

中国药物研发的新机遇：基于医药大数据的系统性药物重定位

王可鉴^①, 石乐明^②, 贺林^①, 张永祥^③, 杨仑^{①*}

① 上海交通大学 Bio-X 研究院, 上海 200095;

② 复旦大学药学院, 药物基因组学研究所, 上海 201203;

③ 军事医学科学院毒物药物研究所, 北京 100850

* 联系人, E-mail: ylemcreast@gmail.com

2013-06-21 收稿, 2014-03-09 接受, 2014-05-27 网络版发表

摘要 在科研资源相对有限的前提下, 如何提高药物研发的投入产出比, 最终为全社会提供安全、有效和廉价的药品, 是一个对我国生物医药基础和应用研究均具有重大意义的问题。面对国际上原创新药研发投入逐年增高且失败风险逐年增大的趋势, 挖掘已有药物新的适应症(旧药重用或药物重新定位)能够在一定程度上有效规避研发风险、降低研发成本、加快药物上市的步伐从而迅速满足临床用药需求, 因而成为众多国际制药企业重视和采用的一种战略。随着药物生物信息学的发展, 药物重定位策略正从依赖临床观察的经验性研究, 朝着基于大规模组学和医药大数据分析的系统性研究转变。本文将着重介绍国际上一系列用于挖掘和指导药物重定位的医药大数据挖掘方法, 如基于基因表达谱联系图、全基因组关联分析、化合物-蛋白互作组、药物副作用等。参考和借鉴上述各种前沿技术, 将有助于我国科研工作者开发和改进适应我国国情的药物重定位技术, 特别是中药的适应症拓展, 从而为我国带来社会和经济效益。

关键词

医药大数据
药物重定位
旧药重用
转化医学
新药研发
计算生物学
生物信息学

目前的药物研究与开发主要遵循“化合物—靶蛋白—表型变化—疾病治疗”这一中心原则, 即设计具有特定结构的化合物, 使其特异性地结合体内的某些蛋白的活性位点, 从而有效改变服药者的生理状态, 最终发挥药物对适应症的预期疗效。通常情况下, 经过长期的生物学和临床医学的研究, 相关领域的专家能够积累针对某一特定疾病的大量知识, 可基于此寻找到在病理通路中发挥核心作用的一系列靶蛋白。围绕靶蛋白可以设计和筛选相匹配的药物分子, 而后通过大规模的临床试验, 在统计学水平上确定药物在安全性和有效性方面的各项指标, 最后向药品监管部门申报, 在得到许可后上市销售^[1]。足见一个全新药物从理论研究到工业开发, 再到临床

应用, 需要走过一条异常漫长艰辛的道路, 而且其中的每个步骤都存在着前功尽弃的失败风险。(i) 新药开发所涉及的研究方法、技术条件和项目管理都十分复杂, 不仅需要庞大的研究队伍, 而且还要依赖于项目中各部分积极有效的协作, 其中任何一个环节出现失误或意外都有可能对开发项目造成重大的负面影响^[2], 这也就使新药开发具有极大的不确定性和风险性^[3]; (ii) 正因为新药开发项目往往周期长, 条件严格, 也就意味着十分高昂的成本投入^[4], 但大多数候选化合物都因为各种原因无法最终成功上市, 之前的研发投入也就无从弥补, 若将失败风险计算在内, 每个成功药物的实际平均投入将更加巨大。以上这些不利因素在近年已经成为了制约新药开发的

引用格式: 王可鉴, 石乐明, 贺林, 等. 中国药物研发的新机遇: 基于医药大数据的系统性药物重定位. 科学通报, 2014, 59: 1790–1796

Wang K J, Shi L M, He L, et al. New opportunity for Chinese pharmaceutical R & D: systematic drug repurposing based on big data (in Chinese). Chin Sci Bull (Chin Ver), 2014, 59: 1790–1796, doi: 10.1360/972013-731

