

精准医学信息学的关键技术及应用发展方向^{*}

刘 爽 弓孟春 朱卫国

史文钊

(北京协和医院 北京 100730)

(神州数码医疗科技股份有限公司 北京 100080)

张抒扬

(北京协和医院 北京 100730)

〔摘要〕 介绍精准医学信息学的关键技术,包括组学数据标准化、数据采集、数据调用、自学习系统开发等方面;阐述精准医学信息学应用发展方向,包括疾病亚型分组技术、临床决策支持系统、电子病历升级换代等方面,以期推动我国精准医学信息学的发展。

〔关键词〕 精准医学信息学;表型化;机器学习;临床决策支持

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.09.002

Key Technology and Application Development Direction of Precision Medicine Informatics LIU Shuang, GONG Meng-chun, ZHU Wei-guo, Digital China Health Technologies Corporation, Beijing 100080, China; SHI Wen-zhao, Digital China Health Technologies Corporation, Beijing 100080, China; ZHANG Shu-yang, Peking Union Medical College Hospital, Beijing 100730, China

〔Abstract〕 The paper introduces the key technologies of precision medicine informatics including omics data standardization, data collection, data transfer and development of self-learning system, and states the application and development direction of the precision medicine informatics including disease subtype grouping technology, clinical decision support system and upgrade of Electronic Medical Record (EMR), in order to promote the development of precision medicine informatics in China.

〔Keywords〕 Precision medical informatics; Phenotyping; Machine learning; Clinical decision support

1 引言

精准医学是一种充分考虑患者在基因、环境、生活方式等多维度的个体差异以期达到更好的预防、诊断及治疗效果的医学模式^[1]。要实现精准医学,必须从循证医学向新的医学模式转变,将临床实践转换为证据,用临床数据来指导研究的方向。精准医学要求对个体特征进行迅速而准确的采集,

〔修回日期〕 2017-09-27

〔作者简介〕 刘爽,初级职称;通讯作者:张抒扬,副院长。

〔基金项目〕 国家重点研发计划精准医学研究重点专项“罕见病临床队列研究”(项目编号:2016YFC0901500)。

通过整合类别、格式不同的海量数据建立知识库并实现持续更新，基于不同类别的知识库建立可行的临床决策系统，这对医学信息学提出新的挑战 and 任务^[2]。精准医学信息学（Precision Medicine Informatics, PMI）是一个新型的信息学领域，主要关注数据采集、管理、分析和系统开发，以期帮助医生、患者及科研人员更好地理解健康与疾病之关键全球范围内，精准医学信息学的内涵不断丰富，正逐步形成一门独立的学科，在多个核心技术领域支撑精准医学的研究进展及临床实践^[3]。2015 年美国医学信息学协会（American Medical Informatics Association, AMIA）在年会上发布白皮书，对高级医学信息学（Advanced Health Informatics）的专业地位进行认证^[4]。此外，信息学知识被认为是临床医生和转化医学科研人员的新核心竞争能力^[5]，一系列关于医学信息学教育方法论的研究表明医学界对信息学人才培养的重视度不断提高^[6-7]。

随着生物医学数据采集和分析技术的发展，更广大的人群可以被纳入研究范围，从而实现对人体疾病状态和发展过程的进一步精细描绘和透彻理解。精准医学理念的发展为医学信息学带来机遇的同时也提出了挑战。未来精准医学信息学将发展为一个独立学科，在大数据管理、精准医疗支持、医疗可及性促进等多个维度实现长足发展。介绍支撑精准医学实施的核心技术领域，见图 1。

```
graph TD; A[数据标准化  
(医学本体系统、  
信息交互操作性、  
数据ETL标准构架)] --> B[数据采集  
(电子健康档案、  
文献文本)]; B --> C[数据调用  
(知情同意、搜索引擎、  
临床研究网络)]; C --> D[数据融合分析]; D --> E[知识库建立与维护]; E --> C; C --> F[数据安全  
(个体识别风险、  
存储安全)]; F --> E; G[自学习系统  
(机器学习、深度学习)] --> C; E --> G;
```

图 1 支撑精准医学实施的核心技术领域

2 精准医学信息学领域关键技术

2.1 组学数据标准化

2.1.1 构建医学本体系统 数据的规范采集和标准化技术可以保证数据的成功集成及交换，从而对分散分布的临床及组学信息进行提取、清洗、储存、传输等处理。数据共享可以克服单个研究中心样本量过小的问题，通过促进数据复用和荟萃分析提高数据利用效率。本体系统是一种知识组织体系，包括核心元素集、元素间交互作用、元素到规范语义间映射关系 3 个要素，通过构建规范术语、明确标准用法、构建逻辑层次、积累同义表述等实现对概念标准化、分层级、有逻辑的描述^[8]。医学本体系统是针对本专业的领域本体，通过明确医学领域内的专业术语、关系及领域公理，实现共享及一定程度的知识复用。目前国际上已经形成一批受到广泛认可、应用效果良好的医学术语集和本体资源，如 ICD-10^[9]、SNOMED-CT^[10]、HPO^[11]等，在规范描述不同类别的临床数据方面发挥重要作用，但格式、配套工具以及编程接口方面的互不兼容问题仍然在一定程度上限制了上述资源的作用。美国梅奥医学中心引领的词汇网格（LexGrid）项目借助标准工具、存储格式及访问/更新机制，为术语系统的分布式网络以及大规模术语集采纳和使用所需的基础结构提供支持^[12]。

