

# 中国精准医学的发展

刘 雷

(复旦大学生物医学研究院 复旦大学基础医学院医学系统生物学系 上海 200032)

**[摘要]** 综述我国精准医学发展情况, 阐述精准医学的生物学基础、核心技术, 介绍测序技术与组学研究、大规模人群队列研究、生物大数据技术研发、个性化治疗技术开发状况, 在此基础上对未来发展进行展望。

**关键词** 精准医学; 基因测序; 组学生物学; 队列研究; 生物大数据; 个体化治疗

**[中图分类号]** R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2017.09.001

**Development of Precision Medicine in China** LIU Lei, Institutes of Biomedical Sciences, Fudan University; Department of Systems Biology for Medicine, Fudan University, Shanghai200032, China

**[Abstract]** The paper summarizes the development situation of precision medicine in China, states the biological foundation and core technology of precision medicine, introduces the situations of research on sequencing technique and omics, cohort study on large populations, R&D of biological big data technology and the development of personalized treatment technology, and looks forward to the future on this basis.

**[Keywords]** Precision medicine; Gene sequencing; Omics biology; Cohort study; Biological big data; Personalized treatment

## 1 引言

精准医学的首次提出源自 2011 年美国国家研究理事会 (National Research Council, NRC) 的一份报告, 此后该理念逐渐受到重视, 相关技术不断发展成熟。2015 年美国将精准医学计划上升为国家战略。精准医学研究集合诸多现代医学科技发展的知识与技术体系, 体现医学科学的发展趋势, 也代表临床实践的发展方向。它的实施将大大提高国民的健康水平、优化国家医疗资源配置、推动相关学科和技术的快速发展, 将带动相关产业的发展。因此, 精准医学研究已成为各国竞相布局的科技战略

制高点。面对激烈的国际竞争, 我国亟待加快启动有关精准医学研究的国家重点研发计划, 在已有深厚工作积累的基础上, 通过完善的顶层设计和跨部门、跨学科的协同创新, 力争在新一轮以实现精准医疗为目标的国际竞争中占领战略制高点, 以精准医学研究的原始创新成果引领我国生物医药与健康产业的变革和超越性发展。

随着生活方式的改变和老龄化的加剧, 我国人群的疾病谱也在发生着变化。肿瘤、代谢性疾病、神经系统疾病、心血管疾病等重大疾病成为危害人民健康的主要威胁, 呈现发病率高、治愈率低、年轻化等特点。一方面, 由于现有的诊断欠精准, 很多药物的有效率低 (如恶性肿瘤的药物整体有效率只有 23%), 常规治疗手段难以显著提高药物治疗的有效率; 另一方面, 一些重大疾病治疗药物的严重不良反应发生率显著高于西方人群, 不仅严

**[收稿日期]** 2017-09-20

**[作者简介]** 刘雷, 教授。

重威胁患者健康，而且由于医治严重不良反应伴生的医疗花费大幅增加，进一步加剧医疗资源的供需矛盾。而“精准医学”研究可以识别对药物无效或者有害的人群，为疾病的预防、诊断和治疗提供新策略和新方法。

## 2 精准医学的生物学基础

精准医学是依据患者内在的生物学与临床症状的特征数据，对患者实施量身定制的临床医疗策略。恶性肿瘤、心脑血管疾病、呼吸系统疾病、代谢性疾病、自身免疫性疾病等的发生发展是外界环境、个体生活习惯和个体基因组等多因素相互影响的结果。生物技术尤其是基因测序技术的飞速发

展，为精准医学的发展创造了条件。根据基因测序结果进行的个性化诊断，把DNA序列与疾病或者体征联系起来。它一方面要求DNA测序更快更准、成本足够低；一方面要求找到生物信息和疾病的联系。目前最先进的测序仪已实现高通量测序，测序成本大幅下降，使该技术大范围临床应用成为可能，许多基因突变与疾病的相关性研究已取得突破。基于大规模多组学生物学数据的技术（如蛋白组学、代谢组学、基因组学、转录组学及表观基因组学等），及相应计算分析大数据工具快速发展，为精准医学提供了强有力的技术基础。临床信息学技术的进步，如电子医疗病例等，也为获得详细临床数据，对接组学生物学大数据提供了可能。我国人群疾病谱和医疗支出比例，见图1。