瓶颈,并逐渐改变着世界各国的制药工业和公共卫生,尤其对于我国这样一个人口众多而人均资源有限的发展中国家来说,其影响显得尤为突出和深远^[5]。

药物重定位,也称为旧药新用,是指利用相关的技术方法对已有的药物进行重新筛选、组合或改造从而发现其未知新用途的过程^[6,7],长久以来都是一个不可或缺的药物开发方式。相比于从零开始的新药研发,药物重定位基于已有药物的重新开发能够节省大量前期研发投入(如药靶发现、化合物筛选、安全性测试等),因而既可能令一个失败的药物“起死回生”也可能进一步扩大一个成功药物的适应症范围和销售市场^[8]。同时,相对于传统的新药发现,药物重定位能够把药物研发的周期从 10~17 年缩短到 2~12 年,因此越来越受到政府部门、制药企业、学术机构等各方面的关注。例如,美国国立卫生研究院(National Institutes of Health, NIH)近期发起大规模的药物重定位项目(<http://www.ncats.nih.gov/research/reengineering/rescue-repurpose/rescue-repurpose.html>),鼓励各机构开放所拥有的候选化合物的部分知识产权,并加强多机构的交叉合作,发现旧化合物的潜在新疗效。包括葛兰素史克、辉瑞、阿斯利康和罗氏等在内的各大制药公司纷纷为 NIH 项目贡献化合物。据咨询公司 BioVista 统计,世界前 20 大制药企业的利润已有至少 30%来自于药物重定位。

尽管随着我国对生物医药领域的投入增多^[9],国内研发的硬件条件得到了很大的提高,但相对于新药研发的巨额投入而言,仍然不得不认真思考如何利用有限的资源实现成果产出最大化和研发风险最小化这一关键问题。药物重定位作为一种国际上被广泛采用的研发战略,具有更高的投入产出效率,值得我国产学研各界的关注和投入。本文首先对药物重定位的概念和发展历史进行了回顾,然后结合最新的研究实例总结了该领域的在世界范围内的发展现状,最后结合具体国情探讨了基于医药大数据的系统性药物重定位在我国的未来前景。

1 基于药物生物信息学的系统性药物重定位

传统的旧药新用主要通过观察来偶尔发现药物引起的未知表型变化。这种传统方法往往具有很大的非理性成分,所以相关的研究方法并不能被大范围推广并从根本上改善药物开发的效率。现代药物

研发需要系统性的和可重复的药物重定位创新方法,而药物生物信息学很可能是解决问题的钥匙之一。药物生物信息学是指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^[10],进而为药物研发提供有质量的假说和线索。在药物重定位方面,可通过组学数据挖掘或电子病历用药记录及病人表型的挖掘,提出全新的“药物-疾病”对应关系,从而为后续实验指明有希望的验证方向。本文将着重介绍一系列基于药物生物信息学和医药大数据的药物重定位新进展。

与传统的药物开发相比,基于药物生物信息学的系统性药物重定位虽然有很多方法和技术上的创新,但在研发的内在逻辑上却有很多相同之处。两者都是通过联通靶点和疾病、或表型和疾病,从而打通从药物到疾病的联系(图 1)。区别在于打通这一联系的过程中,后者可实现通过系统性和客观性的高通量筛选,并且受人为主观偏好影响相对更小,因此能提高效率并降低成本。本文将对当前国际制药工业界采用的一些重要的药物重定位创新方法论进行详细介绍。

1.1 基于基因组表达联系图的系统性药物重定位

转录组数据能够提供一系列在特定实验条件下过表达或欠表达基因,比如疾病相对于健康的对照或药物处理相对于空白的对照等。通过分析这些基因,能够评估因疾病而紊乱的通路或网络。“联系图”(connectivity map, CMap)是目前最完备的基于转录组的研发体系之一^[12],其中包含了不同药物处理人类肿瘤细胞系所得到的基因组表达变化信息。其应用方式主要是通过 gene set enrichment analysis (GSEA)算法^[13]比对不同表达谱之间的相似程度,从而产生两类假设:(i)当目标药物与特定药物具有相似的表达谱,则两个药物可能具有相似的适应症;(ii)当目标药物与特定疾病模型具有相反的表达谱,则该疾病可能是目标药物的潜在适应症。

