DOI: 10.13376/j.cbls/2015074
文章编号：1004-0374(2015)05-0558-06

· 特约综述 ·

建立在系统生物学基础上的精准医学

吴家睿[1,2,3]
(1 中国科学院上海生命科学研究院生物化学与细胞生物学研究所，上海 200031；2 中国科学院上海高等研究院，上海 201210；3 上海科技大学生命科学与技术学院，上海 201210)

摘　要：面对生命复杂性的巨大挑战，研究者提出了整合基因组、蛋白质组和代谢组等多组学数据，以及整合从分子到生理病理表型数据的系统生物学研究策略，利用该策略建立以个体为中心的多层级人类疾病知识整合数据库，并在此基础上形成可用于疾病精确分类的生物医学知识网络，进而发展出未来能够为每个个体提供最好医疗护理的精准医学。

关键词：系统生物学；精准医学；数据库；知识网络
中图分类号：Q-0；R45　　**文献标志码：**A

Precision medicine based on systems biology

WU Jia-Rui[1,2,3]
(1 Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; 2 Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China; 3 School of Life Sciences and Technology, ShanghaiTech University, Shanghai 201210, China)

Abstract: Facing the grand challenges of biological complexity, researchers have developed the systems biology approaches for integrating data derived from genomics, proteomics and metabolomics and other omics, integrating data from molecular level to physiological and pathological levels. Based on the strategy and tools of systems biology, it is possible to build individual-centric biomedical database that consists of multilayered and highly inter-connected biological parameters, and then construct a knowledge network of biomedical research, which could be used for new taxonomic classification of diseases. The biomedical knowledge-network and new taxonomy of diseases should support a so-called precision medicine that provides the best accessible medical-care for each individual.

Key words: systems biology; precision medicine; database; knowledge network

2015 年 1 月 20 日，美国总统奥巴马在美国国会做国情咨文报告时发表了一段激动人心的讲话：“我希望这个消灭了天花、绘制出人类基因组图谱的国家可以引领一个新时代 —— 一个在恰当时机提出正确治疗方法的时代。今晚，我将发起一项‘精准医学’倡议，让我们向治愈癌症与糖尿病等疾病更靠近一步，使每个人都可以得到让我们与家人保持健康所需要的个性化信息。”这段讲话让“精准医学”(precision medicine, 有时被翻译为“精准医疗”)迅速成为了新年伊始世界各国关注的热点。据中国医师协会官方报纸《医师报》3 月 26 日报道：北京天坛医院副院长王拥军教授日前透露，科技部召开了国家首次精准医学战略专家会议，中国精准医疗计划将在 2015 下半年或明年启动。显然，我们有必要思考一下，为什么在这个时候要启动精准医学？怎样才能达到精准医学？

1　为什么要启动精准医学

20 世纪 90 年代初，以美国为主导的国际人类

收稿日期：2015-04-25
基金项目：国家自然科学基金项目(31130034，31470808)
通信作者：E-mail: wujr@sibs.ac.cn

基因组计划 (Human Genome Project, HGP) 启动，目标是测定人类拥有的遗传信息载体 DNA 上 30 亿个核苷酸的排列顺序。2001 年 2 月，人类基因组草图发布；2003 年 4 月 15 日，国际人类基因组计划负责人，现任美国国立卫生院主任 F. Collins 宣布，人类基因组序列图绘制成功。同日，美、英、日、法、德和中国等六国政府首脑联合发表声明，祝贺人类基因组计划的完成。