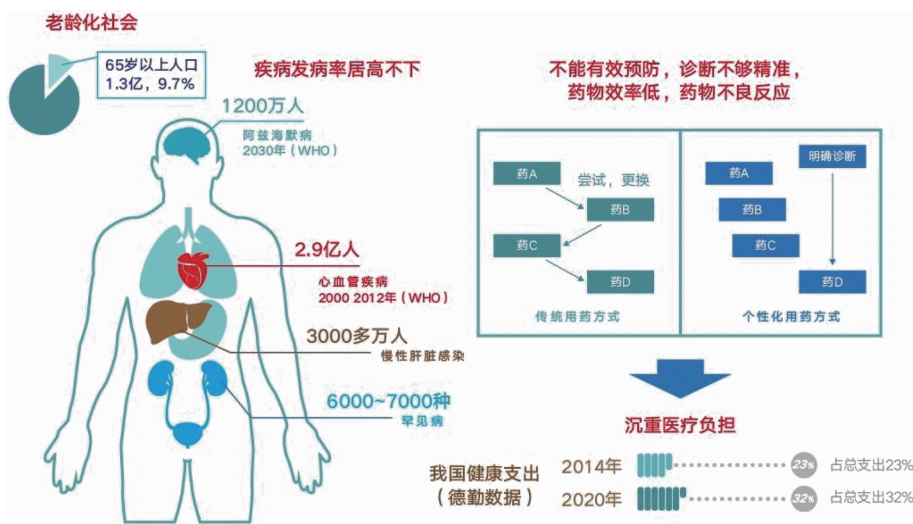


图1 我国人群疾病谱和医疗支出比例

例如作为精准医学的重要组成部分，精准肿瘤学的研究已从癌症基因组的系统研究中获益，这些研究可以揭示癌症基因影响细胞信号、染色体、表观调节以及代谢和谱系成熟等特征。临床表现相同的患者，他们的基因可能由于在分子水平上存在巨大差异，而导致每个人对临床治疗反应不尽相同。肿瘤是一种病因复杂多样的疾病，在分子遗传机理上有很大的异质性，即使相同病理类型的癌症患者，他们对抗癌药物的反应也可能迥异，所以肿瘤学科成为精准医疗最重要的领域之一。目前，精准医疗更多地集中在对恶性肿瘤的早期诊断和治疗

上，基于个体基因检测的肿瘤个体差异化治疗成为重要趋势，肿瘤研究已从细胞生物学水平逐渐转变到分子生物学水平，结合患者自身癌细胞独特的基因特征进行确诊并制定诊疗方案。此外，靶向特异性药物选择性地作用于患者的特异性遗传学靶点的，可提供更安全有效的治疗。治疗靶点一般是信号通路上信号转导或转录活化的关键分子。药物毒性和治疗的抵抗性是靶向药物治疗的巨大挑战。新开发的分子靶向药物毒性可影响心、肺、皮肤、内分泌和胃肠道等器官。因此，计算机数据挖掘是优化疾病候选基因的重要工具。