在一系列研究中,基于 CMap 产生的假设都得到了进一步的有效验证,并为药物重定位提供了许多有价值的线索。例如,有研究分析人类受试者和小鼠(*Mus musculus*)模型的肌肉样本模拟肌肉萎缩的基因组表达模式,从而发现熊果酸具有与肌肉萎缩相反的表达谱。后续实验证明,饮食中加入熊果酸的小鼠

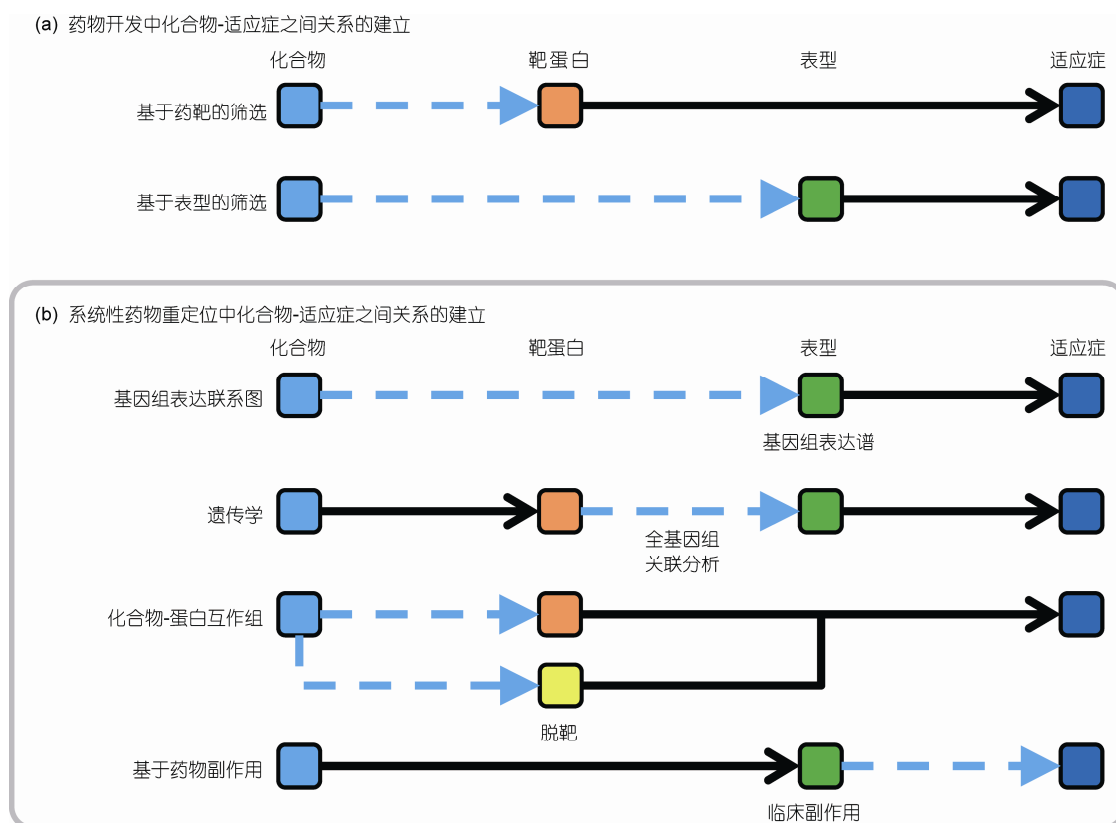


图 1 药物开发(a)与基于药物生物信息学的系统性药物重定位(b)在联通化合物和适应症之间关系的策略

两者在本质上具有很多共同的逻辑, 都是通过各种手段打通药物、药靶、表型和疾病 4 个节点之间的路径. 图中实线表明已有的关系, 而虚线代表有待建立的关系. 改自文献[11]

