从基因组学到放射组学的启示

郑钧正

〔作者简介〕 郑钧正,研究员,教授,博士生导师。1965~1993年工作于中国医学科学院/中国协和医科大学放射医学研究所,曾任副所长6年;后调至中国疾病预防控制中心的辐射防护与核安全医学所。现受聘客座教授于清华大学、复旦大学、首都医科大学等多家单位。1992年获国务院政府特殊津贴。1978~2013年荣获省部级以上的科技成果奖21项(其中负责的国家级奖3项);迄今在国内外正式刊物发表230多篇中、英文学术论文;出版了17本专业书;负责研制起草并报批发布18项国家标准和国家职业卫生标准等。自1998年以来先后以第一、第二作者申请并获批准8项国家自然科学基金的面上项目与重点项目。研究方向为电离辐射剂量学、放射医学与防护、医学物理、核安全、辐射监测等。

中图分类号 R3 文献标识码 A DOI 10.11969/j.issn.1673-548X.2016.02.001

新科学理论、新科技发明、新工程技术往往更多产生在相关学科的交叉结合部,因为有关学科不断交叉融合以及又再分化催生出新分支学科,势必推动科学技术向更深层次和更高水平发展,必然就不断促进科技进步和造福社会^[1]。当历史演进到历经不断积累雄厚基础有条件使科技腾飞的20世纪,诸如核能利用与核科学技术广泛普及,人类首次登上月球,发现DNA双螺旋结构而开启分子生物学时代,20世纪90年代人类基因组计划的实施及重大进展等,均是体现现代科学与人类文明突飞猛进的里程碑式创举。同时又孕育着21世纪的物理学、信息科学、生命科学、材料科学等将有新的突破和必将给人类生活带来崭新巨变。仅纵览当下非常活跃的生命科学(含生物医学)而言,一系列相关组学(omics)的相继陆续崛起并迅速发展尤为突出,非常值得予以充分关注和高度重视。

一、人类基因组计划与基因组学

在荣膺1962年诺贝尔生理学医学奖的沃森(James Dewey Watson)、克里克(Francis Harry Compton Crick)和威尔金斯(Maurice Hugh Frederick Wilkins),于1953年发现DNA双螺旋结构之后。相继于1958年和1980年罕见地两次荣获诺贝尔化学奖的桑格(Frederick Sanger),先后完整定序了胰岛素的氨基酸序列和发明很重要的DNA测序方法,这些划时代的杰出成就于20世纪后半叶完全“打开了分子生物学、遗传学和基因组学研究领域的大门”。于

是20世纪80年代形成了基因组学,在随后20世纪90年代人类基因组计划实施并取得很大进展后,基因组学取得了惊人的长足进展^[2]。

基因(gene)是DNA(脱氧核糖核酸)分子上具有遗传特征的特定核苷酸序列的总称,系具有遗传物质的DNA分子片段。基因位于染色体上,并在染色体上呈线性排列。基因不仅可以通过复制把遗传信息传递给下一代,还可以使遗传信息得到表达。例如不同人种之间头发、肤色、眼睛、鼻子等不同,是基因差异所致。基因是生命遗传的基本单位,不仅是决定生物性状的功能单位,还是一个突变单位和交换单位。由30亿个碱基对组成的人类基因组,蕴藏着生命的奥秘。

基因组(genomes)是一个物种的完整遗传物质,包括核基因组和细胞质基因组。即基因组是生物体内遗传信息的集合,是某个特定物种细胞内全部DNA分子的总和。显然原先只关注单个基因是远远不够的,应当深入研究整个基因组,于是产生了基因组学。