人类基因组计划之所以受到各国政府和科学家的高度重视，是人们希望通过这个计划破译人类的全部遗传信息，从而在保障人类健康和抗击疾病方面提供重大帮助。在 2003 年宣布人类基因组计划完成之后，研究人员就立刻启动了致力于人群水平的遗传变异研究的“国际人类基因组单体型图计划”，要揭示非洲、亚洲及欧洲人群的基因组变异谱图；在 2005 年完成的第一张谱图提供了超过 100 万个的“单核苷酸多态性” (single nucleotide polymorphism, SNP) 位点 [1]。2008 年 1 月，来自美国、英国、中国和德国的研究机构发起了“千人基因组计划”，目标是形成一个更加精细的人类基因组变异数据库；现阶段已经获得来自欧洲、非洲、东亚和美洲的 14 个人类族群的 1 092 个人基因组变异的高分辨率谱图，涵盖了 3 800 万个 SNP 位点、140 万个小片段核苷酸插入或缺失，以及超过 1.4 万个大片段缺失 [2]。

人们对基因组测序技术在临床的应用更是寄予厚望。例如在肿瘤研究领域，美国国立卫生研究院在 2006 年启动了一个耗资 1 亿美元的“癌症基因组图集” (The Cancer Genome Atlas, TCGA) 的科研项目，计划绘制出 1 万个肿瘤基因组变异图谱。2008 年，国际癌症基因组合作体 (International Cancer Genome Consortium, ICGC) 成立，随后有 16 个国家参加了肿瘤基因组变异图谱的研究计划；当时该组织的目标是，针对 50 种不同类型的肿瘤，每种肿瘤采集 500 份样品进行基因组测序研究。TCGA 项目在 2014 年底宣告完成，研究者发现了近 1 000 万种与肿瘤相关的遗传变异 [3]。通过对 TCGA 项目获得的21种癌症突变数据的统计分析，研究者表明，利用基因组测序方法能够找到一些临床上有用的突变位点 [4]。

然而，在基因组测序工作迅速推进的同时，研究者也逐渐认识到基因组知识的局限性。在人类基因组草图发表的第十个年头，人们发表了一系列文章来进行回顾和反思。美国 *Science* 杂志登载了一篇题为《等待革命》的评论文章，其主要观点是：“人类全基因组序列的测定并没有带来基础医疗方面的重大进展，由此促使人们去思考，是什么原因延缓了健康医学领域的基因组革命。” [5] 与此同时，英国 *Nature* 杂志也发表了一篇题为《最好的尚未到来》的社论，指出“在人类全基因组序列测定的十年之后，该计划提出的美好前景仍然需要去努力实现”[6]。

为什么人们的预期目标和现实情况有如此大的距离？在 *Nature* 杂志纪念人类基因组计划 10 周年专辑中，一篇题为《生命是复杂的》评论文章给出了答案：“生物学家看到的越多，显现的就越复杂。”[7] 虽然基因组的 DNA 序列是生物体的遗传基础，但生命活动并不是简单地依靠碱基序列就能够实现的。我们知道，现代生命科学建立在解释遗传信息传递的“中心法则”之上。长久以来人们相信，中心法则确保生物体遗传信息的“高保真性”：mRNA 序列必定严格由 DNA 序列决定，而蛋白质的氨基酸序列也必定由 mRNA 序列上的遗传密码子所严格决定。然而，美国科学家不久前通过比较 27 个个体的免疫细胞的基因组 DNA 序列与转录组 RNA 序列，发现了上万个外显子单核苷酸差异 [8]。一个研究组通过敲除负责 RNA 编辑的酶 ADAR 发现，这种 RNA 与 DNA 序列形成广泛差异的现象被明显抑制，从而证明了在转录组水平上广泛存在的核苷酸变异是由大规模 RNA 编辑导致的；这些 RNA 变异序列因此被称为“RNA 编辑组”(RNA editome) [9]。RNA 编辑组的发现提示我们，仅仅知道基因组序列是不够的，还需要了解转录组 RNA 序列。