发生了体内的代谢变化, 并获得约 7% 的肌肉重量增长^[14]. 尤其值得一提的是, 有科研工作者基于 CMap 体系成功解释了复方中药“四物汤”的作用机理. (i) 使用四物汤处理人类肿瘤细胞系, 并获得相应的基因组表达谱; (ii) 将“四物汤”与 CMap 库中的药物表达谱进行比对, 发现了若干与四物汤相似的西药成分, 例如雌激素等. 这意味着 CMap 可以有效解释中药可能的药理作用^[15], 从而促进对中药新适应症的挖掘.

随着数据量的不断积累, CMap 第二版现在已经发展成一个收录了 1309 种药物表达谱的成熟体系. 理论上讲, 与疾病和药物相关的任何基因表达数据都可以在 CMap 数据库中进行高效率的查询比对, 从而发现与之相“联系”的药物. 而且 CMap 中使用的算法也可以被借鉴和改进, 并广泛应用于其他基因表达数据. 随着国际上针对基因表达谱微阵列质量控制的基础性研究工作的开展^[16,17], 以及我国近年大量药物和疾病相关基因组表达谱数据的积累, 基于基因表达谱的大数据挖掘将为药物重定位、特别是中

药的新适应症的挖掘提供更多的机会.

1.2 基于遗传学的药物重定位

遗传学关联分析多年来主要应用于发掘致病基因. 但从理论上讲, 如果一个药靶蛋白与某个疾病存在遗传学上的关联, 则该蛋白对应的药物很可能影响此疾病的药理学进程. 加之最新的技术进步, 如新一代测序的产生, 能显著降低基因组外显子测序的成本, 从而进一步加速了疾病相关基因作为潜在药靶的发现效率. 一个在药物重定位领域比较成功的范例是利用全基因组关联分析(genome wide association study, GWAS)的数据进行药物重定位的研究^[18]. 他们发现, 在公开的疾病相关 GWAS 基因中, 有 155 条(15.6%)的基因是在研药物或上市药物的靶点. 其中, 63 条基因所对应的 GWAS 性状与药物的适应症匹配, 表明了由 GWAS 技术所发掘的致病基因有较高的概率被作为药靶来直接开发药物. 同时, 作用于这 155 个基因所对应蛋白的药物如果尚未用于治疗

GWAS 相关疾病的话, 则有望用于治疗这些疾病. 例如, 治疗骨质疏松症的地舒单抗并未被批准用于治疗克罗恩氏病. 其靶点 TNFSF11 在 GWAS 研究中被发现与克罗恩氏病发生显著关联. 因此, 作者推测该药物可能对克罗恩氏病具有治疗作用.

这一重要研究表明, 疾病与药物之间并不一定存在明显的界限. 通过适当和有效率的数据分析, 以疾病为目标的研究也能够直接为药物研发提供重要线索^[19]. 从这个意义上讲, 目前已发表的所有 GWAS 研究, 特别是我国人群中所作的 GWAS 研究^[20-23]可以视为一个发现新药靶的宝库, 值得通过药物生物信息学方法对其进行再度挖掘.

1.3 基于化合物-蛋白互作组寻找药物新靶点

在传统的基于靶点的药物研发流程中, 第一步就是高通量的化合物筛选. 即已知某特定的蛋白可作为治疗某疾病的靶, 于是通过实验方法从海量的化合物库中发现与靶蛋白产生相互作用的先导化合物. 在药物重定位中, 化合物库可以是美国食品药品监督管理局(Food and Drug Administration, FDA)审批的药物分子, 也可以是制药公司临床研发链中的分子. 随着计算化学技术的发展和药靶蛋白知识的积累, 对化合物和蛋白结合过程的模拟计算能够在一定程度或特定问题上取代高通量筛选^[24], 并用于挖掘未知的“药物-药靶”关系.