基因组学(genomics)是专门从分子水平系统研究整个基因组的结构(以全基因组测序为目标)、功能(以基因功能鉴定为目标)以及比较(基于基因组图谱和序列分析对已知基因和基因的结构进行比较)的分支学科。基因组学着眼于研究并解析生物体整个基因组的所有遗传信息,突出特点是必须以整个基因组为研究对象,而不是只研究单个基因;同时还要研究如何充分利用基因在各个领域发挥作用。基因组学概括起来涉及基因作图、测序和整个基因组功能分析的遗传学问题。这门分支学科交叉融合了

分子生物学、计算机科学、信息科学等,并以全新视角探究生长与发育、遗传与变异、结构与功能、健康与疾病等生物医学基本问题的分子机制,同时提供基因组信息以及相关数据系统加以利用,进而解决生物、医学和生物技术以及相关产业领域的有关问题^[3]。基因组学的主要目标包括认识基因组的结构、功能及进化规律,阐明整个基因组所涵盖遗传物质的全部信息及相互关系,为最终充分合理利用各种有效资源,以提供预防和治疗人类疾病的科学依据。

人类基因组计划(human genome project,HGP)的确立和实施极大地促进了基因组学的发展。人类基因组计划的提出,可追溯到寻求新方法解决日本广岛长崎原子弹幸存者及其后代的基因突变率检测低于预期问题。1984年12月美国能源部资助召开的环境诱变和致癌物防护国际会议,第一次提出测定人体基因和全部DNA序列,并检测所有的突变,计算真实的突变率。1985年6月,美国能源部正式提出了开展人类基因组测序工作,形成了“人类基因组计划(HGP)”的初步草案。历经几年酝酿与论证,1988年美国国会批准拨款,支持这一被誉为完全可以与“曼哈顿原子弹计划”、“阿波罗登月计划”并列相比美的宏伟科学计划。1990年正式启动后,陆续扩展成为美国、英国、法国、德国、日本和中国共同参加的国际性合作计划^[4]。2000年人类基因组工作框架图(草图)完成,是人类基因组计划成功的标志。

HGP这项规模宏大,跨国又跨学科的大科学探索工程。旨在测定组成人类染色体(指单倍体)中所包含的30亿个碱基对所组成的核苷酸序列,从而绘制人类基因组图谱,并且辨识其载有的基因及其序列,达到破译人类遗传信息,解码生命奥秘,探索人类自身的生、老、病、死规律,揭示疾病产生机制以提供疾病诊治的科学依据^[5]。截至2005年,人类基因组计划的测序工作已经完成,但基因组学等研究工作一直在不断深入和扩展。例如,2006年启动了肿瘤基因组计划力求揭示人类癌症的产生机制以及癌症预防与治疗的新理念。当下已经迈进后基因组时代,从揭示生命所有遗传信息转移到在分子整体水平上对功能的研究(功能基因组学)^[6]。21世纪的生命科学以新姿态和新方法阔步向着纵深发展,同时有力推进了基础与临床医学、生物信息学、计算生物学、社会伦理学等相关学科的蓬勃发展。为促进这些相关学科及其应用的更好发展,尤其推动在人类健康与疾病、个性化医疗、农业、环境、微生物等诸多领域的广泛应

用,自2006年以来已经召开了十届国际基因组学大会(ICG)。第10届国际基因组学大会于2015年10月在中国深圳举行,特别就临床基因组学、生育健康、癌症、衰老、精准医疗、人工智能与健康、农业基因组学、合成生物学、生命伦理和社会影响、相关组学及生物产业等热点问题进行深入研讨,展现了相关组学的旺盛活力。

二、转录组学、蛋白质组学、代谢组学等与基因组学相辅相成

基因组学作为研究生物基因组的组成,组内各基因的精确结构、相互关系及表达调控的科学,又必须从系统生物学角度与方法,着眼于整体出发去研究人类组织细胞结构、基因、蛋白质及其分子间相互作用,并通过整体分析研究人体组织器官的功能代谢状态,从而才能更有效地探索解决人类疾病发生机制及其诊治与保健问题^[2]。