2.1.2 实现信息交换互操作性的规范化 互操作性技术是对分布式数据进行交换、集成的前提，只有不同信息系统之间的互操作性得以保证，信息共享才具有实际意义。实现互操作性需要对数据结构、文件格式、传输标准、接口协议进行统一。HL7（Health Level 7）是目前国际上最通用的卫生信息交换标准，规定一系列临床和管理信息传输标准，包括标准信息格式、医院数据信息传输协议、应用编程界面等^[13]。其中，基因组应用编程界面通过一系列信息交换协议使不同来源的基因组数据更易与其他信息整合，为基因组数据在不同应用、平台间的安全、模式化、可操作的获取创造条件^[14]。目前已有针对肿瘤的应用程序被开发出来^[15]。

2.1.3 数据提取、转化、加载过程的标准架构

医疗大数据提取、转化、加载 (Extraction、Transform、Load, ETL) 需要一套标准架构, 即一套提取数据、清洗数据、加载数据到预设模型中的标准流程, 从而将半结构化或非结构化的海量临床数据的数据格式统一起来, 使其具备可整合分析性^[16]。同时, 质量控制、应用扩展性等也需要被充分考虑。哈佛大学引领的 I2B2 (Informatics for Integrating Biology and the Bedside) 项目可以作为这一 ETL 架构的范例。I2B2 基于医学本体系统对电子病历中的信息进行标准处理、整合、管理、交换, 同时开发出高效的数据检索工具, 为后续精准医学研究提供可靠支持^[17]。类似的数据处理架构标准是精准医学研究部署中的重要内容。

2.2 数据采集

2.2.1 基于电子病历提取表型信息 储存在电子病历中的大量表型数据具有重要的研究意义, 一方面有利于加深人们对疾病基因基础的理解, 另一方面可以验证过去发现的遗传变异在未经选择的广泛人群中是否与特定疾病具有显著相关性^[18]。由于临床数据量大、时间跨度长、成本-收益比高, 电子病历是十分理想的表型数据来源^[19]。Halpern Y 等利用电子病历中的数据构建 8 组表型数据库, 发现这些表型数据可以为急诊患者提供良好的实时临床决策支持^[20]。此外, 在大量电子病历中进行的全表型组关联分析被证实可以重复过去由全基因组关联分析论证过的变异-疾病联系, 同时揭示了一些潜在的多效性联系^[21]。近年来, 一系列基于电子病历的研究在新药研发^[22]、揭示种族特异疾病网络^[23]等方面显示出电子病历广阔的研究应用前景。

2.2.2 基于文献文本提取信息 寻找疾病风险和基因突变之间的关系是构建精准医学决策支持数据库的重要任务, 然而有些新发现通过数量庞大且非结构化的文献文本呈现, 难以及时被整合入临床决策支持系统^[24]。近年来有一些针对如何从文献文本中自动化或半自动化提取变异信息的研究, 关注特定变异或者寻找遗传变异和特定疾病之间的联系^[25]。Singhal 利用 tmVar 和 DNorm 工具对文献文

本中的变异和疾病信息进行挖掘, 以距离信息为指导开发出一系列评测指标, 得到可以自动获取变异-疾病关联的算法模型^[24]。类似的研究还有自动化识别药物的分子效应^[26]、在临床注释中自动化识别心脏病患者的受伤情况^[27]等。

2.2.3 数据长期追踪技术 精准医学知识库的持续更新有赖于各种数据的长期追踪, 因为受特定时间间隔所限的研究可能会得出不完整的结论^[28]。从数据源 (如电子健康档案) 中持续提取有效信息、定期进行重复分析并报告任何可能的显著结果十分必要。Warner 等开发的 CUSTOM-SEQ 可以从电子病历中持续获得数据, 基于基因突变追踪生存结局^[28]。开发具有连续学习能力的系统, 将多个数据源的信息持续整合, 对推动医疗模式转变起到重要支持作用。