值得注意的是, 小分子药物往往和一系列非药靶的人类蛋白(off-target, 脱靶)发生相互作用^[25]. 对于这些脱靶的挖掘不仅有助于了解不良反应的发生机理^[26], 更可以创造药物重定位的机会. 我国科研工作者为此开发出了“化合物-蛋白互作组”(chemical-protein interactome, CPI)^[27]. 不同于以往仅针对化合物与某个蛋白之间的模拟, CPI 的覆盖范围尽可能穷尽全部的上市药物和全部的人类可药性蛋白, 进而得到包含一系列靶药结合强度的矩阵. 通过对作用强度的统计分布进行矩阵变换, 可实现对特定“化合物-蛋白”配对的结合强度与该背景分布进行系统性矫正, 从而有效规避模拟中常见的偏差, 进而降低数据中的噪声并突出有价值的信号^[28,29].

该方法已经成功用于阐明导致药物副反应的脱靶效应^[27,30], 其预测的药物脱靶以及药物和蛋白的非预期性结合已经被另一个独立研究组通过 X 射线晶体衍射所证实^[31]. 而药物的非预期药理学效应(药

物重定位的潜在提示)本身也是属于药物副作用的范畴. 因此, 在 CPI 的基础上, 用于药物重定位的在线服务器 DRAR-CPI(<http://cpi.bio-x.cn/drar/>)^[32]被开发出来. 用户在线提交一个感兴趣的化合物分子, 后台程序便会自动在一个云计算平台上模拟计算其与人类蛋白组的结合强度, 从而行成一个“结合谱”. 具有显著相似结合谱的药物之间很可能具有相似的作用机理和适应症, 因此上述方法可直接用于寻找用户分子和所有库中的 FDA 所批药物间的相似性, 进而辅助药物重定位的研究.

1.4 基于药物副作用的药物重定位

临床中所观测到的药物副作用可以作为药物重定位挖掘的另一个重要来源(side effect-based repositioning). 在以往的临床研究中, 副作用信息通常被理解为药物安全的“风险”; 但随着药物生物信息学的发展, 已经能从副作用信息中发现潜在的旧药新用线索, 将风险成功转化成“机会”. 一个经典的案例是西地那非(商品名: 万艾可)原本在临床试验中用于治疗心绞痛, 然而通过仔细审视临床中的副反应发现其具有治愈勃起障碍的潜在功效^[33]. 再比如, 醋酸艾塞那肽原本用于治疗 2 型糖尿病, 在临床试验中却被发现具有减轻体重的副作用^[34].

然而, 上述发现均为“妙手偶得”, 具有很大的偶然性. 而近期, 一项系统的药物重定位研究则全面收集了与药物相关的副反应, 并评估了用这些副反应来推测药物新用途的精度^[35]. 该研究的基本假定是, 药物的治疗效果和副反应都是药理学效应, 这两者往往紧密相连, 因此副反应这种表型可作为一种“药物反应的生物标志物”来表征药物潜在的治疗用途. 例如, 治疗器官移植排异的药物在临床上往往总会导致较高比率的巨细胞病毒感染. 其背后的关联机理是排异药物往往是免疫抑制剂, 而免疫抑制的同时会增加病毒感染率. 例如, 甲氨蝶呤尚未被批准用于治疗器官移植排异, 但它会增加病人的巨细胞病毒的感染几率, 因此, 该药物有可能用于治疗器官移植排异. 事实上, 确已有研究在尝试使用该药作为移植排异的治疗, 并且疗效显著^[36]. 综上所述, 系统地研究药物上市后的不良反应作用谱可以为药物重定位找到临床层面的直接证据, 这种临床证据因远强于细胞水平和动物模型水平的药理学证据, 而被葛兰素史克采用来进行其药物重定位的开发^[35].

我国是具有全世界最庞大的医药市场之一, 国内外生产的大多数药物在我国都有规模可观的服药人群. 而且随着医疗体系的不断升级和完善, 越来越多的药物导致的副作用信息将被收集和记录在医药大数据中. 这些信息不仅有助于监控用药安全, 还将有望成为促进我国新药研发的重要资源.