虽然人类基因组图揭示了人类遗传密码,而对生命活动起调节作用的是蛋白质。基因组研究本身不能体现蛋白质的表达水平、表达时间、存在方式以及蛋白质自身独特活动规律等。因此,自从基因和基因组学问世以后,分子生物学的组学大家庭中,不断延伸分化形成了相互密切关联的转录组学(transcriptomics)、蛋白质组学(proteomics)、代谢组学(metabolomics),以及脂类组学(lipidomics)、免疫组学(immunomics)、糖组学(glycomics)、RNA组学(RNAomics)等,这些相互密切关联的组学构成丰富的系统生物学以及组学生物技术基础^[7]。

转录组学是一门在整体水平上研究细胞中基因转录情况以及转录调控规律的分支学科。也即转录组学是从RNA水平研究基因表达的情况。转录组即一个活细胞所能转录出来的所有RNA的总和,是研究细胞表型和功能的一个重要手段。可见在整体水平上研究所有基因转录及转录调控规律的转录组学,乃是功能基因组学研究的重要组成部分。

蛋白质组(proteome)是指一个基因、一个细胞或组织所表达的全部蛋白质。而蛋白质组学研究不同时间、空间发挥功能的特定蛋白质及其群体;从蛋白质水平上研究蛋白质表达模式和功能模式及其机制、调节控制及蛋白质群体中各个组分。蛋白质组本质上指的是在大规模水平上研究蛋白质的特征,包括蛋白质的表达水平,翻译后的修饰,蛋白与蛋白相互作用等,由此获得蛋白质水平上的关于疾病发生,细胞代谢等过程的整体而全面的认识。基因组相对稳定,

而蛋白质组是动态的概念。研究蛋白质组学是基因组学研究不可缺少的后续部分,也即生命科学进入后基因时代的特征^[2]。

代谢组学的概念源于代谢组,代谢组是指某一生物或细胞在一特定生理时期内所有的低分子量代谢产物。代谢组学则是对某一生物或细胞在一特定生理时期内所有低分子量代谢产物同时进行定性和定量分析的一门新分支学科。代谢组学以组群指标分析为基础,以高通量检测和数据处理为手段,以信息建模与系统整合为目标的系统生物学的一个分支。继基因组学和蛋白质组学之后新发展起来的代谢组学,是借助基因组学和蛋白质组学的研究思想,对生物体内所有代谢物进行定量分析,并寻找代谢物与生理病理变化的相对关系。基因组学和蛋白质组学分别从基因和蛋白质层面探寻生命的活动,而实际上细胞内许多生命活动是发生在代谢物层面的。因此有研究者认为“基因组学和蛋白质组学告诉你什么可能会发生,而代谢组学则告诉你什么确实发生了”。所以,代谢组学迅速发展并渗透到诸多领域,例如疾病诊断、医药研制开发、营养食品科学、毒理学、环境学、植物学等与人类健康密切相关的各领域^[8]。

三、放射组学在交叉融合中应运而生

2015年是伦琴发现X射线120周年,正如简明不列颠百科全书所评价:X射线的发现“宣布了现代物理学时代的到来,使医学发生了革命”^[9]。近40多年来计算机科学技术的交叉融合,以X射线透射开始并不断拓展许多种类型的医学成像技术,又经历了数字化革命而呈现出跨越式发展。数字化医学影像学已经成为现代医学不可或缺的重要手段和必不可少的组成部分。医学影像学在保健查体、疾病预防、疾病筛查、早期诊断、病情评估、治疗方法选择、康复疗效评价等,以及生命科学研究方面发挥了越来越大的不可替代作用^[10]。随着多排螺旋CT、双源CT、能谱CT、磁共振成像(MRI)、单光子和正电子计算机断层显像(SPECT与PET)、图像融合一体机成像(PET/CT等等)诸多影像医学新设备、新技术、新方法层出不穷,医学影像学已经从结构成像发展到功能成像,又迈向分子影像学的新阶段。尤其进入21世纪后,分子影像学方兴未艾地蓬勃发展,已经成为分子生物学的重要手段。当前数字化医学影像学所形成的大数据又密切相关到相关基因组学,应运而生了放射组学(radiomics)^[11]。如果说20世纪驱动医学