2.3 数据调用

2.3.1 知情同意与样本管理 随着与电子病历相关联的生物样本库逐步普及, 研究者从机构内或机构间调用记录数据并选择患者构建队列成为可能, 但进一步利用数据必须事先确保患者已签署相关的知情同意^[29]。结构化、电子化的知情同意格式可以让研究的受试者查看、复核及修改其偏好, 兼顾数据、生物样本复用价值最大化和受试者隐私保护^[30]。本地标本的有效管理则涵盖标本提取、处理、质控、监管等多个方面, 而这是效力显著且可重复性高的研究的必要条件^[31]。信息学领域对上述工作的支持包括: 开发结构化知情同意书的格式、相关术语体系的部署、用户界面的设计、入组后跨机构的受试者交互所需要的基础设施等。目前数个大規模研究课题都在进行有益的技术尝试, 包括 Sage 生物网络^[32]、基因联盟及全球基因组健康联盟等。此外, OpenSpecimen^[33]、PopMedNet^[34]等课题正在持续深入研究基于角色分布式的队列和样本查询。

2.3.2 搜索引擎与跨库检索技术 数据量级的激增、数据模态的多样性、储存位点的分散为快速获取有效信息带来了巨大挑战。搜索引擎需要对多种模态的数据分别建立索引, 对返回结果进行合理排

序,开发可视化、用户友好型搜索界面。此外,随着多个区域性医疗大数据中心和多类型医学信息知识库的建立,实现跨中心、跨数据库的搜索也成为亟待解决的问题。目前已有一些课题对跨库检索医疗大数据的 IT 架构进行探索。如哈佛大学于 2009 年联合 5 家机构启动共享的健康研究信息网络 (SHRINE) 项目,实现对分布于 5 家机构中的 600 万患者超过 10 亿条临床数据的快速搜索,同时对搜索的可视化、交互性、精确程度等方面进行了优化^[35]。

2.3.3 临床研究网络构建 通过构建临床研究网络实现数据共享是目前兴起的一大策略^[36]。在临床研究网络中,从电子病历提取的数据存储在各医院并在中心的公用数据模型中进行匹配,由委员会或协作组的成员来远程执行调用^[37]。目前实施临床研究网络构建策略的研究包括 MiniSentinel (FDA 支持的药品安全检测系统)^[38]、OHDSI (观察性的医疗数据科学及信息学)^[39]、PCORNet (以美国患者为中心的临床研究网络)^[40]等。

2.3.4 通用分析工具开发 由美国 RTI 研究中心领衔建立的表型和暴露组通用度量工具箱 (Consensus Measures for Phenotypes and eXposures Toolkit, PhenX Toolkit) 是一项促进数据调用的重要工作^[41]。PhenX 工具箱提供高质量且完善的表性特征及环境暴露的标准测量方式,使科研工作者可以将多项研究的数据进行综合分析,从而提高研究的统计学效力^[42]。此类通用分析工具的开发有利于信息复用的实现,对扩大研究范围、提高数据利用效率具有重要意义^[43]。

2.4 自学习系统开发

大数据由于其体量、维度和复杂程度都远远超过人类本身的处理能力,必须需要通过一系列复杂的算法揭示藏在大数据背后的信息^[44]。传统的计算工具根据预设程序判断某一特定关系是否存在,而机器学习则是从标记或未标记过的数据中提取隐藏模型、用这一模型处理新数据。机器学习可以根据输入量是否带有标记,分为监督方法 (输入量有标记、用于训练)、无监督方法 (输入量无标记)、半

监督方法 (输入量是有或无标记的混合数据) 3 种,在疾病分型、药物复用等方面具有重要应用^[45]。深度学习是机器学习中一个模拟神经网络的研究领域,指通过组合低层特征形成更抽象的高层表示,从而呈现出学习对象的属性类别或特征^[46]。深度学习在处理高维、海量临床数据和组学数据时具有良好应用前景。Dudley 团队于 2016 年提出一种新型无监督深度特征学习方法,使用 3 层降噪自动编码方法从 70 万名患者的电子病历信息中提取出通用患者信息,发现深度学习可以更好地从电子病历中呈现患者特征,为提高临床预测水平提供优化临床决策系统的机器学习框架^[47]。

2.5 数据融合分析

各种测序技术的发展带领临床医学进入组学时代,基因组、转录组、蛋白组、代谢组、表型组和暴露组等多维度的数据大大丰富临床数据的内容,同时也增加将各种组学数据进行整合的难度^[45]。组学数据的复杂性体现在数量和异质性两个维度:对个体来说,可获得的临床数据量大大增加,同时各种数据之间由于测试技术、测试深度等差异很难具有统一格式^[45]。Kho 指出将基因组数据整合入电子健康记录的现实挑战包括基因测试结果的庞大体量和复杂性、临床和基因组数据标准不足、电子健康记录储存以及分析基因数据的限制等,提示基因组数据的成功整合可能需要对现行的电子健康记录系统进行重大修改^[48]。信息学界能否在数据整合方面提供有力支持对精准医学的发展十分关键。