更重要的是生命的复杂性远远不是简单地测定基因组核酸序列就能阐明的。生命复杂性涉及表观遗传现象，以及代谢小分子和糖脂的参与；生命复杂性还涉及细胞、组织和器官等不同层次。如果把一类生物分子或一种表型视为一个变量，相同变量的数据形成一个信息层，那么生物（例如患者）个体就是由很多变量组成的多层级结构，每一层包含一个与疾病相关的变量信息。这样的生物医学知识反映出来的正是其核心特征——多变量的整合。科学家首先要把生物系统内不同种类的分子组成成分整合在一起进行研究；其次，对于多细胞生物而言，还要实现从基因到细胞、到组织、到个体的各层次的整合。精准医学需要构造的生物医学知识是建立在生物系统各层次基础上的。转化医疗研究是实现精准医学强有力的手段。转化医疗在基础研究和临床医疗之间建立了更直接的联系，打破了基础医学与药物研发、临床及公共卫生之间的固有屏障，将它们直接关联起来。转化医疗应用生物学和临床医学大数据，将基础医学研究成果向最有效和

最合适的疾病预防诊断、治疗和预防模式的转化，推动精准医学的发展。

### 3 精准医学的核心技术

精准医学研究是指在大样本研究中获得疾病分子机制的知识体系基础上，以生物医学数据，特别是组学数据为依据，根据“患者个体”在基因型、表型、环境和生活方式等各方面的特异性，应用现代遗传学、分子影像学、生物信息学和临床医学等手段，制定个性化的精准预防、精准诊断和精准治疗方案。我国已初步建立重大疾病临床标本库及临床试验网络，获得了一批针对重大疾病的易感人群筛选、预警、早期诊断、预后判断和指导个体化治疗的相关分子标志物/谱；初步建立了具有各种疾病特色的临床研究队列和生物样本库等，部分领域已处于与国际水平并列的地位，这些工作为我国精准医学研究重点专项的实施奠定了坚实的基础。精准医学体系，见图 2。

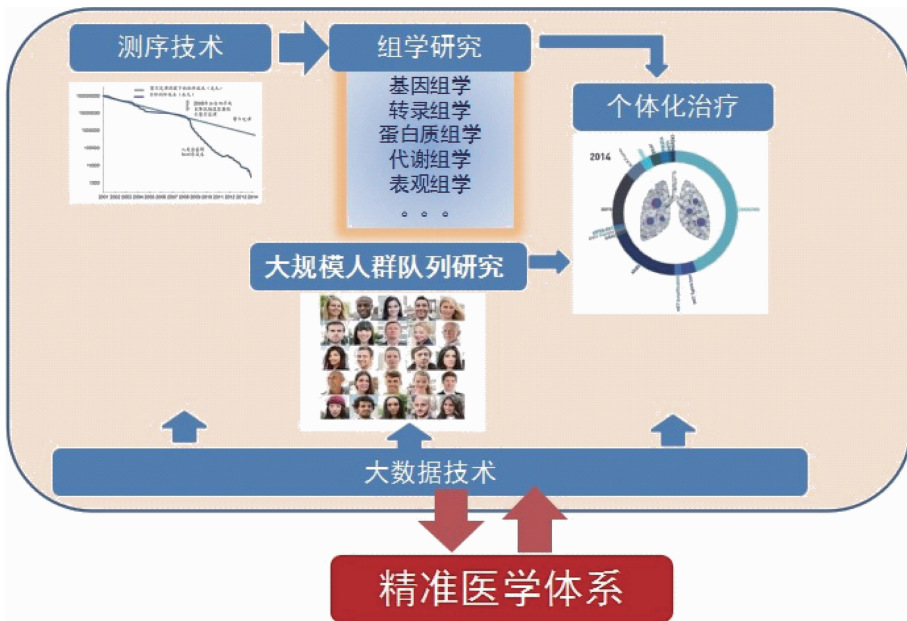


图 2 精准医学体系

### 4 测序技术与组学研究

测序技术的进步和组学研究的快速发展为精准

医学提供了关键技术和科学基础。第 2 代测序技术 (Next-generation Sequencing, NGS) 是指一个能生成高达 500 千兆碱基数据的大规模平行测序技术。该技术是肿瘤基因组计划 (The Cancer Genome