影像学的发展主要是依靠物理学和计算机科学技术、电子工程科学技术等,而21世纪则迫切需要与医学、分子生物学(包括基因组学等诸多组学)等相关学科进一步深入交叉融合相辅相成。

放射组学(亦有称之为影像组学)、分子影像学完全是与基因组学、蛋白质组学等相关组学彼此关联并相互促进而不断发展的。整合各种技术实现运用影像学手段显示人体组织水平、细胞和亚细胞水平的特定分子,并能反映活体状态下分子水平变化,从而对其生物学行为在分子影像层面进行定性和定量研究,无论在人体保健与疾病的诊断治疗,或者在药物研究开发,以及在基因功能分析与基因治疗研究等方面,都凸显了巨大优势和良好前景。

包含分子影像学的数字化医学影像学迅速发展,可提供越来越丰富的多层次医学影像数据资料,显然必须加以深度发掘并充分利用这些极其庞大的数字化信息^[12]。通过放射组学研究,解码隐含在医学影像信息中的因患者的细胞、生理、遗传变异等多因素共同决定的综合影像信息,并客观且量化将其内涵呈现在临床诊治、预后分析的整个过程,这无疑会成为临床医学具有重大意义的革命^[13]。应运而生的放射组学,就是致力于应用大量的自动化数据特征化算法将感兴趣区域(region of interest, ROI)的影像数据转化为具有高分辨率的可发掘的特征空间数据。数据分析是对大量的影像数据进行数字化的定量高通量分析,得到高保真的目标信息来综合评价肿瘤的各种表型(phenotypes),包括组织形态、细胞分子、基因遗传等各个层次。例如近期文献报道,放射组学可揭示肿瘤预测性的信号,能够捕获肿瘤内在的异质性,并与潜在的基因表达类型相关联^[14]。

美国的国家癌症研究所(National Cancer Institute, NCI),已经建立量化研究网络(quantitative research network, QIN),旨在共享数据、算法和工具,以加速影像信息量化的合作研究网络^[15]。他们将放射组学的建设及应用框架分为5部分:①图像的获取及重建;②图像分割及绘制;③特征的提取和量化;④数据库建立及共享;⑤个体数据的分析^[13]。当然这些均是很有挑战性的工作。

放射组学通过标准化的图像获取以及自动化的图像分析等,能为疾病的诊断、预后及预测提供有价值的信息。近期的研究还提示放射组学能有效预测不同患者中的肿瘤基因异质性等,可见放射组学有着广阔应用前景。

四、发展相关组学更好共促精准医疗

从基因组学、转录组学、蛋白质组学、代谢组学等直到新形成的放射组学,均是在相关学科交叉融合中,当条件与时机发展到一定程度而瓜熟蒂落催生。这些相互关联的组学全部都兼备着学科分化以及整合的特色。学科交叉融合根据发展需要分化催生出新分支,而所有这些组学分支学科又都从系统生物学角度出发,注重对形成的分支学科自身整体开展研究。正是如此辩证统一的现代科技发展特点,如同DNA的螺旋结构一样在不断深化中而螺旋式上升,推动科学技术向更深层次和更高水平发展。