2.6 知识库建立与维护

基于对数据的持续分析,可以建立一个灵活、可扩展、支持广泛应用的知识库,包含疾病亚型、疾病风险、诊断、预后等重要信息。信息学领域在构建精准医学知识库方面要提供的支持包括:(1) 开发信息技术解决方案,保证数据共享,构建对于不一致的多层的数据的临床解读的共识。(2) 提供联合查询和灵活的计算机辅助的分析能力,确保在数十亿的知识单元内进行查询、构建因果关联并提供决策支持,分别为临床医生和研究者提供应用。

(3) 确保有效地术语建模、知识管控, 构建一套新的更新及维护这一整合性的知识库的方法论, 以确保生物医学知识持续进入集成的相互连接的知识库, 基于最新的基因组学发现及临床证据辅助精准的诊断和治疗。一些研究团队已经进行积极的探索, 例如从生物医学文献中自动化识别药物的分子效应的 AIMED^[26]、持续追踪携带不同突变的肿瘤患者的生存结局的 CUSTOM-SEQ^[28] 等。

2.7 信息安全与隐私保护技术

信息安全有两个维度: 个体的重识别化风险和信息存储安全。精准医学要求的广泛数据整合涉及大量临床描述性信息及组学信息, 这使个体始终面临一定的重识别化风险^[49]。有效的去身份化处理可以起到保护隐私的作用, 是保证信息安全的重要一环^[50]。保证数据安全不意味着把数据锁死, 而是加大基因组数据加密和共享技术的开发力度, 以便在保证研究协同的情况下避免患者隐私的泄露^[51]。目前已有一些关于匿名化共享个人信息的技术手段研究^[52], 包括电子病历中的病人信息及实验室检测涉及的个人信息等。此外, 美国还出台了健康保险携带和责任法案, 规定医疗组织要在证实重识别化风险极小或列举出要变更的属性 (如年份) 才能进行数据共享, 以在个体隐私保护和数据价值最大化之间达到平衡^[53]。

3 精准医学信息学应用发展方向

3.1 疾病亚型分组技术: 改进生物标记物发现及临床转化方法

精准医学要求对患者进行更精确的分型, 以阐释疾病表现及预后中的各种变异^[45]。这要求基于目前疾病分类制定计算机可追溯的、在各个机构间具有可重复性的表型定义, 同时结合分子生物学研究厘清疾病的底层分子机制。在疾病的表型定义方面达成共识目前看来极为困难。将一个具有可干预性的生物标记物转化为临床指南的一部分并纳入临床实践前需要回答许多重要的问题, 包括基因的功能特性、生物标记物影响的通路、证据的级别、经济

承受能力的评估等^[54]。合理的框架设计可以促进针对可干预分子标记物的临床决策制定, 如对一个变异的临床意义 (是否、如何产生临床诊疗意义, 证据级别) 进行整理、将可干预的临床信息纳入变异释读报告等, 这有助于临床决策支持系统提供可理解并可用于进一步干预的患者特异性报告^[55]。针对该领域的进一步研究方向包括计算技术辅助的表型化^[56-57]、基于多个异质性数据源的数据的生物标记物发掘以及评估临床可干预性和可用性的框架体系。与此同时, 亚表型化的方法学层面有新的进展, 包括对高维度数据的分析、不同类型数据的整合以及跨越多重时间及空间的疾病发展的模拟^[58]。

3.2 临床决策支持系统: 及时、可靠的行动支持

随着数据量呈指数增长、数据更新速度不断加快, 某一特定领域内所有常见及罕见的治疗相关分子生物学类型已经超出人工掌握的范围, 精准医学知识迅速及时的传递和使用需要依赖计算机驱动的临床决策支持系统^[59]。临床决策支持是实现精准医学的核心步骤, 这一临床决策环的构建涉及多种类型的医疗数据收集 (依赖于 EHR 和数据标准管理)、基因型-表型-药物等多重数据对应关系的构建 (依赖于精准医学数据库)、开发应用于真实场景的临床决策工具等^[60]。在此基础上, 临床决策支持系统与患者个体数据相结合, 可以在诊疗过程的各个环节提供可干预的信息^[61]。一些有前景的临床决策支持工具已经被开发, 如 IMPACT^[62] (根据全外显子测序预测可干预药物), Antibiotic 临床决策支持系统^[63] 等。Halpern^[20] 等近期在《美国医学信息学会会刊》上介绍了他们利用电子病历中的数据进行实时临床决策支持的创新性工作。研究人员在锚定标记帮助下建立了一个含有 42 种表型定义的表型数据库, 利用这些信息分别对分诊时和就诊时得到的临床表型信息进行分析, 发现对包括感染、肿瘤、肺炎等 8 种病况有非常好的诊断效力。临床决策支持工具在真实场景应用中也可能会出现, 如进入重症监护室和出现肾功能损害的患者往往由于自身状况复杂而对抗生素临床决策支持系统推荐方案的依从度较低, 提示临床决策工具可能