Atlas, TCGA) 和国际癌症基因组联盟 (International Cancer Genome Consortium, ICGC) 绘制完整人类癌症基因图谱的主要工具, 可检测单核苷酸变异、插入或缺失、拷贝数异常、结构变异、基因融合、甲基化及表达。其中 DNA 测序技术正在向高效率、低成本、高通量和高精确度方向发展。目前, 已可以针对单细胞进行基因组、转录组等的测序分析。宏基因组技术的发展使研究疾病与人体微生物组的关系成为可能, 为疾病防治提供新的视角和策略, 目前已开展了肠道、皮肤、口腔等菌群的宏基因组测序分析。定向蛋白质组学技术以其高灵敏度、重复性好、定量准确及可操作性高的特点, 将蛋白质组学研究的思维方式从发现式带入定向式, 大大提高了蛋白质分析的精确性。超灵敏高覆盖代谢组原位测量分析技术体系为突破临床代谢组学的瓶颈提供了可能, 实现了系统分析代谢物组成及其网络调节变化规律。测序技术的快速发展使得组学研究得以快速发展。继人类基因组计划完成后, 国际上先后开展了从基因组、转录组、蛋白质组, 到代谢组、表观基因组、表型组等一系列组学计划。

我国参与了国际人类基因组计划, 这大大提高了我国组学的研究水平, 也使我国在基因组学、蛋白质组学、代谢组学、表型组学研究中处于国际领先行列, 在基因组测序、疾病发病机制与药物设计靶点等方面也取得了重大进展, 形成有国际竞争力的研究团队。在完成人类基因组计划任务后, 我国先后参与了国际人类基因组单体型图计划、DNA 元件百科全书计划、千人基因组计划等多项国际重大基因组研究项目。2004 年启动了人类第 1 个组织器官 (肝脏) 的蛋白质组计划, 这也是中国第 1 次领导大型国际合作计划。

## 5 大规模人群队列研究

大规模人群队列研究和海量临床样本为建立精准医疗体系奠定了基础。大型队列是一种系统的病因学研究方法, 主要针对数十万人群, 在数十年内持续对人群健康状况和疾病进行追踪、随访调查和

相关研究, 以了解人群健康状况和疾病发生情况随社会经济改变而发生的变化和相关影响因素。其中, 电子健康档案是包含有关社会人口统计学、医疗状况、遗传学和治疗等信息的大数据来源。临床医生需要利用计算机模型来帮助组织数据、识别模式、解释结果和设置行动阈值, 通过大数据分析, 产生新知识、改善临床护理、改进公共卫生监控。由于生物组学技术、转化医疗、流行病学、影像技术、信息科技的快速发展, 人们对健康与环境、经济、社会、文化的复杂关系的认识日益深刻, 综合性、前瞻性的大型健康队列建设意义进一步凸显, 队列研究的维度和深度大大增加。

2006 年 5 月的香山科学会议将大型队列建设上升到国家战略高度。“十二五”科技支撑计划“区域人口健康大型队列关键技术研究”等项目顺利实施为建设国家大型健康队列奠定了初步基础。2014 年 4 月, 以“国家大型健康队列建设与应用”为主题的香山科学会议上, 与会专家呼吁应尽快建立国家大型健康队列研究, 以期明确我国社会快速转型期产生的致病和危险因素, 探索发病机制, 开发、应用和推广防控策略和工具, 提高防控效果, 降低和减轻慢性疾病导致的沉重经济和社会负担。与此同时, 中国科学院等机构建立了我国首个包括近两万人份的 2 型糖尿病全基因组关联分析研究数据库。大规模人群队列研究在国内已有了显著的发展。

## 6 生物大数据技术研发

生物信息技术方面突飞猛进的发展, 特别是在各种组学方面提供了前所未有的机会去进行大数据处理, 从而认识肿瘤的遗传特征和发病分子机理, 对癌症的个体性预防或治疗带来准确有力的指导。组学研究产生的海量数据引领生物医学研究进入大数据时代。现代生物医学的发展模式已转向以数据为驱动的数据密集型科学发现模式。生物数据正以惊人的速度增长, 麦肯锡公司的报告指出, 全球每年新增与医疗相关的生物数据达到数十艾字节 (1