侧重生物学角度来纵览这些组学分支学科及应用,都是密切关系到促进精准医疗问题。精准医疗(precision medicine)本质上就是个性化医疗(personalized medicine),必然需要基因组学及放射组学信息。就是要用每个个体的遗传学信息来指导对其疾病的预防、诊断和治疗。精准医疗基于系统生物学和基因组学、蛋白质组学、代谢组学等组学知识网络,针对个体疾病在分子水平分析疾病成因根源,利用靶向技术进行个性化诊断治疗的过程。个性化其实就是医学实践的正常形式,而分子水平信息的运用必然会使医疗更精准。精准医疗已成为国际上医学领域研究热点,代表了未来医疗行业的发展方向。实现精准,首先在于高精度测量技术和高灵敏度医疗检测设备,体内检测和成像技术将是关键核心技术;其次是数据的获取处理和综合挖掘能力。因此,包括基因组学与放射组学在内的相关组学,必须更深入研究探索,特别要用系统生物学整合各种组学的海量数据。

参考文献

1 路甬祥. 认知客观规律,促进科学发展[M]. 中国科学院. 2008 科

学发展报告. 北京:科学出版社, 2008:i-vi

- 2 Watson JD(沃森), Baker TA, Gann A, et al(杨焕明译). 基因的分子生物学[M]. 北京:科学出版社, 2009
- 3 Peltonen L, Mckusick VA. Genomics and medicine: dissecting human disease in the postgenomic era [J]. Science, 2001, 291(5507): 1224-1229
- 4 冷方伟. 中国基因组学研究进展与发展态势[J]. 生物化学与生物物理进展, 2010, 12:1261-1264
- 5 陈竺. 基因组科学与人类疾病[M]. 北京:科学出版社, 2001
- 6 王琳芳, 方福德. 疾病基因组学研究——后基因组时代的主旋律[J]. 中国医学科学院学报, 2002, 24(3):217-218
- 7 张雁明, 邢国芳, 刘美桃, 等. 全基因组关联分析:基因组学研究的机遇与挑战[J]. 生物技术通报, 2013, 33(6):1-6
- 8 周秋香, 余晓斌, 涂国全, 等. 代谢组学研究进展及其应用[J]. 生物技术通报, 2013, 33(1):49-55
- 9 《简明不列颠百科全书》中美联合编审委员会. 简明不列颠百科全书, 第5卷[M]. 北京:中国大百科全书出版社, 1986:459
- 10 郑钧正. 医学影像数字化所面临新课题应对当议[J]. 世界医疗器械, 2010, 16(2):72-75
- 11 Virendra K, Yuhua G, Satrajit B, et al. Radiomics: the process and the challenges [J]. Magnetic Resonance Imaging, 2012, 30(9):1234-1248
- 12 Lambin P, Rios - Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical imaging using advanced feature analysis [J]. Eur J Cancer, 2012, 48(4):441-446
- 13 苏会芳, 周国锋, 谢传森, 等. 放射组学的兴起和研究进展[J]. 中华医学杂志, 2015, 95(7):553-556
- 14 Arts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach [J]. Nat Commun, 2014, 5:4006
- 15 Levy MA, Freymann JB, Kirby JS, et al. Information methods to enable sharing of quantitative imaging research data [J]. Magnetic Resonance Imaging, 2012, 30(9):1249-1256

(收稿日期:2015-12-05)

(修回日期:2015-12-10)

欢迎订阅 2016 年《医学研究杂志》

《医学研究杂志》(原名《医学研究通讯》)于1972年创刊,是由国家卫生和计划生育委员会(原卫生部)主管、中国医学科学院主办的国家级医学学术期刊。本杂志为“中国精品科技期刊”以及“领跑者 F5000——中国精品科技期刊顶尖论文”来源期刊,中国科技论文统计源期刊,中国科技核心期刊,世界卫生组织西太区医学索引(WRPIM)收录期刊。月刊。CN11-5453/R, ISSN1673-548X。

本杂志信息量大,装帧精美。每册定价10元,全年120元(含邮费)。国内外公开发售。邮发代号:2-590。编辑部电话:010-52328677, 52328678, 52328679;传真:010-65230946。投稿网址: <http://www.yx-yjzz.cn>。编辑部地址:北京市朝阳区雅宝路3号(邮编:100020)。