(Action)几项对数据库进行组合查询。BCTD ID 为 BCTD 数据库赋予该库中每一种药物的唯一编号。本库中药物名称均采用国际非专有药名(International Nonproprietary Name, INN), 若该药物未列入 INNs 则采用英国批准的名称(British Approved Name, BAN)、美国采用的名称(the United States Adopted Name, USAN)或者默克索引(Merck Index)。CAS 登录号是由美国化学文摘社(Cheical Abstracts Service)为每一个化学物质所提供的唯一编号。药物作用于靶点的药理作用可能尚不明确或存在争议, 用户可选择“yes”、“no”或者“unknown”。用户还可以选择药物对靶点的作用, 如抑制剂(inhibitor)、拮抗剂(antagonist)、激动剂(agonist)等。现选择在药物名称栏输入“Acetylsalicylic acid”(乙酰水杨酸), 在基因名称栏输入“P35354”, 结果显示如图 2。在图 2 所示结果中, 选择 BCTD ID 或 Drug 均可显示该药物详细信息, 如图 3 所示; 选择 Target 或 Gene 均可显示靶点详细信息, 如图 4 所示。

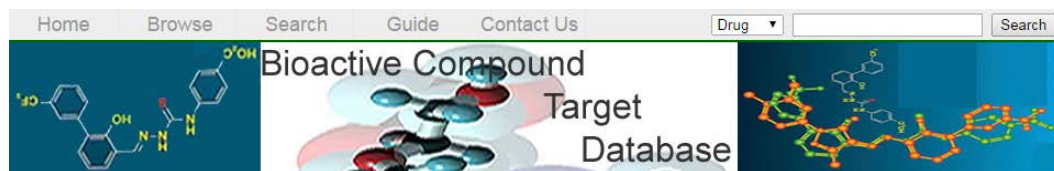


Figure 1. The navigation bar of BCTD database

图 1. BCTD 数据库导航栏

>>Home>>Search
BCTD: Bioactive Compound Target Database

BCTD ID: (BCTD ID, such as BCTD00036)
 Drug Name: (The name of drug, such as Acetylsalicylic acid)
 CAS Number: (CAS Number, such as 50-78-2)
 DrugBank ID: (DrugBank ID, such as DB00945)
 Target: (The name of target, such as Prostaglandin G/H synthase 2)
 Uniprot ID: (Uniprot ID, such as P35354)
 Gene name: (The name of gene, such as PTGS2)
 Pharmacological Action: (The pharmacological action whether is known yes, no, or unknown)
 Species: (The species of target belonging to, such as Human)
 Action: (How drug acts on target, such as Inhibitor)

>>Home>>Search
BCTD: Bioactive Compound Target Database

BCTD ID	DRUG	Target	Gene	Action	Pharmacological Action
BCTD00036	acetylsalicylic acid	Prostaglandin G/H synthase 2	PTGS2	inhibitor	yes

Figure 2. An example of the search interface and result interface of BCTD database

图 2. BCTD 数据库组合查询页面及查询结果示例

4. 讨论

目前 BCTD 数据库中含有 766 个药物和 746 个药物靶点以及 2862 条药物与靶点相互作用的信息。现在对于药物靶点的定义和到底有多少分子可成为药物靶点尚无统一看法[15],但可以确定大多数靶点均为蛋白[16]。Drews [17]于 2000 年最早总结出当时已知的所有药物的分子靶点为 483 个,其中人体内的蛋白靶点有 417 个(不包括抗菌、抗病毒和抗寄生虫等抗感染药物)。Russ [18]等的研究则表明人体内有 2000~3000 种蛋白可成为药物靶点。Peter [19]等则将已知的药物靶点缩减到仅为 218 个。此外,还有众多通过生物信息学手段预测靶点的研究,如 STITCH (Search Tool for Interactions of Chemicals) [20]即为化学物质-蛋白质相互作用在线数据库,其中包含了海量预测信息。总之,目前已知的药物靶点大约有数百个,同时还有更多潜在的药物靶点有待发现。为了提高药物重定位研究的准确性,我们排除了不准确的和预测的靶点结果。本库中现有的 746 个靶点均有相关实验或文献支持,包括作用机制已研究清楚的和正在研究中的。其中 98.5% (738 个)为蛋白,蛋白中有 84.1% (618 个)为人体蛋白,其余均为细菌、真菌、病毒、寄生虫等抗感染药物靶点。

BCTD 目前为 1.0 版本,数据量较小且信息不够完善,应进一步升级和丰富数据库内容。首先应增加药物数量和种类,建立药物分类。其次,应完善药物相关信息,如适应症(Indication)、别名(Synonyms)、详细的作用机制(Mechanism of Action, MOA)等。同时也应优化查询接口,丰富组合查询功能以方便研究者的使用。