需要持续进行反馈优化以处理更为复杂的情况^[64]。

3.3 电子病历升级换代：将精准医学融入临床实践

目前的电子病历记录结构还不够完善，不足以满足对特定病症的信息进行提取、研究的要求^[65]。电子病历需要更好的整合及显示患者的信息，以确保使用者可以看到来自不同数据源的每个患者的信息^[66]；电子病历也需要将整合的关于每位个体患者的知识以结构化的形式显示^[67]。实现这一目标需要电子病历系统提供基于标准的编程接口，保证外部数据及知识库的整合，开发支持特定工作流程、新的分析、数据可视化以及数据集成汇总的工具^[48]。开放的结构可以促进单一的电子病历厂商系统基础上模块化的基因组学临床决策支持系统的开发，促进针对每个元素的平行性的改进^[68]。如电子病历系统与精准医学知识库的连接可以提供基于基因组的风险预测、预后评估、在特定临床环节的药物剂量制定等建议^[69]。信息学界必须与电子病历系统厂商通力合作，构建用例及示范，开发接口，开创合作共赢的局面。

3.4 面向个体的信息工具及服务模式：促进精准医疗服务的可及性

精准医学所涵盖的内容远远超出医生诊室中所提供的临床诊疗服务。事实上，绝大部分人群不会经常出现在医生的诊室中。精准医学需要接受这一事实，将“用户参与”放在更高的位置，以确保消费者了解自己的健康状态，这些信息纳入到健康决策制定的过程中。这涉及在就诊间隔期间收集更多关于个人环境及生活方式的信息，包括活动量水平、营养信息、暴露信息及睡眠状态等，这些数据的有效整合可以实现靶向性的治疗及预防。制定消费者数据类型标准十分重要，开发针对信息采集、可视化及集成的用户友好界面也是实现非传统数据源可能价值的重要条件。将更广大人群的健康数据纳入知识库是精准医学的一大目标。Xu 等近期开发一个半自动化的信息提取框架，可以从政府网站中寻找临床试验并且提取相关的基因突变信息，从而在针对特定基因突变的临床试验和携带特定基

因突变的患者之间搭建桥梁^[25]。Eubank 团队也报道 DCMS (DARWIN Cohort Management System) 在自动筛选基于基因型药物试验的受试者方面的成功应用^[70]。同时，Ni 等尝试利用机器学习算法预测患者对临床试验邀请的态度，而这很有希望在未来提高临床试验招募的效率^[71]。随着医生开立医嘱的增多和直接面向消费者的遗传诊断变得更为普及且成本更低，基因检测对消费者的可及性会大大增加^[45]。然而，精准医学的重要理念并非被全部人群一致接受，如在美国不同种族对基因信息的态度及基因研究的参与度就有显著差异^[72]。如何确保消费者能够理解遗传学的基本原理、基因检测的可能后果、信息检测的准确度，如何帮助消费者结合遗传检测结果和家族史等信息选择更合理的治疗方案，将成为精准医学发展中必须解决的重要议题^[73]。此外，患者参与过程涉及的伦理、法规及社会学问题也必须得到妥善解决^[74]。

5 结语

精准医学是当今世界范围内医学研究的重点和热点，受到国家卫计委和科技部的高度重视。精准医学的逐步发展，使医学信息学的基础性支撑作用日益凸显。如何构建全方位、多层次、持续更新的知识体系并将其运用于临床情境，是医学信息学在精准医学时代要解决的核心问题。同时，信息安全与隐私保护也是精准医学信息学的重要议题。本文希望通过对中国精准医学信息学关键技术进展及发展方向的初步探讨，与医学界和信息学界人士共同把握历史机遇，推动中国精准医学信息学的快速发展。

参考文献

- 1 Francis S. Collins, Harold Varmus. A New Initiative on Precision Medicine [J]. N Engl J Med, 2015, 372 (9): 793 - 795.
- 2 Shoner A, Elemento O. A Primer on Precision Medicine Informatics [J]. Brief Bioinform, 2016, 17 (1): 145 - 153.
- 3 Frey LJ, Bernstam EV, Denny JC. Precision Medicine In-