2 系统性药物重定位的其他创新策略

除了上述策略之外, 生物医学研究的不同阶段还会产生其他与药物重定位相关的数据, 其中都存在着可以促进药物重定位的潜在创新机遇. 比如, 有许多研发工作旨在确定潜在的药靶, 而药靶发现会产生大量信号通路^[37]和蛋白-蛋白相互作用的数据^[38], 其中许多数据经过处理后可以整合成基因网络. 然后, 通过衡量不同疾病在基因网络上的相似度, 可以在一定程度上预测不同疾病被同一药物治愈的可能性^[39]. 再比如, 体内动物模型(包括基因敲除模型等)作为一种世界范围内被广泛使用的标准化成熟技术, 完全有机会成为潜在的药物重定位信息资源. 目前仅在小鼠表型数据库中就已经收录了至少 1400 余种疾病模型信息^[40], 而直接将基因敲除模型数据应用于药物重定位的研究仍非常少见, 其中的创新机遇不言而喻. 此外, 随着医学信息学技术的进步, 各种与疾病和药物相关的电子化数据正在飞速生成. 目前已有研究通过对电子医疗记录(如电子病历)和服药者社交媒体等医药大数据的信息挖掘, 成功地发现了药物重定位的机会^[41,42]. 总而言之, 数据和药物重定位之间存在着广泛和直接的联系, 为各种药物生物信息学和统计学方法提供了广阔的用武之地.

3 总结与展望

事实证明, 即便集中最优质的资源和大量的资本, 基于传统流程的新药开发仍是一项高投入、长周期且高风险的研发工作. 即使国际领军制药企业投

入巨资并聘请最优秀的人才, 1999~2008 年美国食品药品监督管理局也只批准了 50 个小分子一类新药, 且其中仅有 17 个是通过基于药靶的药物筛选开发的^[43]. 因此可以预见, 系统性药物重定位及其相关的创新开发方法在未来的药物研发中将不仅是一项替代策略, 而且会随着药物生物信息学和统计学的不断介入而逐渐成为一种主流战略.

这一趋势的变化, 对我国的药物研究和制药产业都将具有深远的影响. (i) 我国是一个人口众多而人均科研和医疗资源都非常有限的发展中国家, 如何提高的药物的研发效率并降低失败率显得尤为重要. 随着国内外生物技术和医药科学的长足发展, 产学研各界已经积累了并且还在产生着大量的与疾病和药物相关高通量组学数据和以电子病历数据为代表的医药大数据. 如果能加大在药物生物信息学领域的投入, 建立能够处理医药大数据的云计算平台, 并开发一系列对高通量数据进行分析 and 解读的方法, 就能够通过低成本和系统性的方式(而非基于偶然的实验或临床发现)发掘出数据的最大价值, 并最终降低药物开发的成本和风险; (ii) 众所周知, 每一种原创的新药都会享受约 10~20 年的专利保护期, 尽管在此期限内新药在未经知识产权方授权的情况下难以被再度开发. 但随着这些药物相关医药大数据的积累, 一旦超出专利保护期后, 任何机构均可以对其进行再开发. 通过系统性药物重定位和医药大数据的挖掘手段, 可以对非专利或专利即将到期的药物这一全人类共同的宝库进行深度挖掘, 从而有助于降低开发成本进而降低药价, 促进我国的医疗卫生事业的长足发展; (iii) 我国还拥有一种独特的药物研究资源, 即丰富且宝贵的中医药开发经验和数据积累. 如果能够进一步完善中药组学数据的产生和挖掘技术, 则有望利用其中蕴藏的巨大机会, 使历史悠久的中医药更加焕发生机, 带来更大的社会和经济效益.