BCTD 数据库的建立大大简化了研究人员查询与寻找药物靶点信息的流程,缩短了前期的研究时间,并且随着 BCTD 数据库的完善与丰富,这种简化效果将越发明显。BCTD 数据库的建立使研究者从繁琐的信息搜集与筛选中解脱出来,更加关注于药物重定位的策略与方法研究,提高了药物重定位研究的效率,将有助于加快药物重定位的研究。当前研究者可以通过两种策略来使用 BCTD 进行药物重定位研究:首先,通过对 BCTD 数据库的查询,研究者能够迅速找到作用于同一靶点的不同药物信息,结合生物学实验验证这些不同药物是否具有相同的功能,依此发现药物的新功能;或者依据查询出的不同药物的靶点信息,通过实验或其它方法寻找能够对相应靶点起调节作用的分子,从而推断出此分子可能具有对应

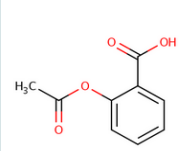
BCTD ID	Name	INN	Synonym	CAS	DrugBank ID	Structure
BCTD00036	acetylsalicylic acid		aspirin, USP	50-78-2	DB00945	
Target				Action	Pharmacological Action	
Prostaglandin G/H synthase 2				inhibitor	yes	
Prostaglandin G/H synthase 1				inhibitor	yes	
Aldo-keto reductase family 1 member C1				inhibitor	unknown	
78 kDa glucose-regulated protein				binder	unknown	
Nuclear factor NF-kappa-B p100 subunit				antagonist	unknown	
Nuclear factor NF-kappa-B p105 subunit				antagonist	unknown	
5'-AMP-activated protein kinase subunit gamma-3				activator	unknown	

Figure 3. The detailed information of acetylsalicylic acid, this figure has not displayed all targets of the drug while the whole information is on the website

图 3. 药物乙酰水杨酸详细信息, 本图未列出该药物全部靶点, 完整数据见数据库网站

Name:	Prostaglandin G/H synthase 2
Kind:	Protein
Uniprot:	P35354
Gene Name:	PTGS2
GeneBank Protein:	291988
GenBank Gene:	L15326
Uniprot Title:	PGH2_HUMAN
GeneCard ID:	PTGS2
GenAtlas ID:	PTGS2
HGNC ID:	HGNC:9605
Species:	Human

Drugs

BCTD ID	Name	Action	Pharmacological Action
BCTD00032	thalidomide	antagonist	unknown
BCTD00023	LM-1685	inhibitor	yes
BCTD00755	NS-398	inhibitor	yes
BCTD00024	SC-58125	inhibitor	yes
BCTD00367	aceclofenac	inhibitor	yes
BCTD00449	acemetacin	inhibitor	yes
BCTD00036	acetylsalicylic acid	inhibitor	yes
BCTD00249	benzylamine	inhibitor	yes
BCTD00022	celecoxib	inhibitor	yes

Figure 4. The detailed information of prostaglandin G/H synthase 2, this figure has not displayed all drugs acting on the target while the whole information is on the website

图 4. 靶点前列腺素 G/H 合酶 2 详细信息，本图未列出作用于该靶点全部药物，完整数据见数据库网站

药物的功能。本研究组已从“联系图” (the Connectivity Map, CMap) [5] 获得 BCTD 库中现有 766 种药物处理多种人类细胞系所得基因表达谱，以期得到每种药物的作用特征谱并加入到 BCTD 数据库中，联合药物靶点相互作用信息和特征谱数据为药物重定位提供线索。随着 BCTD 数据库的不断完善，相应的研究策略也会得到相应扩展，为药物重定位提供更多线索和方向。

5. 结论

BCTD 数据库是一个简洁明了的在线药物靶点数据库，目前包括 766 个药物的 2862 条靶点信息，用户可自由查询和检索。本数据库内容整合自多个公开数据库和科学文献。我们将持续对 BCTD 进行升级和完善，加入更多信息和功能，期望能建成一个基于药物靶点和基因特征谱的数据库，为药物重定位提供方向和线索，同时能为科研工作者、临床医生、学生等来自不同背景的人员所利用。

参考文献 (References)

- [1] Rask-Andersen, M., Almen, M.S. and Schioth, H.B. (2011) Trends in the exploitation of novel drug targets. *Nature Reviews Drug Discovery*, **10**, 579-590. <http://dx.doi.org/10.1038/nrd3478>
- [2] Gottlieb, S. (2012) Part II—Drug shortages: Big pharma’s new business model. *Mo Med*, **109**, 100-101.