对转录组 RNA 的深入研究还发现了许多不依赖基因组 DNA 序列的变异现象。过去人们认为，因为每一个编码蛋白质的基因都有着明确的起始序列和终止序列，所以从一个基因转录形成的所有 mRNA 转录本彼此间的序列长度都应该是一致的。但是，2013 年发表在 *Nature* 杂志的一项工作对这个传统观点提出了挑战。一个研究组利用一种全新的方法将酵母细胞的近 6 000 个基因表达的近 200 万条转录本进行了精准地测序，发现大多数基因的各条转录本之间长短不一，即一个基因在转录过程中产生的数十或数百条转录本的长度有很大的变化。该文作者由此认为，“转录本的边界存在变异是一个基本的规则而并非是一个例外” [10]。

在人类基因组中，编码蛋白质的基因序列只占 1.5% 左右，其余的都属于非编码序列。这些基因组的大部分非编码序列在转录过程中都产生了相应

的非编码 RNA (no-coding RNA)，例如 microRNA、siRNA、piRNA、snoRNA 和 lncRNA 等。这些非编码 RNA 参与了调控各种生命活动的过程。近年来的研究还发现，在真核生物中编码蛋白质的外显子序列可以通过一种反向剪接反应，使外显子来源的 RNA 序列首尾连接形成环形 RNA[11]。也就是说，编码蛋白质的基因组序列有可能产生非编码的环形 RNA。中国科学院研究人员 2014 年在 *Cell* 杂志上发表一篇研究论文，揭示出在人源胚胎干细胞中存在近万条环形 RNA，并且发现了同一个基因可以产生多个环形 RNA 的可变环化 (alternative circularization) 现象 [12]。这些研究工作表明，转录组虽然源于基因组，但 RNA 的复杂性和多样性显然是不能从基因组 DNA 序列简单推导出来的。

根据 RNA 制造蛋白质的翻译过程同样也不是过去人们想象的那样一种“高保真”过程。据估计，蛋白质翻译错误广泛存在于从大肠杆菌到哺乳动物之中 [13]，其发生概率大约从千分之一到万分之一 [14]；在平均长度为 400 个氨基酸的蛋白质群体中，大约有 18% 的蛋白质包含至少 1 个氨基酸的错义替换 [15]。研究者认为，某些物种拥有高频的翻译错误是进化选择决定的，这些翻译错误可以帮助这些生物体增加其蛋白质组的多样性，从而能够更好地适应环境变化 [13]。导致蛋白质翻译异常的分子机制目前还不是很清楚。在 2015 年，*Science* 杂志报道了一个在真核细胞中发现的非中心法则的翻译机制：酵母细胞在蛋白质翻译过程中，能够利用一种核糖体调控蛋白 Rqc2p，直接招募携带丙氨酸或者苏氨酸的 tRNA 进入核糖体 60S 大亚基的接受氨酰 -tRNA 的“A”位上，实现不依赖 mRNA 链的新合成蛋白链的延伸 [16]。显然，这种在蛋白质合成过程中不依赖 mRNA 添加氨基酸的现象已经不能简单地用“翻译错误”来表述了。

单个碱基变异引起的单核苷酸多态性 (SNP) 是个体在基因组水平变异的最普遍方式；在基因编码区的 SNP 还包含两个亚类：不改变氨基酸序列的同义 SNP 和改变氨基酸序列的非同义 SNP[17-18]。由此引出了在表型水平的个体差异的另一类分子：蛋白质氨基酸序列差异，即蛋白质上的单氨基酸多态性 (single amino-acid polymorphism, SAP)[19-20]。不久前，中国科学院研究人员首次利用靶向蛋白质组定量方法，揭示了汉族人群血浆样本中 SAP 在群体水平的定量特征，表明不同的 SAP 分别与生理或者病理性状存在不同的相关性 [21]。因此，人群的分子多态性研究不能仅仅停留在基因组水平的分子多态性，蛋白质组水平的分子多态性也需要加以关注。