- formatics [J]. J Am Med Inform Assoc, 2016, 23 (4): 668 – 670.
- 4 Gadd CS, Williamson JJ, Steen EB, et al. Creating Advanced Health Informatics Certification [J]. J Am Med Inform Assoc, 2016, 23 (4): 848 – 850.
- 5 Valenta AL, Meagher EA, Tachinardi U, et al. Core Informatics Competencies for Clinical and Translational Scientists: what do our customers and collaborators need to know [J]. J Am Med Inform Assoc, 2016, 23 (4): 835 – 839.
- 6 Berner ES, Dorsey AD, Garrie RL, et al. Assessment – based Health Informatics Curriculum Improvement [J]. J Am Med Inform Assoc, 2016, 23 (4): 813 – 818.
- 7 Tremblay MC, Deckard GJ, Klein R. Health Informatics and Analytics – building a Program to Integrate Business Analytics Across Clinical and Administrative Disciplines [J]. J Am Med Inform Assoc, 2016, 23 (4): 824 – 828.
- 8 Mayer G, Jones AR, Binz PA, et al. Controlled Vocabularies and Ontologies in Proteomics: overview, principles and practice [J]. Biochim Biophys Acta, 2014, (1844): 98 – 107.
- 9 Martiez I, Escayola J, Martinez – Espronceda M, et al. Seamless Integration of ISO/IEEE11073 Personal Health Devices and ISO/EN13606 Electronic Health Records into an End – to – end Interoperable Solution [J]. Telemed J E Health, 2010, 16 (10): 993 – 1004.
- 10 Millar J. The Need for a Global Language – SNOMED CT Introduction [J]. Stud Health Technol Inform, 2016, (225): 683 – 685.
- 11 Groza T, Kohler S, Moldenhauer D, et al. The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease [J]. Am J Hum Genet, 2015, 97 (1): 111 – 124.
- 12 Pathak J, Solbrig HR, Buntrock JD, et al. LexGrid: a framework for representing, storing, and querying biomedical terminologies from simple to sublime [J]. J Am Med Inform Assoc, 2009, 16 (3): 305 – 315.
- 13 Bowens FM, Frye PA, Jones WA. Health Information Technology: integration of clinical workflow into meaningful use of electronic health records [J]. Perspect Health Inf Manag, 2010, (7): 1d.
- 14 Swaminathan R, Huang Y, Moosavinasab S, et al. A Review on Genomics APIs [J]. Comput Struct Biotechnol J, 2016, (14): 8 – 15.
- 15 Warner JL, Rioth MJ, Mandl KD, et al. SMART Precision Cancer Medicine: a FHIR – based app to provide genomic information at the point of care [J]. J Am Med Inform Assoc, 2016, 23 (4): 701 – 710.
- 16 Denney MJ, Long DM, Armistead MG, et al. Validating the Extract, Transform, Load Process Used to Populate a Large Clinical Research Database [J]. Int J Med Inform, 2016, (94): 271 – 274.
- 17 Klann JG, Abend A, Raghavan VA, et al. Data Interchange Using i2b2 [J]. J Am Med Inform Assoc, 2016, 23 (5): 909 – 915.
- 18 Ackerman JP, Bartos DC, Kapplinger JD, et al. The Promise and Peril of Precision Medicine: Phenotyping Still Matters Most [J]. Mayo Clin Proc, 2016, (16): 30463.
- 19 Wei WQ, Denny JC. Extracting Research – quality Phenotypes from Electronic Health Records to Support Precision Medicine [J]. Genome Med, 2015, 7 (1): 41.
- 20 Halpern Y, Horng S, Choi Y, et al. Electronic Medical Record Phenotyping Using the Anchor and Learn Framework [J]. J Am Med Inform Assoc, 2016, 23 (4): 731 – 740.
- 21 Denny JC, Bastarache L, Ritchie MD, et al. Systematic Comparison of Phenome – wide Association Study of Electronic Medical Record Data and Genome – wide Association Study Data [J]. Nat Biotechnol, 2013, 31 (12): 1102 – 1110.
- 22 Hovlid ML, Winzeler EA. Phenotypic Screens in Antimalarial Drug Discovery [J]. Trends Parasitol, 2016, 32 (9): 697 – 707.
- 23 Glicksberg BS, Li L, Badgeley MA, et al. Comparative Analyses of Population – scale Phenomic Data in Electronic Medical Records Reveal Race – specific Disease Networks [J]. Bioinformatics, 2016, 32 (12): 1 – 10.
- 24 Singhal A, Simmons M, Lu Z. Text Mining for Precision Medicine: automating disease – mutation relationship extraction from biomedical literature [J]. J Am Med Inform Assoc, 2016, (23): 766 – 772.
- 25 Xu J, Lee HJ, Zeng J, et al. Extracting Genetic Alteration Information for Personalized Cancer Therapy from ClinicalTrials.gov [J]. J Am Med Inform Assoc, 2016, 23 (4): 750 – 757.
- 26 Fathiamini S, Johnson AM, Zeng J, et al. Automated Identification of Molecular Effects of Drugs (AIMED) [J]. J Am Med Inform Assoc, 2016, 23 (4): 758 – 765.
- 27 Topaz M, Lai K, Dowding D, et al. Automated Identifica-