EB =  $10^{18}$  B) 量级, 在 2020 年左右达到 35 泽字节 ( $1 \text{ ZB} = 10^{21} \text{ B}$ ) 的规模。早在 20 世纪 80 ~ 90 年代, 美国、欧洲和日本即已分别建立世界 3 大生物数据中心: 美国国家生物技术信息中心、欧洲生物信息研究所和日本 DNA 数据库。这 3 大生物数据中心掌握并管理着全世界的生物数据和知识资源, 处于垄断地位。近年来的大型转化医疗项目还包括肿瘤基因图谱计划, DNA 元件百科全书 (Encyclopedia of DNA Elements, ENCODE) 和人类蛋白质组计划 (Human Proteome Project, HPP) 等。

我国自人类基因组计划以来, 对国际生物数据资源的贡献日益增大, 各类测序形成了海量的组学数据; 健康档案、电子病历医学影像等医疗数据等呈指数级的增长; 基于各类疾病患者组织样本的组学研究数据也在大量产生, 我国生物医学数据的产出量已稳居世界第 2 位, 成为名副其实的数据生产大国。在精准医学快速发展的背景下, 面向精准医疗需求, 需要建立以生命组学数据、临床信息和健康数据为核心的多层次数据整合的大数据中心; 研发多维动态大数据标准、管理和共享系统; 研发系列大数据技术、工具和平台; 研发海量个人基因组信息管理、注释、可视化与应用系统; 建立精准医疗大数据的存储、加密、分析、解读和应用的全产业链技术体系, 以获得大数据整合分析能力和提取核心信息的能力; 整合已有临床大样本资源, 采集基因组、蛋白质组、代谢组、表型组等组学信息, 建立可共享的组学和临床信息库等等。

## 7 个体化治疗技术开发

个体化治疗技术的开发与临床方案制定是精准医学理念的具体体现。随着人类基因组计划的完成, 现代医疗模式发生了重大变革, 疾病的预测和治疗开始步入基因水平, 在此背景下产生的个体化治疗, 就是以个人基因组信息为基础, 结合蛋白质组、药物代谢组等相关内环境信息, 为患者量身设计最佳治疗方案, 以期达到治疗效果最大化、不良反应最小化的定制医疗模式。基因组学和基因组测

序技术的发展推动了个体化治疗的发展, 药物基因组学、蛋白质组学、代谢组学以及生物信息学、生物芯片技术、纳米生物技术等组成个体化治疗的基础; 分子水平生物标志物使疾病分类从宏观形态学转向以分子特征为基础的分类体系, 为疾病的分子分型和预后发展提供了重要参考; 具有统一标准的数据和高品质生物样本的获取和存储是个体化治疗实现的必要条件。从广义的角度看, 个体化医疗包含六维概念, 即疾病、环境、基因、药物治疗、医疗保健和信息, 它们之间存在着相互联系相互影响的关系。疾病发展、环境 (微生物) 和基因表达整合成一个概念性个体化空间, 其中癌症、糖尿病、心血管疾病等多基因疾病的个性化治疗成为研究焦点。新疗法聚焦于改变患者异常的基因融合和突变、甲基化、乙酰化作用、畸变或者蛋白过度表达。基于人类遗传学和基因组学、下一代基因测序技术、信号通路、基因交互及其网络、分子的监管和控制等研究成果, 精准医学可用于创建更有效更安全的疗法。精准医学的新意体现在生物医学新技术对许多疾病有更深入的理解。