由于在 RNA 转录和蛋白质翻译过程中存在着广泛的变异，因此蛋白质的氨基酸差异如 SAP 的来源显然就不会仅仅局限于基因组的 SNP，完全有可能来自 RNA 编辑组或者是蛋白质的翻译错误 [22]。也就是说，在蛋白质组水平上出现的许多 SAP 很有可能不依赖于基因组 DNA 序列。中国科学院研究人员最近发展了一种不依赖于基因组 SNP 信息而直接检测蛋白质水平氨基酸差异的蛋白质组技术，并应用该技术在人脑样本中找到了许多与基因组核酸序列无关的单氨基酸差异 [23]。肿瘤基因组计划 TCGA 在 2012 年报道了结直肠癌的基因组分析 [24]；在此基础上研究者又进行了结直肠癌的蛋白质组分析，在 86 个肿瘤样本中总共鉴定出了 796 个单氨基酸变异 (single amino acid variants, SAAVs)[25]。值得关注的是，研究者发现了 162 个以前没有报道过的全新的单氨基酸变异，占全部单氨基酸变异总数的 20%；他们认为这些新的氨基酸变异中至少有一部分可能就来自于 RNA 编辑 [25]。这些工作进一步表明，个体差异的分子水平研究不能停留在基因组 DNA 序列分析。

综上所述，生命的复杂性远远不是简单地测定基因组核酸序列就能够阐明的。即使只从中心法则直接涉及到的 DNA、RNA 和蛋白质分子水平来看，基因组核酸序列不过是生命复杂性的“冰山一角”；更不用提，生命复杂性涉及到表观遗传现象，以及代谢小分子和糖脂的参与；更不用提，生命复杂性还涉及到细胞、组织和器官等不同层次。*Cell* 杂志在 2014 年 3 月为纪念创刊 40 周年发行了专辑，其主题就被定为“复杂性”(complexity)。在该专辑中，美国著名肿瘤生物学家 R. Weinberg 发表了一篇题为《完整的循环 —— 从无尽的复杂性到简单性再回到复杂性》的评论文章，着重指出：在过去的 40 年里，从事肿瘤研究的科学家从最初面对无数难以理解的病理现象的困惑到树立了还原论必胜的信念，再到最近几年重新面对肿瘤这个疾病无尽的复杂性 [26]。

面对当前生物医学领域急需解决的生理和病理的复杂性问题，有人看到了挑战，有人看到了机会。“精准医学”正是在生命科学和医学实践处于这样一个重要转折点时应运而生。美国科学院研究理事会在 2011 年发布了一本有 100 多页的研究报告《迈向精准医学 —— 构建生物医学研究的知识网络和

新的疾病分类法》(以下简称《迈向精准医学》)，第一次明确提出了“精准医学”的概念并系统地讨论了为实现该目标所需要开展的核心任务 [27]。在该报告的作者看来，要想实现“精准医学”的前提是，构建基于生物学大数据的生物医学研究知识网络和基于分子生物学的全新疾病分类方法；通过建立一个整合了各种类型的生物学数据和知识、以个体为中心的信息共享平台，就可以形成用来获取对个人健康具有决定性的高度复杂影响因素或发病机理的生物医学知识网络；而利用生物医学研究知识网络将有助于建立新的疾病分类体系，从而定义新型疾病或对疾病进行分子分型和药物分层，实现对疾病的精确诊断和准确治疗 (图 1)。该报告的作者强调指出，“所提议的疾病知识网络和新分类法带来的主要收益正是‘精准医学’”[27]。