参考文献

- 1 Gozner M. Accelerated drug approval: FDA may get tougher; companies cite hurdles. *J Natl Cancer Inst*, 2011, 103: 455–457
- 2 Hengstler J G, Bolt H M. Failure in drug development: The role of inhibition and induction of cytochrome P450 enzymes. *Arch Toxicol*, 2008, 82: 665–666
- 3 Kola I. The state of innovation in drug development. *Clin Pharmacol Ther*, 2008, 83: 227–230
- 4 Morgan S, Grootendorst P, Lexchin J, et al. The cost of drug development: A systematic review. *Health Policy*, 2011, 100: 4–17

- 5 Nyigo V A, Malebo H M. Drug discovery and developments in developing countries: Bottlenecks and way forward. *Tanzan Health Res Bull*, 2005, 7: 154–158
- 6 Chong C R, Sullivan D J Jr. New uses for old drugs. *Nature*, 2007, 448: 645–646
- 7 谢达菲, 李鹏, 李非, 等. 药物重定位的计算分析方法: 研究综述. *生物化学与生物物理进展*, 2012, 11: 1029–1036
- 8 张永祥, 程肖蕊, 周文霞. 药物重定位——网络药理学的重要应用领域. *中国药理学与毒理学杂志*, 2012, 6: 779–786
- 9 Zhang Y S. The development of biochemistry and molecular biology in China. *IUBMB Life*, 2009, 61: 549–554
- 10 Whittaker P A. What is the relevance of bioinformatics to pharmacology? *Trends Pharmacol Sci*, 2003, 24: 434–439
- 11 Hurler M R, Yang L, Xie Q, et al. Computational drug repositioning: From data to therapeutics. *Clin Pharmacol Ther*, 2013, 93: 335–341
- 12 Lamb J, Crawford E D, Peck D, et al. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 2006, 313: 1929–1935
- 13 Subramanian A, Tamayo P, Mootha V K, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, 2005, 102: 15545–15550
- 14 Kunkel S D, Suneja M, Ebert S M, et al. mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. *Cell Metab*, 2011, 13: 627–638
- 15 Wen Z, Wang Z, Wang S, et al. Discovery of molecular mechanisms of traditional Chinese medicinal formula Si-Wu-Tang using gene expression microarray and connectivity map. *PLoS One*, 2011, 6: e18278
- 16 Shi L, Reid L H, Jones W D, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*, 2006, 24: 1151–1161
- 17 Shi L, Campbell G, Jones W D, et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol*, 2010, 28: 827–838
- 18 Sanseau P, Agarwal P, Barnes M R, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol*, 2012, 30: 317–320
- 19 Wang Z Y, Zhang H Y. Rational drug repositioning by medical genetics. *Nat Biotechnol*, 2013, 31: 1080–1082
- 20 Shi Y, Hu Z, Wu C, et al. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet*, 2011, 43: 1215–1218
- 21 Lu X, Wang L, Chen S, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet*, 2012, 44: 890–894
- 22 Han J W, Zheng H F, Cui Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet*, 2009, 41: 1234–1237
- 23 Wang F, Xu C Q, He Q, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat Genet*, 2011, 43: 345–349
- 24 Kitchen D B, Decornez H, Furr J R, et al. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat Rev Drug Discov*, 2004, 3: 935–949
- 25 Li H, Gao Z, Kang L, et al. TarFisDock: A web server for identifying drug targets with docking approach. *Nucleic Acids Res*, 2006, 34 (Web Server issue): W219–W224
- 26 Lounkine E, Keiser M J, Whitebread S, et al. Large-scale prediction and testing of drug activity on side-effect targets. *Nature*, 2012, 486: 361–367
- 27 Yang L, Wang K, Chen J, et al. Exploring off-targets and off-systems for adverse drug reactions via chemical-protein interactome—clozapine-induced agranulocytosis as a case study. *PLoS Comput Biol*, 2011, 7: e1002016
- 28 Yang L, Chen J, Shi L, et al. Identifying unexpected therapeutic targets via chemical-protein interactome. *PLoS One*, 2010, 5: e9568
- 29 Yang L, Luo H, Chen J, et al. SePreSA: A server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a chemical-protein interactome. *Nucleic Acids Res*, 2009, 37 (Web Server issue): W406–W412
- 30 Yang L, Chen J, He L. Harvesting candidate genes responsible for serious adverse drug reactions from a chemical-protein interactome. *PLoS Comput Biol*, 2009, 5: e1000441
- 31 Ostrov D A, Grant B J, Pompeu Y A, et al. Drug hypersensitivity caused by alteration of the MHC-presented self-peptide repertoire. *Proc Natl Acad Sci USA*, 109: 9959–9964
- 32 Luo H, Chen J, Shi L, et al. DRAR-CPI: A server for identifying drug repositioning potential and adverse drug reactions via the chemical-protein interactome. *Nucleic Acids Res*, 2011, 39 (Web Server issue): W492–W498
- 33 Ghofrani H A, Osterloh I H, Grimminger F. Sildenafil: From angina to erectile dysfunction to pulmonary hypertension and beyond. *Nat Rev Drug Discov*, 2006, 5: 689–702
- 34 Buse J B, Henry R R, Han J, et al. Effects of exenatide (exendin-4) on glycemic control over 30 weeks in sulfonylurea-treated patients with type 2 diabetes. *Diabetes Care*, 2004, 27: 2628–2635