- tion of Wound Information in Clinical Notes of Patients with Heart Diseases: developing and validating a natural language processing application [J]. *Int J Nurs Stud*, 2016, (64): 25 – 31.
- 28 Warner JL, Wang L, Pao W, et al. CUSTOM – SEQ: a prototype for oncology rapid learning in a comprehensive EHR environment [J]. *J Am Med Inform Assoc*, 2016, 23 (4): 692 – 700.
- 29 Kondylakis H, Koumakis L, Hanold S, et al. Donor's Support Tool: Enabling informed secondary use of patient's bio-material and personal data [J]. *Int J Med Inform*, 2017, (97): 282 – 292.
- 30 Taber C, Warren J, Day K. Improving the Quality of Informed Consent in Clinical Research with Information Technology [J]. *Stud Health Technol Inform*, 2016, (231): 135 – 142.
- 31 Smith ME, Aufox S. Biobanking: The Melding of Research with Clinical Care [J]. *Curr Genet Med Rep*, 2013, 1 (2): 122 – 128.
- 32 Kaiser J. Profile: Stephen Friend. The visionary [J]. *Science*, 2012, 335 (6069): 651 – 653.
- 33 McIntosh LD, Sharma MK, Mulvihill D, et al. CaTissue Suite to OpenSpecimen: Developing an extensible, open source, web – based biobanking management system [J]. *J Biomed Inform*, 2015, (57): 456 – 464.
- 34 Davies M, Erickson K, Wyner Z, et al. Software – Enabled Distributed Network Governance: The PopMedNet Experience [J]. *EGEMS (Wash DC)*, 2016, 4 (2): 1213.
- 35 McMurphy AJ, Murphy SN, MacFadden D, et al. SHRINE: enabling nationally scalable multi – site disease studies [J]. *PloS One*, 2013, 8 (3): e55811.
- 36 Richter TA. Clinical Research: a globalized network [J]. *PloS one*, 2014, 9 (12): e115063.
- 37 Weitzel KW, Alexander M, Bernhardt BA, et al. The IG-NITE network: a model for genomic medicine implementation and research [J]. *BMC Med Genomics*, 2016, (9): 1.
- 38 Psaty BM, Breckenridge AM. Mini – Sentinel and Regulatory Science – big data rendered fit and functional [J]. *N Engl J Med*, 2014, 370 (23): 2165 – 2167.
- 39 Hripsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers [J]. *Stud Health Technol Inform*, 2015, (216): 574 – 578.
- 40 Fleurence RL, Curtis LH, Califf RM, et al. Launching PCORnet, a National Patient – centered Clinical Research Network [J]. *J Am Med Inform Assoc*, 2014, 21 (4): 578 – 582.
- 41 Hamilton CM, Strader LC, Pratt JG, et al. The PhenX Toolkit: get the most from your measures [J]. *Am J Epidemiol*, 2011, 174 (3): 253 – 260.
- 42 Pan H, Tryka KA, Vreeman DJ, et al. Using PhenX measures to identify opportunities for cross – study analysis [J]. *Hum Mutat*, 2012, 33 (5): 849 – 857.
- 43 Hendershot T, Pan H, Haines J, et al. Using the PhenX Toolkit to Add Standard Measures to a Study [J]. *Curr Protoc Hum Genet*, 2015, (86): 1 – 17.
- 44 Belle A, Thiagarajan R, Soroushmehr SM, et al. Big Data Analytics in Healthcare [J]. *BioMed Res Int*, 2015, (2015): 370194.
- 45 Gligorijevic V, Malod – Dognin N, Przulj N. Integrative Methods for Analyzing Big Data in Precision Medicine [J]. *Proteomics*, 2016, 16 (5): 741 – 758.
- 46 Bengio Y DO. On the Expressive Power of Deep Architectures. *Proc of the 14th International Conference on Discovery Science [C]*. Berlin: Springer – Verlag, 2011: 18 – 36.
- 47 Miotto R, Li L, Kidd BA, et al. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records [J]. *Sci Rep*, 2016, (6): 26094.
- 48 Kho AN, Rasmussen LV, Connolly JJ, et al. Practical Challenges in Integrating Genomic Data into the Electronic Health Record [J]. *Genet Med*, 2013, 15 (10): 772 – 778.
- 49 Altman RB, Clayton EW, Kohane IS, et al. Data Re – identification: societal safeguards [J]. *Science*, 2013, 339 (6123): 1032 – 1033.
- 50 Li M, Carrell D, Aberdeen J, et al. De – identification of Clinical Narratives Through Writing Complexity Measures [J]. *Int J Med Inform*, 2014, 83 (10): 750 – 767.
- 51 Xia W, Heatherly R, Ding X, et al. Efficient Discovery of De – identification Policies Through a Risk – Utility Frontier [C]. *CODASPY*, 2013: 59 – 70.
- 52 El Emam K, Rodgers S, Malin B. Anonymising and Sharing Individual Patient Data [J]. *BMJ*, 2015, (350): h1139.
- 53 Xia W, Heatherly R, Ding X, et al. R – U Policy Frontiers for Health Data De – identification [J]. *J Am Med Inform*