2　怎样才能达到精准医学

《迈向精准医学》的作者认为，“知识网络的建立以及对其进行研究和临床应用，都取决于是否拥有一个大型的、多层级的、充分整合的关于人类疾病的知识数据库”[27]。在这样的数据库里，人类疾病知识不仅包含了临床诊断和病理分析等表型信息，还具有各种生物分子信息，包括基因组、转录组、蛋白质组、代谢组、脂质组和表观遗传组等 (图 1)。也就是说，开展精准医学的基础是需要有尽可能完整的个体生物学数据。美国国立卫生院主任 F. Collins 和美国国立癌症研究所所长 H. Varmus 在描述拟开展的美国精准医学计划时表达了同样的观点：“我们准备建立一个有一定时间跨度的 100 万人以上的美国人群‘队列’，他们自愿参加该项研究。参加者被要求同意对其进行全面地生物学分析 (包括细胞种类、蛋白质、代谢分子、RNA 和 DNA，当经费允许时可进行全基因组测序) 和行为分析，所有这些分析数据都将连接到他们的电子健康档案。”[28]

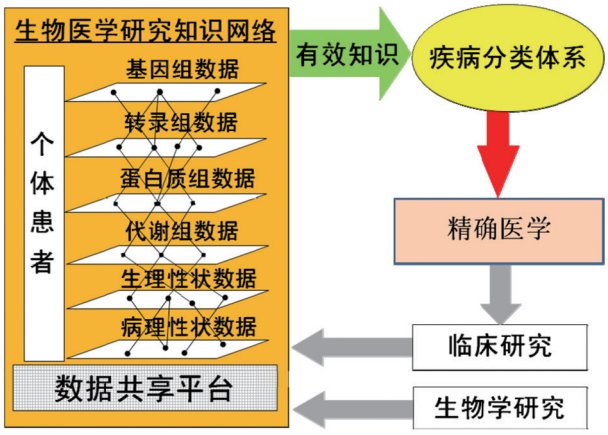

图1　精准医学与个体为中心的数据库以及疾病分类之关系示意图

这种数据库并不是一个把某一种类型的生物学数据简单地收集在一起，形成像 GenBank 那样的常规生物信息学数据库。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，那么这个数据库就是由很多变量组成的多层级的结构，每一层包含一个与疾病相关的变量信息 (图 1)。需要强调的是，利用生物信息学和计算生物学技术，人们能够发现各种分子之间的相互关系，建立起各种不同类型生物学数据层之间的高度的内部连结，从而形成一个复杂的生物医学知识网络 (图 1)。例如，基因组的突变与表观遗传改变相联系，或者与蛋白质组表达变化相联系，等等。理想的情况下，每个信息层都与其他信息层形成紧密的联系。这种不同种类生物分子之间、生物分子与表型 / 临床症状之间的高度整合将有利于人们发现传统方法不能挖掘到的致病因子或者诊断标记物，有利于人们对特定的个体患者进行准确地个性化诊断和治疗。

显然，这样的生物医学知识网络反映出来的正是系统生物学的核心特征 —— 多变量的整合。系统生物学 (systems biology) 是 21 世纪生命科学领域出现的一门新兴的交叉学科。系统生物学的创始人之一美国科学家 L. Hood 认为，系统生物学的特点是研究一个生物系统中基因、mRNA、蛋白质等所有组成成分的构成以及在特定条件下这些组分间的相互关系 [29]。因此，系统生物学的核心就是整合：首先是要把生物系统内不同种类的分子组成成份整合在一起进行研究；其次，对于多细胞生物而言，系统生物学还要实现从基因到细胞、到组织、到个体的各个层次的整合。也就是说，“迈向精准医学”需要构造的生物医学知识网络是建立在系统生物学的基础之上。

欧盟委员会 (European Commission) 为了在医学领域推进系统生物学，专门成立了一个“系统医学行动协调组织”(Coordinating Action Systems Medicine Consortium, CASyM)，涉及到 9 个欧洲国家的研究组织、基金会和企业。2014 年 6 月，欧盟委员会发布了《CASyM 路线图》，包括了近期 (2.5 年) 和长期 (10 年) 开展系统医学 (systems medicine)