我国在“十一五”期间通过部署“常见重大疾病的全基因组关联分析和药物基因组学研究”重大专项, 建立了常见重大疾病药物基因组关键技术体系, 获得了一批与药物有效性安全性相关分子标志物, 包括 30 种阿尔茨海默病和帕金森病致病基因及单核苷酸多态性 (SNP) 位点, 并应用于帕金森病和阿尔茨海默病的基因诊断及风险预测的临床试验研究; 与食管癌分子分型和个体化诊疗相关的 2 种食管癌易感基因和 18 个 SNP 位点; 与铂类药物治疗相关的肺癌 20 个 SNP 位点; 与胃癌病变密切相关的 3 个基因多态性。精准医疗计划有望列入国家“十三五”科技发展规划。

## 8 展望

精准医疗的产业前景巨大。一方面, 大数据技术与传统诊疗体系的结合与升级改造, 产生新的健康服务产品; 另一方面, 建立在基因测序基础上的



精准诊疗方式和个性化药物研发, 将带来整个健康医疗领域产业链的全面创新, 带动疾病诊疗产品与服务的全面升级, 形成新的健康经济增长点。同时, 针对未知病因和治疗方案的罕见疾病, 全基因组测序也展示了明确病因寻找治疗方案的成功案例。因此, 精准医学的开展将推动预防为主的健康医学发展。与传统医学不同, 精准医学的发展可以从根本上精准地优化诊疗效果, 避免医疗资源浪费, 减少无效、有害和过度医疗, 降低医疗成本, 优化医疗资源配置。精准医学研究的实施也是新的疾病分类体系和诊疗标准话语权的再分配过程。

精准医学的发展将大大提升相关学科研究和技术水平。它是集合了诸多现代医学科技发展的知识与技术体系, 体现了医学科学发展趋势, 也代表了临床实践发展的方向。精准医学研究的开展需要大规模人群和特定疾病队列研究、各类组学及其测序技术、生物大数据分析及其整合技术、分子影像等

相关技术研发, 在应用方面需要临床医学研究、检测与诊断研发、个体化治疗技术开发等。精准医学概念提出之前, 呈现“碎片化”, 精准医学研究的实施无疑将高效整合这些学科和技术, 并促进其快速发展, 形成整体性解决方案, 最终提高疾病的预防和诊治效率。精准医学研究已成为新一轮国家科技竞争和引领国际发展潮流的战略制高点。

综上所述, 在基因组测序、定量蛋白质组、高灵敏代谢组等新一代生命组学技术进入大规模应用的窗口期, 发挥我国临床资源优势, 在精准医学领域快速布局, 对于提高疾病的防诊治效率, 优化医疗资源配置, 提高国民健康水平, 抢占未来医学及相关产业发展的主导权具有至关重要的意义, 将为我国科技创新和健康产业带来新的驱动力。

(本文转载自《科学》2016 年 11 月 (68 卷 6 期), 略有改动。)

## 2018 年《医学信息学杂志》征订启事

《医学信息学杂志》是国内医学信息领域创刊最早的医学信息学方面的国家级期刊。主管: 国家卫生和计划生育委员会; 主办: 中国医学科学院; 承办: 中国医学科学院医学信息研究所。中国科技核心期刊 (中国科技论文统计源期刊), RCCSE 中国核心学术期刊 (武汉大学中国科学评价研究中心, Research Center for Chinese Science Evaluation), 美国《化学文摘》、《乌利希期刊指南》及 WHO 西太区医学索引 (WPRIM) 收录, 并收录于国内 3 大数据库。主要栏目: 专论, 医学信息技术, 医学信息研究, 医学信息组织与利用, 医学信息教育, 动态等。读者对象: 医学信息领域专家学者、管理者、实践者, 高等院校相关专业的师生及广大医教研人员。

2018 年《医学信息学杂志》国内外公开发行, 每册定价: 15 元 (月刊), 全年 180 元。邮发代号: 2-664, 全国各地邮局均可订阅。也可到编辑部订购: 北京市朝阳区雅宝路 3 号 (100020) 医科院信息所《医学信息学杂志》编辑部; 电话: 010-52328673, 52328674, 52328671。

《医学信息学杂志》编辑部