- [3] Chong, C.R. and Sullivan Jr., D.J. (2007) New uses for old drugs. *Nature*, **448**, 645-646. <http://dx.doi.org/10.1038/448645a>
- [4] Boolell, M., Allen, M.J., Ballard, S.A., Gepi-Attee, S., Muirhead, G.J., et al. (1996) Sildenafil: An orally active type 5 cyclic GMP-specific phosphodiesterase inhibitor for the treatment of penile erectile dysfunction. *International Journal of Impotence Research*, **8**, 47-52.
- [5] Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., et al. (2006) The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929-1935. <http://dx.doi.org/10.1126/science.1132939>
- [6] Sanseau, P., Agarwal, P., Barnes, M.R., Pastinen, T., Richards, J.B., et al. (2012) Use of genome-wide association studies for drug repositioning. *Nature Biotechnology*, **30**, 317-320. <http://dx.doi.org/10.1038/nbt.2151>
- [7] Yang, L. and Agarwal, P. (2011) Systematic drug repositioning based on clinical side-effects. *PLoS One*, **6**, e28025. <http://dx.doi.org/10.1371/journal.pone.0028025>
- [8] Qin, C., Zhang, C., Zhu, F., Xu, F., Chen, S.Y., et al. (2014) Therapeutic target database update 2014: A resource for targeted therapeutics. *Nucleic Acids Research*, **42**, D1118-D1123. <http://dx.doi.org/10.1093/nar/gkt1129>
- [9] Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., et al. (2014) DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research*, **42**, D1091-D1097. <http://dx.doi.org/10.1093/nar/gkt1068>
- [10] von Eichborn, J., Murgueitio, M.S., Dunkel, M., Koerner, S., Bourne, P.E., et al. (2011) PROMISCUOUS: A database for network-based drug-repositioning. *Nucleic Acids Research*, **39**, D1060-D1066. <http://dx.doi.org/10.1093/nar/gkq1037>
- [11] Davis, A.P., Grondin, C.J., Lennon-Hopkins, K., Saraceni-Richards, C., Sciaky, D., et al. (2015) The comparative toxicogenomics database's 10th year anniversary: Update 2015. *Nucleic Acids Research*, **43**, D914-D920. <http://dx.doi.org/10.1093/nar/gku935>
- [12] Gao, Z., Li, H., Zhang, H., Liu, X., Kang, L., et al. (2008) PDTD: A web-accessible protein database for drug target identification. *BMC Bioinformatics*, **9**, 104. <http://dx.doi.org/10.1186/1471-2105-9-104>
- [13] Roider, H.G., Pavlova, N., Kirov, I., Slavov, S., Slavov, T., et al. (2014) Drug2Gene: An exhaustive resource to explore effectively the drug-target relation network. *BMC Bioinformatics*, **15**, 68. <http://dx.doi.org/10.1186/1471-2105-15-68>
- [14] Seiler, K.P., George, G.A., Happ, M.P., Bodycombe, N.E., Carrinski, H.A., et al. (2008) ChemBank: A small-molecule screening and cheminformatics resource database. *Nucleic Acids Research*, **36**, D351-D359. <http://dx.doi.org/10.1093/nar/gkm843>
- [15] Bull, S.C. and Doig, A.J. (2015) Properties of protein drug target classes. *PLoS ONE*, **10**, e0117955. <http://dx.doi.org/10.1371/journal.pone.0117955>
- [16] Xu, H., Xu, H., Lin, M., Wang, W., Li, Z., et al. (2007) Learning the drug target-likeness of a protein. *Proteomics*, **7**, 4255-4263. <http://dx.doi.org/10.1002/pmic.200700062>
- [17] Drews, J. (2000) Drug discovery: A historical perspective. *Science*, **287**, 1960-1964. <http://dx.doi.org/10.1126/science.287.5460.1960>
- [18] Russ, A.P. and Lampel, S. (2005) The druggable genome: An update. *Drug Discovery Today*, **10**, 1607-1610. [http://dx.doi.org/10.1016/S1359-6446\(05\)03666-4](http://dx.doi.org/10.1016/S1359-6446(05)03666-4)
- [19] Imming, P., Sinning, C. and Meyer, A. (2006) Drugs, their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery*, **5**, 821-834. <http://dx.doi.org/10.1038/nrd2132>
- [20] Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T.H., von Mering, C., et al. (2014) STITCH 4: Integration of protein-chemical interactions with user data. *Nucleic Acids Research*, **42**, D401-D407. <http://dx.doi.org/10.1093/nar/gkt1207>