的研究规划[30]。该报告指出，“系统医学就是将系统生物学的方法策略应用到医学概念、研究和实践之中”[30]。该报告认为，“系统医学将在下一个10年围绕着以患者为中心这个概念来进行医疗研究和实践，这些活动的开展需要整合不同的学科，包括数学、计算机科学、数据分析、生物学以及临床医学、伦理和社会实践”[30]。显然，这份路线图与《迈向精准医学》报告称得上是“异曲同工”。

需要强调的是，以患者为中心的观念也正是《迈向精准医学》的作者建设疾病知识数据库和知识网络的关键——“需要强调的是，这个信息共享平台的新颖性和能力就在于以‘个体为中心’”[27]。该报告认为，这种数据库的信息储存和管理与现存的生物学数据库有着根本的不同。目前国际上最大的生物学数据库是美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)。NCBI管理的各种数据库一般只包含一个单一类型的生物学数据或者一种疾病变量，例如，基因组数据保存在GenBank，而转录组数据则存储在Gene-Omnibus数据库；即使是来自同一个人的不同信息也会进入到多个数据库中，这些信息在不同数据库间并没有联系。也就是说，一个研究人员难以知道这些数据是来自同一个人；如果一个个体中决定疾病状况的涉及到多个变量，那么根本不可能抽提出它们之间的关系[27]。与此相反，“精准医学”所需要的数据库，就是要在从一个个体获取的各种类型的生物学数据之间建立起高度的内部连结(图1)。

如何建立以个体为中心的数据信息库？*Cell*杂志在2012年发表的一篇文章可以作为一个范本。美国斯坦福大学科学家M. Snyder对自己进行了连续14个月的表型监测和血液样本分析，获得了表型组谱、基因组序列、转录组表达谱、蛋白质组表达谱和代谢组表达谱等一个完整的个体“多组学”数据，并通过生物信息学的工具将这些不同种类的数据进行整合，建立了“整合的个人多组学谱”(integrative personal omics profile, iPOP)[31]。作为类似的工作，2014年3月，L. Hood和他领导的系统生物学研究所发起了“The Hundred Person Wellness Project”，计划用9个月的时间，选择100个健康人进行从分子到表型的个体化多组学研究[32]。L. Hood认为，“这种个体化组份的基础在于：每个个体在遗传和环境方面都是独一无二的，在不同时间段需要用他们自己作为对照(n = 1)来分析个体从健康到疾病的转变”[33]。该研究所计划在未来的5到10年内启动一个更大的称为“100K”的研究计划，要针对10万个健康人来开展这种多组学研究工作[33]。美国国立卫生研究院在2015年计划启动的精准医学计划也是以个体为中心的多组学数据整合研究，只是将研究的人数扩大到了100万[28]。

也就是说，以个体为中心的、整合了不同数据层的生物学数据库和高度关联的知识网络是迈向精准医学的必要条件。“‘精准医学’是用来为每个个体提供可得到的最好医疗护理。如果不对研究者和医疗保健提供者所依赖的信息系统进行巨大的重新定位，是无法达到这个目标的。这些信息系统就像它们准备支持的医学类型那样必须是个体化的。普遍性必须建立在大量个体信息的基础之上；而与这样一个过程相反的做法都将会失败。显然，如果在分析调查过程刚刚开始时就将生物分子表达谱、个体特定情况相关的数据和健康史从个体中剥离出来，那么要用来判定健康和疾病决定因素所必需的信息就会丢失。”[27]

3 结语与展望

通过以上讨论，我们可以看到，“精准医学”是一个有着丰富内涵的复杂概念，需要人们认真地思考和谨慎地解读。例如，“精准医学”不能简单地等同于“个体化医学”(personalized medicine)，因为中医是个体化医学，但不是精准医学；又例如，基因组测序是实现“精准医学”的主要任务之一，但不能把实现“精准医学”局限于基因组测序。另一方面，我们更要认识到，“精准医学”的出现将对生物医学研究和医疗实践产生重大影响，有可能改变人类维护健康和抗击疾病的传统模式。

[参 考 文 献]