- 35 Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. *PLoS One*, 2011, 6: e28025
- 36 Cahill B C, O'Rourke M K, Strasburg K A, et al. Methotrexate for lung transplant recipients with steroid-resistant acute rejection. *J Heart Lung Transplant*, 1996, 15: 1130–1137
- 37 叶浩杨, 曹志伟, 唐凯临, 等. 基于 pathway 谱对药物重定位的探讨. *科学通报*, 2012, 57: 534–541
- 38 Huang J, Niu C, Green C D, et al. Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS Comput Biol*, 2013, 9: e1002998
- 39 Suthram S, Dudley J T, Chiang A P, et al. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput Biol*, 2010, 6: e1000662
- 40 Grubb S C, Maddatu T P, Bult C J, et al. Mouse phenome database. *Nucleic Acids Res*, 2009, 37 (Database issue): D720–D730
- 41 Wicks P, Vaughan T E, Massagli M P, et al. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol*, 2011, 29: 411–414
- 42 Jensen P B, Jensen L J, Brunak S. Mining electronic health records: Towards better research applications and clinical care. *Nat Rev Genet*, 2012, 13: 395–405
- 43 Swinney D C, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov*, 2011, 10: 507–519

New opportunity for Chinese pharmaceutical R & D: systematic drug repurposing based on big data

WANG KeJian¹, SHI LeMing², HE Lin¹, ZHANG YongXiang³ & YANG Lun¹

¹*Bio-X Institute, Shanghai Jiao Tong University, Shanghai 200095, China;*

²*Institute of Pharmacogenomics, College of Pharmacy, Fudan University, Shanghai 201203, China;*

³*Institute of Pharmacology and Toxicology, Academy of Military Medical Sciences, Beijing 100850, China*

Improvement in the cost-efficiency of drug development to provide people with safe, effective, and affordable medicines is of particular importance in China. Novel drug discovery is associated with high cost and risk. In contrast, identification of a novel indication for an established drug (*i.e.*, drug repositioning or repurposing) has become a widely implemented strategy in the pharmaceutical industry because of its lower risk, decreased cost, and shorter development period. Along with the development of translational bioinformatics and biomedical big data technologies, drug repurposing is evolving from empirical research that relies on clinical observation to systematic research that is based on large-scale data analysis. In this article, we introduce a series of newly established methods of big data oriented drug repurposing with a particular focus on the details of chemical-protein interactomes, side effects, gene-expression connectivity mapping, and genome-wide association studies. Taking these cutting-edge achievements as references and examples, Chinese scientists can better develop and improve drug repurposing techniques that meet the specific needs of China, thus leading to great socioeconomic benefits.

biomedical big data, drug repositioning, drug repurposing, translational medicine, drug discovery, computational biology, bioinformatics

doi: 10.1360/972013-731