- Assoc, 2015, 22 (5): 1029 – 1041.
- 54 Collins DC, Sundar R, Lim JS, et al. Towards Precision Medicine in the Clinic: From Biomarker Discovery to Novel Therapeutics [J]. Trends Pharmacol Sci, 2017, 38 (1): 25 – 40.
- 55 Meric – Bernstam F, Johnson A, Holla V, et al. A Decision Support Framework for Genomically Informed Investigational Cancer Therapy [J]. J Natl Cancer Inst, 2015, 107 (7), 98.
- 56 Pathak J, Bailey KR, Beebe CE, et al. Normalization and Standardization of Electronic Health Records for High – throughput Phenotyping: the SHARPN consortium [J]. J Am Med Inform Assoc, 2013, 20 (e2): e341 – 348.
- 57 Oniki TA, Zhuo N, Beebe CE, et al. Clinical Element Models in the SHARPN Consortium [J]. J Am Med Inform Assoc, 2016, 23 (2): 248 – 256.
- 58 McBride WA, Mossey PA, McIntyre GT, et al. Reliability, Completeness and Accuracy of Cleft Subphenotyping as Recorded on the CLEFTSiS (Cleft Service in Scotland) Electronic Patient Record [J]. Surgeon, 2013, 11 (6): 313 – 318.
- 59 Castaneda C, Nalley K, Mannion C, et al. Clinical Decision Support Systems for Improving Diagnostic Accuracy and Achieving Precision Medicine [J]. J Clin Bioinforma, 2015, (5): 4.
- 60 Ohno – Machado L. Data and the Clinical Decision Support Loop [J]. J Am Med Inform Assoc, 2016, 23 (e1): e1.
- 61 Bietz MJ, Bloss CS, Calvert S, et al. Opportunities and Challenges in the Use of Personal Health Data for Health Research [J]. J Am Med Inform Assoc, 2016, 23 (e1): e42 – 48.
- 62 Hintzsche J, Kim J, Yadav V, et al. IMPACT: a whole – exome sequencing analysis pipeline for integrating molecular profiles with actionable therapeutics in clinical samples [J]. J Am Med Inform Assoc, 2016, 23 (4): 721 – 730.
- 63 Zaidi ST, Marriott JL. Barriers and Facilitators to Adoption of a Web – based Antibiotic Decision Support System [J]. South Med Rev, 2012, 5 (2): 42 – 50.
- 64 Chow AL, Lye DC, Arah OA. Patient and Physician Predictors of Patient Receipt of Therapies Recommended by a Computerized Decision Support System When Initially Prescribed Broad – spectrum Antibiotics: a cohort study [J]. J Am Med Inform Assoc, 2016, 23 (e1): e58 – 70.
- 65 Moskow JM, Cook N, Champion – Lippmann C, et al. Identifying Opportunities in EHR to Improve the Quality of Antibiotic Allergy Data [J]. J Am Med Inform Assoc, 2016, 23 (e1): e108 – 112.
- 66 Kuqi K, Eveleigh T, Holzer T, et al. Design of Electronic Medical Record User Interfaces: a matrix – based method for improving usability [J]. J Healthc Eng, 2013, 4 (3): 427 – 452.
- 67 Chen TL, Chung YF, Lin FY. A Study on Agent – based Secure Scheme for Electronic Medical Record System [J]. J Med Syst, 2012, 36 (3): 1345 – 1357.
- 68 Hyponen H, Saranto K, Vuokko R, et al. Impacts of Structuring the Electronic Health Record: a systematic review protocol and results of previous reviews [J]. Int J Med Inform, 2014, 83 (3): 159 – 169.
- 69 Lang RD. In Search of the Missing Link: data access and the next generation of CDSS [J]. J Healthc Inf Manag, 2002, 16 (4): 2 – 4.
- 70 Eubank MH, Hyman DM, Kanakamedala AD, et al. Automated Eligibility Screening and Monitoring for Genotype – driven Precision Oncology Trials [J]. J Am Med Inform Assoc, 2016, 23 (4): 777 – 781.
- 71 Ni Y, Beck AF, Taylor R, et al. Will they participate? Predicting patients' response to clinical trial invitations in a pediatric emergency department [J]. J Am Med Inform Assoc, 2016, 23 (4): 671 – 680.
- 72 Dye T, Li D, Demment M, et al. Sociocultural Variation in Attitudes Roward Use of Genetic Information and Participation in Genetic Research by Race in the United States: implications for precision medicine [J]. J Am Med Inform Assoc, 2016, 23 (4): 782 – 786.
- 73 Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future [J]. Genet Med, 2013, 15 (10): 761 – 771.
- 74 Hazin R, Brothers KB, Malin BA, et al. Ethical, Legal, and Social Implications of Incorporating Genomic Information into Electronic Health Records [J]. Genet Med, 2013, 15 (10): 810 – 816.