[1] The International HapMap Consortium. A haplotype map of the human genome. Nature, 2005, 437: 1299-320
[2] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1092 human genomes. Nature, 2012, 491: 56-65
[3] Ledford H. End of cancer atlas prompts rethink. Nature, 2015, 517: 128-9
[4] Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature, 2014, 505: 495-501
[5] Marshall E. Waiting for the revolution. Science, 2011, 331: 526-9
[6] Editorial. Best is yet to come. Nature, 2011, 470: 140
[7] Hayden EC. Life is complicated. Nature, 2010, 464: 664-7
[8] Li M, Wang IX, Li Y, et al. Widespread RNA and DNA

sequence differences in the human transcriptome. Science, 2011, 333: 53-8
[9] Bahn JH, Lee JH, Li G, et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res, 2012, 22: 142-50
[10] Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. Nature, 2013, 497: 127-31
[11] Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs area large class of animal RNAs with regulatory potency. Nature, 2013, 495: 333-8
[12] Zhang XO, Wang HB, Zhang Y, et al. Complementary sequence-mediated exon circularization. Cell, 2014, 159: 134-47
[13] Pouplana LR, Santos MA, Zhu JH, et al. Protein mistranslation: friend or foe? TiBS, 2014, 39: 355-62
[14] Ogle JM, Ramakrishnan V. Structural insights into translational fidelity. Annu Rev Biochem, 2005, 74: 129-77
[15] Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell, 2008, 134: 341-52
[16] Peter S, Shen PS, Park J, et al. Rqc2p and 60*S* ribosomal subunits mediate mRNA-independent elongation of nascent chains. Science, 2015, 347: 75-8
[17] Cargill M, Altshuler D, Ireland J, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet, 1999, 22: 231-8
[18] Li Y, Vinckenbosch N, Tian G, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet, 2010, 42: 969-72
[19] Cavallo A, Martin AC. Mapping SNPs to protein sequence and structure data. Bioinformatics, 2005, 21: 1443-50
[20] Valentine SJ, Sevugarajan S, Kurulugama RT, et al. Split-field drift tube/mass spectrometry and isotopic labeling techniques for determination of single amino acid polymorphisms. J Proteome Res, 2006, 5: 1879-87
[21] Su ZD, Sun L, Yu DX, et al. Quantitative detection of single amino acid polymorphisms by targeted proteomics. J Mol Cell Biol, 2011, 3: 309-15
[22] Wu JR, Zeng R. Molecular basis for population variation: From SNPs to SAPs. FEBS Lett, 586: 2841-5
[23] Su ZD, Sheng QH, Li QR, et al. *De novo* identification and quantification of single amino-acid variants in human brain. J Mol Cell Biol, 2014, 6: 421-33
[24] The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature, 2012, 487: 330-7
[25] Zhang B, Wang J, Wang X, et al. Proteogenomic characterization of human colon and rectal cancer. Nature, 2014, 512: 382-7
[26] Weinberg RA. Coming full circle-from endless complexity to simplicity and back again. Cell, 2014, 157: 267-71
[27] National Research Council. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. Washington, DC: National Academies Press, 2011 (http://www.nap.edu/catalog/13284/)
[28] Collins FS, Varmus H. A new initiative on precision medicine. New Engl J Med, 2015, 372: 793-5
[29] Hood L. A personal view of molecular technology and how it has changed biology. J Proteome Res, 2002, 1: 399-410
[30] The CASyM Consortium. The CASyM roadmap: Implementation of systems medicine across Europe. 2014 (http://www.casym.eu/publications)
[31] Chen R, Mias G I, Li-Pook-Than J, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell, 2012, 148: 1293-307
[32] Gibbs W. Medicine gets up close and personal. Nature, 2014, 506: 144-5
[33] Hood L, Price ND. Demystifying disease, democratizing health care. Science Transl Med, 2014, 6: 1-3